REPUBLIC OF TURKEY
MUĞLA SITKI KOÇMAN UNIVERSITY
INSTITUTE OF EDUCATIONAL SCIENCES
FOREIGN LANGUAGES EDUCATION DEPARTMENT
ENGLISH LANGUAGE TEACHING PROGRAM

USING THE COH-METRIX TOOL FOR LINGUISTIC ANALYSIS OF MUĞLA SITKI KOÇMAN UNIVERSITY'S PREPARATORY CLASS STUDENTS' ESSAYS IN ORDER TO PREDICT AND COMPARE THEIR LANGUAGE PROFICIENCY LEVELS

VILÉM NÁJEMNÍK

MA THESIS

SEPTEMBER, 2017
MUĞLA

T.C.
MUĞLA SITKI KOÇMAN ÜNİVERSİTESİ
EĞİTİM BİLİMLERİ ENSTİTÜSÜ
YABANCI DİLLER EĞİTİMİ ANABİLİM DALI
İNGİLİZ DİLİ EĞİTİMİ BİLİM DALI

USING THE COH-METRIX TOOL FOR LINGUISTIC ANALYSIS OF MUĞLA SITKI KOÇMAN UNIVERSITY'S PREPARATORY CLASS STUDENTS' ESSAYS IN ORDER TO PREDICT AND COMPARE THEIR LANGUAGE PROFICIENCY LEVELS

VILÉM NÁJEMNÍK

Eğitim Bilimleri Enstitüsünce

"Yüksek Lisans"

Diploması Verilmesi İçin Kabul Edilen Tezdir.

Tezin Sözlü Savunma Tarihi: 02.10.2017

Tez Danışmanı: DOÇ. DR. EDA ÜSTÜNEL

Jüri Üyesi: DOÇ. DR. ŞEVKİ KÖMÜR

Jüri Üyesi: YRD. DOÇ. DR. SELAMİ OK

Enstitü Müdürü:

EKİM, 2017

TUTANAK

Muğla Sıtkı Koçman Üniversitesi Eğitim Bilimleri Enstitüsü'nün 27/08/2017 tarih ve 210/.. sayılı toplantısında oluşturulan jüri, Lisansüstü Eğitim-Öğretim Yönetmeliği'nin (26/4) maddesine göre, İngiliz Dili Eğitimi Bilim Dalı Yüksek Lisans öğrencisi Vilem Najemnik'nin "Using The Coh-Metrix Tool for Linguıstic Analysis Of Muğla Sıtkı Koçman Unıversity's Preparatory Class Students' Essays In Order To Predict And Compare Their Language Proficiency Levels'' (Muğla Sıtkı Koçman Üniversitesi Hazırlık Sınıfı Öğrencilerinin Dil Yeterlik Düzeylerinin Dilbilimsel Bir Analiz Yazılımı olan Coh-Metrix ile İncelenmesi) başlıklı tezini incelemiş ve aday tarihinde saat 10.00...'de jüri önünde tez savunmasına alınmıştır.

Adayın kişisel çalışmaya dayanan tezini savunmasından sonra 60 dakikalık süre içinde gerek tez konusu, gerekse tezin dayanağı olan anabilim dallarından sorulan sorulara verdiği cevaplar değerlendirilerek tezin **kabul** edildiğine oybirliği... ile karar verilmiştir.

Doç. Dr. Eda ÜSTÜNEL

Tez Danışmanı

Doç. Dr. Şevki KÖMÜR

Üye

Yrd. Doç. Selami OK

Üye

ETİK BEYANI

Muğla Sıtkı Koçman Üniversitesi Eğitim Bilimleri Enstitüsü Tez Yazım Kılavuzuna uygun olarak hazırlanan "Using The Coh-Metrix Tool for Linguistic Analysis Of Muğla Sıtkı Koçman University's Preparatory Class Students' Essays In Order To Predict And Compare Their Language Proficiency Levels'' (Muğla Sıtkı Koçman Üniversitesi Hazırlık Sınıfı Öğrencilerinin Dil Yeterlik Düzeylerinin Dilbilimsel Bir Analiz Yazılımı olan Coh-Metrix ile İncelenmesi) başlıklı Yüksek Lisans tez çalışmasında;

- Tez içinde sunulan veriler, bilgiler ve dokümanların akademik ve etik kurallar çerçevesinde elde edildiğini,
- Tüm bilgi, belge, değerlendirme ve sonuçların bilimsel etik ve ahlak kurallarına uygun olarak sunulduğunu,
- Tez çalışmasında yararlanılan eserlerin tümüne uygun atıfta bulunarak kaynak gösterildiğini,
- Kullanılan verilerde ve ortaya çıkan sonuçlarda herhangi bir değişiklik yapılmadığını,
- Bu tezde sunulan çalışmanın özgün olduğunu,

bildirir, aksi bir durumda aleyhime doğabilecek tüm hak kayıplarını kabullendiğimi beyan ederim. 02 / 10 / 2017

VILÉM NÁJEMNÍK

Bu tezde kullanılan ve başka kaynaktan yapılan bildirişlerin, çizelge, şekil ve fotoğrafların kaynak gösterilmeden kullanımı, 5846 sayılı Fikir ve Sanat Eserleri Kanunu'ndaki hükümlere tabidir.

ABSTRACT

USING THE COH-METRIX TOOL FOR LINGUISTIC ANALYSIS OF MUĞLA SITKI KOÇMAN UNIVERSITY'S PREPARATORY CLASS STUDENTS' PORTFOLIOS IN ORDER TO PREDICT AND COMPARE THEIR LANGUAGE PROFICIENCY LEVELS

VILÉM NÁJEMNÍK

Master's Thesis / Department of Foreign Languages Teacher Education / English Language Teaching Education Department

Supervisor: Assoc. Prof. Dr. Eda ÜSTÜNEL

October 2017, 126 pages

This study aimed to investigate the phenomenon of automated evaluation of natural language texts, namely usage of the Coh-Metrix 3.0 system for automated linguistic analysis for that purpose. The study provides a brief literature review of relevant topics and gives a brief introduction to metrics provided by the Coh-Metrix 3.0 web tool analysis. The main part of the study tries to identify scores provided by the Coh-Metrix system that would correlate with scores produced by human evaluators using standardised grading called CERMAT, which has different dimensions of grading such as vocabulary or cohesion. In order to find out which Coh-Metrix scores would correlate with human scores, quantitative methodology, namely Linear Discriminant Analysis, was implemented. The data was obtained from 60 English preparatory class students' essays and 17 volunteers from ELT department freshmen of Muğla Sıtkı Koçman University who attended preparatory classes in the preceding school year. The study identifies strongest discriminators that play statistically significant role in subsequent prediction models and uses them to find accuracy under which is a single prediction model able to assign correct scores to the students' essays. In the study, the final prediction model could achieve accuracy higher than 60%. The study then uses the prediction model to try to assign levels of writing proficiency according to CEFR with accuracy over 70%. The study concludes that as up to the date of the study, linguistic analysis, as provided by the Coh-Metrix, should not be used for evaluation purposes, and suggests further research.

Key Words: Coh-Metrix, automated evaluation, natural language processing

ÖZET

MUĞLA SITKI KOÇMAN ÜNİVERSİTESİ HAZIRLIK SINIFI ÖĞRENCİLERİNİN DİL YETERLİK DÜZEYLERİNİN DİLBİLİMSEL BİR ANALİZ YAZILIMI OLAN COH-METRIX İLE İNCELENMESİ

VILÉM NÁJEMNÍK

Yüksek Lisans Tezi / Yabancı Diller Eğitimi Anabilim Dalı / İngiliz Dili Eğitimi Bilim Dalı

Danışman: Doç. Dr. Eda ÜSTÜNEL

Ekim 2017, 126 sayfa

Bu çalışma, doğal dil metinlerinin otomatik olarak değerlendirilme olgusunu, yani bu amaçla otomatik dil analizi için Coh-Metrix 3.0 sisteminin kullanımını incelemeyi amaçladı. Çalışma, ilgili konularda kısa bir literatür taraması yapıyor ve Coh-Metrix 3.0 web aracı analizi tarafından sağlanan metriklere kısa bir giriş yapıyor. Çalışmanın ana kısmı, Coh-Metrix sistemi tarafından sağlanan, insan değerlendiriciler tarafından üretilir. CERMAT olarak adlandırılır. Kelime dağarcığı veya kaynaşma gibi derecelendirmenin farklı boyutlarına sahip standartlaştırılmış derecelendirmeyi kullanarak, üretilen puanlarla ilişkili puanlar belirlemeye çalışmaktadır. Hangi Coh-Metrix puanlarının insan puanı ile köreleceğini bulmak için nicel metodoloji, yani Doğrusal Ayırt Etme Analizi uygulanmıştır. Veriler, bir önceki okul yılının hazırlık sınıflarına katılan 60 ingilizce hazırlık sınıfı öğrencisinin denemelerinden ve 17 yeni gönüllü Muğla Sıtkı Koçman Üniversitesi İngilizce Öğretmenliği Bölümünden alınmıştır. Çalışma, sonraki tahmini modellerde istatistiksel açıdan önemli bir rol oynayan en güçlü ayrımcıları tanımlar, bunları öğrencilerin denemelerine doğru puanlar atayabilen tek bir tahmini modelin altında doğruluğu bulmak için kullanır. Çalışmada, nihai tahmin modeli % 60'tan daha yüksek bir doğruluk elde edebildi. Çalışma daha sonra, CEFR'e göre yazma yeterlilik düzeylerini % 70'in üzerinde doğrulukla atamaya çalışmak için tahmini modeli kullanır. Çalışma, çalışma tarihine kadar, Coh-Metrix tarafından sağlanan dilsel analizin değerlendirme amacıyla kullanılmaması ve daha ileri araştırmalar önermesi sonucuna varıyor.

Anahtar Kelimeler: Coh-Metrix, otomatikleşmiş değerlendirme, doğal dil işleme

ACKNOWLEDGEMENTS

I would like to give my sincere gratitude to my advisor Assoc. Prof. Dr. Eda ÜSTÜNEL for her contribution, support and guidance through this process, as well as for her support and guidance throughout my whole master degree study program in Turkey in both academic and professional development outside her duties as my supervisor.

I would like to thank to Assoc. Prof. Dr. Şevki KÖMÜR for being supportive, helping me with my academic development and making the whole experience of studying in Turkey possible.

I would like to mention Yrd. Doç. Selami OK's considerable contribution to this thesis by providing wide and constructive feedback, without which it would be much harder to finalize the thesis. For this I would like express my sincere gratitude to him.

I would also like to thank to all who were great support for me whilst adjusting to a life in a foreign country, the people thanks to whom I consider Turkey being my second home. For this I would like to especially thank Ms. Demet KUTUCUOĞLU, MA, her family, Mr. Mohammed Altar CANBEK and the whole Canbek family.

My gratitude goes to people who were inspiration to me trough-out my years in university, Mgr. Jan SUK, Ph.D., PhDr. Věra TAUCHMANOVÁ, MA, Ph.D, and late PhDr. ThDr. Helena TAMPIEROVÁ, Th.D., may the God bless her soul.

Special thanks goes to my dear friends Merve MUTLU MA, and Emine SAYAR for helping me with formal aspects of this paper and for providing all information that would be otherwise hard to reach due to the language barrier.

I would like to thank my friends from all countries and continents for their understanding and support. Special thanks to my partner Bc. Tereza FABIÁNKOVÁ who supported me trough-out the process of writing this study.

Lastly, I would like to thank my father Ing. Petr NÁJEMNÍK, my mother Mirka ORLOVSKÁ and my brothers Tomáš and Adam for always providing limitless support.

Vilém Nájemník

TABLE OF CONTENTS

Page

CHAPTER III
METHODOLOGY

CHAPTER IV
RESULTS

CHAPTER V

DISCUSSION, CONCLUSION AND IMPLICATIONS

LIST OF TABLES

LIST OF APPENDICIES

CHAPTER I

INTRODUCTION

1.1. Background of the Study

Being able to write successfully is undoubtedly a crucial skill that learners of any language have to get a hold on their way on becoming successful in mastering the given language (Terk, 2014). Moreover, storing knowledge in the form of written sources has been historically seen as a necessity. Writing is, therefore, a vital skill for anyone who desires to study and learn in a university environment (Çubukçu, 2012). There is no surprise then, that academics, as well as practitioners agree that good written communication skills are an essential skill for academic as well as a professional career (Russ, 2016).

As writing skills in the population are always shifting, with time and technology, also are needs on populations' literacy. The Turkish Republic has in recent years made various efforts to increase literacy, to move from the basic level of literacy, as defined by UNESCO, higher through changes in teaching methods and adoption of the approaches giving priority to language and literacy. In recent years, it has been shown through research that only 19% of students own more than 25 books in Turkey. The money spent on books annually is 45 cents, children in Turkey range 26th out of 35 countries in terms of reading habits, and only 8% of the population going to libraries goes there to read books (Aşici, 2015). These aspects are evident later, in practice as some students lack basic knowledge about how to write appropriately on an academic level. In effect of this, the teachers of English language are tied to teach the language but also have them teaching their students the basic rules of writing academic texts.

Writing can be one of the most difficult aspects of EFL teaching due to its complexity, and it is a challenging task for both a teacher and a student to tackle even in ideal circumstances (Çubukçu, 2012). Writing and expressing one's thoughts is challenging even in the most natural and supportive environment of native language and safety of the classroom and therefore when teaching (and perhaps even more in self-learning process) psychological and emotional attitudes toward writing must be considered as well (Chastain, 1998). Writing is also crucial for language development and acquisition, learning vocabulary, getting a hold on and understanding grammar via large chunks of written language the students have to deal with while writing (Bello, 1997). Students have to get familiar with appropriate word choice, use of appropriate grammar (such as Subject-Verb agreement and tense) syntax, mechanics (e.g. punctuation or spelling) and good organisation of ideas into coherent and cohesive form. There are also aspects of higher levels of writing, such as focus on audience and purpose as well as discovering of meaning (Gebhard, 2006). It is also evident that demand for language skills and standards is higher than in for example speaking, for it requires more careful construction, more precise and varied vocabulary and more correctness of expression in general. Writing also gives more space and time for development of language for it enables a learner to devote more time to it to organise the utterances and learn all aspects of language in real time (Ur, 1996).

Muğla Sıtkı Koçman University, School of Foreign Languages, English language preparatory classes department is responsible for preparation of students who wish to undertake their studies in the English language. To give the students opportunity to learn how to write effectively, the department has had writing portfolios prepared for students. The School of Foreign Languages of Muğla Sıtkı Koçman University is aware of the necessity and importance of good writing communication skills, crucial for students' academic career, as well as professional. It follows steps to improve the overall situation of students. Despite that students may be having difficulty in achieving an adequate level of writing to perform in following studies. In some cases students are not well-prepared for writing in English; they enter university lacking writing, reading and reasoning abilities (Jameson, 2007).

To effectively tackle the problem, the self-learning portfolio for students called 'Enhancing Writing Skills – From Paragraph to Essay' was developed. The portfolio guides students from very elemental writing rules, such as rules of writing in paragraphs, and utterances, such as writing about one's self, to the final stage where the student should be able to independently write an intermediate language level essay, reports and cause and effect essays, with relative ease. It proved to be a success to a certain extent, but many teachers said that correcting students' writings began to be dramatically time-consuming. For this reason it have been argued that the portfolio might be supplemented using computer technologies to make writing training less time consuming for teachers and possibly more beneficial for students' learning.

This may turn to be considerably hard as Chastain (1998) explained, assistance during this challenging process to overcome hardships of writing is crucial, whether it is in pre-writing, writing or post-writing phases of writing activities.

Modern technology may prove useful at times for students who have to improve their language level as it is, but the technology may also provide a supplementary aid to teachers. As one of the most time-consuming aspects considering writing is the fact that a lot of teachers' time is devoted to correcting students' writing and giving feedback. Being able to use online and automated devices may prove useful in conserving time and rerouting it into other aspects of teaching.

It has been proposed in various studies (e.g. Connor, 1990; Engber, 1995; Ferris, 1994; Jarvis, 2002; Reid, 1992) that surface and text base analysis of L2 students' writing can yield data that may be later used to predict students' L2 proficiency. The text base analysis consists of text length, lexical diversity, word frequency and repetition. As pointed out by McNamara, Crossley & McCarthy (2010) however, little research has been done on deeper-level linguistic measures used in the field of predicting L2, or EFL proficiency level, and they addressed this research gap by using the computational tool Coh-Metrix.

The developers explain on the official website of Coh-Metrix that *''Coh-Metrix is a system for computing computational cohesion and coherence metrics for written and spoken texts. Coh-Metrix allows readers, writers, educators, and researchers to instantly gauge the difficulty of written text for the target audience.''* (McNamara, Louwerse, Cai, & Graesser, 2005) It could be, therefore, possible to use the system to gain extra data on students' writing with little to no effort for a teacher.

There have been several debates on the topic. McNamara, Crossley, & McCarthy in 2010 hypothesised, that such system, as Coh-Metrix, could, thanks to the range of metrics it provides, serve as a useful tool for grading written utterances of various texts. They conducted a study with a wide range of high-school level students, which aims to validate this theory. They have been succesful, and they could adequately predict students' levels of writing, and the program has attributed the ability to provide scores that may serve to predict human grading.

The theory was put to trial in 2016 by Perin & Lauterbach who tried it on native-speaking adults. They replicated the study on two sets of samples produced by native speaking university candidates, and they sought to verify findings of McNamara, Crossley & McCarthy

(2010). Their results could not replicate the success of McNamara, Crossley, & McCarthy in 2010, they, however, targeted ten scores, as provided by the Coh-Metrix, which had a relationship to the scores of low-proficiency and high-proficiency adults. The current study will try to add to these researches by changing samples to FL students of English in Turkish environment, extending the dimensions of evaluating and will try to correlate wider scale of the Coh-Metrix system's data with scores produced by human evaluators.

1.1. Significance of the Study

This study has been sparked from both practical, as well as theoretical reasons. With theoretical background laid for simplification of linguistic evaluation processes as provided by currently available linguistic analysis tools a practical outcome of the study is expected as well. In this particular case the analysis tool in question is the Coh-Metrix system. The system has proved its ability to provide detailed and accurate linguistic analysis data on numerous occasions before (e.g. Greaser, McNamara, Louwerse, & Cai, 2004; Kirkgöz & Ünaldi, 2012).

The theoretical significance of this study lays in further validation of the Coh-Metrix system abilities in the linguistic analysis field but also in exploring the current state of machine evaluation processes. Furthermore, this study hopes to lay the theoretical background for the further development of the so-called mechanism for computer evaluation of written exams of students of English as a foreign language.

From the practical point of significance, the study will use the linguistic analysis tools to get some concrete data on the language levels of students of the preparatory classes of the Muğla Sıtkı Koçman University and hopes to provide enough data to be used in corpora for the purpose of future studies. Moreover, if the theory can be proven, the study could have an impact on the use of automated systems by the teachers, which could, in turn have a positive influence on time spent by teachers for grading their students.

Most importantly, the study will attempt to find a correlation between standardised human scores on writing and indexes of automated linguistic analysis tool, the Coh-Metrix.

1.2. Purpose of the Study

The study aims to present a detailed investigation of machine semantic analysis evaluation and its correlation with a real human scoring of written texts produced by Turkish learners of

English, for which quantitative measures will be used. The study aims to validate the readiness of abilities of currently available systems of algorithms for real-life classroom purposes. As a secondary objective, this study tries to identify exact scores provided by the system that would correlate with human grading.

1.3. Research Questions

1) What, if any, relationship is there between linguistic features and usage of cohesive devices in final essays as rated by Coh-Metrix and grading provided by human evaluators?
   a. Is there a relationship between different parts of human grading (Cohesion, Vocabulary, Content score) and those of Coh-Metrix relevant to this grading?
2) Can the Coh-Metrix system be used for grading the final essays of Turkish FL students?
   a. Which parts of Coh-Metrix analysis system can be used for grading and to what extent?
   b. Can the system be used for targeting exact level of essays regarding Common European Framework of Reference for Languages?

1.4. The Scope of the Study

The scope of the study is limited to randomly selected samples of written essays of sixty students attending the English language preparatory classes at Muğla Sıtkı Koçman University's School of Foreign Languages during the school year of 2015-2016. The second set of samples consists of seventeen student-volunteers of the ELT department of the university, in their first year of training, in the school year 2016-2017. Only a negligible portion of the samples was produced by nationals other, then of Turkish origin, and all of the students undertook a year of English language preparatory classes.

The study will use quantitative research methods to find correlations between scores of human evaluators and scores on linguistic analysis as provided by the Coh-Metrix system. The main chunk of this task will be done using Linear Descriptive Analysis and prediction models produced by this technique.

1.5. Terminology

To avoid misleading information that may arise out of a different usage of language in the thesis and by a possible reader, it is necessary to set a precise terminology as it will be used in the paper.

Literacy used in this context as defined by Aşici (2015) and stands for the concept as accepted by UNESCO after 1987. In a 1997 research examination called PISA (Programme for International Student Assessment) developed by the Organisation for Economic Co-operation and Development (OECD) the term covers wide range of literacies (such as information literacy, culture literacy, history literacy, environment literacy, art literacy, finance literacy, and universal literacy) and not just in its basic meaning – ability to write and read

English as a Foreign Language (EFL) means learning English in a formal classroom setting, with limited or no opportunities for use outside the classroom, in a country English does not play a major role in internal communication. (Richards & Schmidt, 2010)

Discourse cohesion refers to the basis of an article. Cohesion method typically includes two types, grammatical and lexical cohesion and incorporates all kinds of term relationships. It is the main way of linking appropriate grammar or terms to form them into an article. It shows the relative semantic functions, constructing the unique aesthetic feeling of the language. (Li, 2013)

Coh-Matrix is a web tool developed and designed by McNamara, Louwerse, Cai, & Graesser (2005) The purpose of the web tool is to provide data on texts fed to it instantly. The program yields over 100 different types of linguistic data on linguistic qualities in several categories.

Deep learning in the thesis is understood as in the field of machine learning, and it represents specific methodology used to teach algorithms to understand complicated matters and abstract principles of human interactions, for example, understanding of language (Deng & Yu, 2013)

Abbreviations

EFL: English as a Foreign Language

ELT: English Language Teaching

FL: Foreign Language

LDA: Linear Discriminant Analysis (Also known as Fisher's LD)

NLP: Natural Language Processing

MT: Machine translation

MTLD: Measure of textual language devices

CEFR: Common European Framework of Reference for Languages

vocd: Vocabulary diversity

CHAPTER II

REVIEW OF THE LITERATURE

2.1. Writing

In this section, history of writing, its importance, process of writing, cohesion and natural language processing issues will be briefly discussed.

2.1.1. Brief History of Writing

Written language is nowadays perceived as a common feature in our lives. We all are after all reading every day, may it be a billboard or a commercial, book or news. Therefore it may come as a surprise to some, that looking to writing from a longer time perspective; writing is a very new feature to the humankind.

If we look backwards, as early as modern people, Homo sapiens, are considered, we had emerged some 100,000 years ago. Since then modern humans have learned how to master agriculture, and began to develop first tools, but it was not until about 5,500 years ago when the first recorded written language is known to emerge, in the year 1999 at a place called Harappa, in the region where great Harappan or Indus civilisation once flourished. The meaning of symbols found is yet to be agreed upon; these records are however the Indus scripts – the first recognised written language (Harmer, 2004).

In the 5,500 years that came, countless protolanguages and languages developed their written systems, some of them into such a level that they are still being used although their spoken form disappeared centuries ago (e.g. Latin). However, for an extended time after it emerged, scribing was the occupation of a small minority of people. Writing and reading have been seen as a skill needed only for the management of Church and State and that is why writing was reserved almost exclusively for rulers of states and clergy. Too much knowledge in populations has been seen as a bad thing among labour force. Just some two hundred years ago or so, it became apparent that literacy – being able to write and read – as well as numeracy widespread in a population can mean the difference between a successful country or empire and an unsuccessful one. General literacy allowed the bureaucracy to work more efficiently than ever before. Education began to be desirable for the whole population, not just for the efficient running society anymore, but also for personal advancement and fulfilment of individuals (Harmer, 2004). Nowadays we consider writing not just as a virtue, but more than ever as a right of every individual. Being deprived of this right may lead to the person being excluded from a wide range of social roles, needless to say, that access to the positions of power is straight away impossible to get to without being able to write (Chris, 1997).

2.1.2. Importance of Writing

Children everywhere are acquiring spoken language from their surroundings. Ability to speak at least a language is therefore taken for granted. Learning how to write is, however, an academic skill, which needs hours of practice and mastering. As mentioned above, writing is, however, something we take as a right, but all over the World many people are deprived of this very right. According to the World Literacy Canada (WLC) organisation, there are at least 875 million illiterate adults in the world and at least a hundred million children who have little or no access to primary education.

We have to realise how deprived such individuals are. Would not there be for the personal development, illiterate people are still at a significant disadvantage to their literate counterparts. When we move to the education context itself, for example, we may realise that almost every school measures, stores and shares knowledge via written texts. When a school wishes to measure students' knowledge learned up to a point in the institution, in a foreign language or any other subject, they very often rely on students' writing proficiency in an examination. Needless to say that many jobs later in life except for those based on hard manual labour also require an ability to write. It is, therefore, undoubtedly a thing of high importance to improve the circumstances of the teaching of writing.

2.1.3. Process of Writing

Any written utterance, may it be a shopping list, a letter, an essay or a novel, undergoes a certain process of its creation. The process of any writing has several layers and may be affected by the content (subject matter), the type and the medium. In all these cases it is suggested that the process consists of four elements, as described by Harmer (2004) and listed below.

The elements as described below are listed in a way that may suggest that these are in linear, roughly like Planning -> Drafting -> Editing -> Final version. That is, however, in most cases not accurate. As most authors will often re-plan, re-draft, and re-edit and even reconsider the state of the final version of their writing, the process is rather recursive. Some authors may even leave the planning part out (e.g. stream of consciousness). Just after a final version becomes the final version, the process reached its end. Up until then, the author will typically move from an element to another one as required. (McDonald & Russel, 2002)

2.1.3.1. Planning

Before writing anything, most writers and people who do not consider themselves as writers will plan what it is that they are about to write. Some may make and write down detailed elaborative plan, some writers will go on with few simple lines, and some will just plan their writing in their mind. According to Harmer (2004), in the planning stage of the writing, writers have to consider three main issues.

Firstly, they have to think about the purpose of the writing. The purpose will amongst other things which have an impact on the type of text they will produce, also language that is to be used, and last, but not least the information they choose to include.

Secondly, the writer has to consider what kind of audience s/he is writing for. The audience will influence not only choice of language (formal or informal) but also overall shape of the utterance (how it should be laid out, structure of paragraphs, etc.)

Thirdly, the content structure of the utterance must be considered. Mainly, how facts, ideas, or arguments are going to be sequenced.

2.1.3.2. Drafting

Drafts are pieces of writing, which are usually not intended to be read by the audience. These may be pieces of text that are to be amended and later used or even discarded altogether. Before the process gets to the editing part, a considerable number of drafts may be produced before having sufficient amount of text to be able to create the final form.

2.1.3.3. Editing

After writing a sufficient draft or a number of them, the writer typically starts the editing part. He or she would normally go over the writing, seeing what works and what does not. Checking whether ideas are composed in a meaningful way, whether paragraphs are organised as it was desired or whether some parts have to be rewritten from scratch.

The writer may also rely on reflections of a fellow author or an editor, as an extra pair of eyes may provide valuable feedback on things that may have been missed out by the writer entirely.

2.1.3.4. Final version

After finishing editing drafts, the author gets the writing ready for sending it to the audience to read. The final version may be considerably different than what was originally planned at the planning stage, or from a first draft.

As mentioned above, it is not very important what type of a written utterance we are working on, may it be a simple e-mail, one’s memoirs or a to-do-list. We will typically plan what we are about to write, check what we have written and revised it before sending it to whomever it is meant to be sent to, no matter how casual.

2.1.4. Difference between a Written Utterance and Spoken One

Most of the time, differences between spoken and written utterances are easy to be marked, however, the differences may sometimes be minimal.

One could say that main difference is in time and space, as spoken communication is tied to here and now, whereas written piece can be preserved for hundreds or even thousands of years. On the other hand, if we, consider essentials and take a more detailed look, we may discover that differences are not as marginal as one could think. Taking modern media and text messaging into account, messages are often transferred in a fast phase, being deleted shortly after reaching its destination (chat, Instagram). In such cases, people may refer to texting as to ‘talking’ and indeed modern technologies may in some cases bridge this difference. Some speech may also take the quality of a written text, as it can be read aloud real-time.

When talking about participants, as Harmer (2004) reminds us of both spoken and written communication, the difference may seem to be easy to draw. In spoken communication, we usually have certain personal contact with the receiver of the message. We can often see them face-to-face, we hear their responses, and we can read their body language, which may lead to changes in all aspects of our spoken utterance. Also, typically in speaking, the participants may change roles and negotiate meaning etc. In written form, we generally assume who will be the receiving audience, and we adjust the written part accordingly. However, these qualities of both utterances may break; similarly into above mentioned time and space. Sometimes a speaker talking to a large audience will not be able to react to receivers' reactions, or just on a limited basis. In the same way, messaging technologies mentioned earlier break this boundary, as participants will very often switch roles and thus accustom their message or its form accordingly. This is a quality, which a novel writer or a journalist cannot enjoy.

A key feature of spoken communication is the process. In a basic communication, the participants can react instantly, they use the time employing techniques, and spoken form of language is often unstructured. As mentioned above, lots of forms of communication break this barrier. For instance, we have public speeches that may be rehearsed before the actual speech involving sort of 'drafts' and 'editing'. Writing is, however, more often easier to be structured correctly, for it is not as instant as speech, as a result, the writer has better chance to modify and plan his message before revealing it to whoever it is intended to. *'The process of writing is usually more complex than the process of speaking, but not always'* (Harmer, 2004).

We also have to mention distinct differences in the language used for each type of communication.

2.2. Cohesion

Considering the research methods of the thesis currently presented, we should get an insight on cohesion, as well as basics of writing, as it is the basis of the Coh-Metrix system all around. When we read or hear a longer chunk of language, that we master, we can typically quickly understand whether it forms a unified one or whether it is just a collection of unrelated sentences (Halliday & Hassan, 1976). This part is about the differences of the two.

2.2.1. Defining Concept of Text

To understand cohesion, we should first explain the concept of text. Halliday and Hassan (1976) in their book *Cohesion in English* describe a text as "*... used in linguistics to refer to any passage, spoken or written, of whatever length, that does form a unified whole.*" (p.1) And later they specify that a text may be anything that is spoken or written, long or short; a text is simply a unit of language in use. To them, a text is a semantic unit, which has its parts linked together by explicit *cohesive ties*. Cohesion, therefore, defines a text as a text. (Witte & Faigley, 1981)

2.2.1.1. Texture and ties

The property of being a text, as described above, can be entirely appropriately expressed by the concept of *texture.* Texture, as defined by Halliday & Hassan (1976) refers to a unity of text with its environment. Cohesive devices are in fact the resources that English uses to a total unity, thus creating texture. When we use *anaphoric*, *cataphoric* and *exophoric* references in a text, such as definite article or pronouns, we change or rather create, the texture of given text. These references are what give sentences cohesive relation, creating text.

References create presuppositions, and after, or while, satisfying these presuppositions, we give the text cohesive qualities, giving it texture, thus creating text. Halliday & Hassan (1976) use an opening sentence produced by a radio comedian *'So we pushed him under the other one.*' to demonstrate a sentence filled with presuppositions, references *so, him, other* and *one*, that are yet to be satisfied to give it cohesive qualities.

The connection between a reference and the item it refers to is called a *tie*, or as used by Witte & Faigley (1981) a *cohesive tie*. The tie can also be mere repetition of the item in the text. Repetition also has a cohesive effect, even if it does not refer to the same item. The concept of a tie gives us a tool to analyse a text's cohesive properties and systematic accounts of its patterns of texture.

2.2.1.2. References

The semantic connection between elements, reference, can be of three types. As mentioned above, Halliday & Hassan (1976) use terms *endophoric* and *exophoric* for references. The first one refers to a reference within the text. It can be later divided to an *anaphoric* reference, meaning reference to something already mentioned that is being referred to later (as in: I met

John. He is very nice – *He* anaphoricaly refers to *John* mentioned in the first sentence), and *cataphoric* for something that is yet to be referred to (as in: I met him before. His name is John. – *Him* in the first sentence is a cataphoric reference to John in second sentence). Exophoric reference is used to refer to thing outside the text. Witte & Faigley (1981) provide an example of exophoric reference in editorial ‘‘we‘‘ in a newspaper and they explain, that such references are exophoric because antecedent cannot be recoverable within the text.

2.3. Natural Language Processing

Writing, cohesion and automated systems for processing language are the main topics of this research. Let us now, briefly, have a closer look at natural language processing. The area of investigation that is called Natural Language Processing (NLP) explores how computers can be used to understand and manipulate natural language, written or spoken. The field includes technologies for machine translation, natural language text processing and summarisation, user interfaces, multilingual and cross-language information retrieval, speech recognition, artificial intelligence and expert systems. (Chowdhury, 2003)

The field’s origins, in concept at least, can be traced as far back as to seventeenth century, when first ideas of mechanical translators were proposed. Patents for devices that would mechanically convert words of one language to another were first issued in 1933 in France by Georges Artsrouni and in Russia by Peter Troyanskii (Hutchins J., 1993). It was not until 1940’s, however, when Machine translation (MT), the first computer-based application related to natural language, was proposed by Warren Weaver. (Hutchins J., 1997) Early pioneers of the field had a tremendous experience from the World War II, conducting breaking enemy codes. According to Liddy (2001), the Weaver’s memorandum of 1949 brought the idea of MT to general notice that led to many future projects. Weaver suggested that ideas from information theory and cryptography should be used for language translation.

Simplistic views on language influenced first attempts to employ MT. It was assumed that the only differences between languages reside in different vocabularies and word orders that are permitted within languages. Systems developed with these views only replaced words with their dictionary equivalent in the target language and restructured word order; needless to say, the results were poor, as the lexical ambiguity inherent in natural language has not been taken into account. The researchers realised that the task to translate natural languages is going to be much harder than anticipated and they lacked an adequate theory of language. This theory came in the form of idea of generative grammar, that was introduced in the book 'Syntactic Structures' (Chomsky, 1957) and provided better insight into how and if mainstream linguistics could help MT (Liddy, 2001).

Around this time other fields of NLP emerged, such as speech recognition. Due to the syntactic theory of language and parsing algorithms developments in the 1950's, there was too much enthusiasm that made people believe that fully automatic translation systems would be able to replace human translators in the scope of few years entirely. This was not only unrealistic due to the available linguistic knowledge and computer systems at the time, but also impossible in principle (Hutchins J., 1997). This was proven by the Automatic Language Processing Advisory Committee of National Academy of Science (ALPAC) that was formed by the U.S. government in 1965 to evaluate development up to then and in future (Pirce, et al., 1966). This research gained infamy for being very sceptical of research done on the field so far and led to cuts in government funding of NLP studies in the United States. Despite the substantial decrease in the research during the years after the ALPAC report, significant developments, in both theoretical issues and construction of prototype systems date to these times. In late 1960's and early 1970's academic work focused on how to represent meaning and development of computationally traceable solutions that available theories of grammar were unable to produce. Late 1970's saw the attention shift to semantic issues, discourse phenomena and communicative goals and plans and 1980's researchers re-examined non-symbolic approaches that lost popularity in early days (Liddy, 2001).

Years of 1990's saw a rapid growth in the field, mainly due to increased availability of large amounts of electronic text, availability of computers with increased speed and memory and the advent of the Internet (Liddy, 2001). Statistical approaches were re-introduced into the field, and this meant that many generic problems in computational linguistics such as part-of-speech identification, word sense disambiguation etc. were to be considered as standard in NLP field. By the end of 1990's emergence of new approaches in field of neural networks have taken the field of NLP and started to accelerate new development (Hochreiter & Schmidhubner, 1997).

2.3.1. Neural Networks and Deep Learning

Recent developments in machine learning and neural network technologies based on it have changed a great variety of research fields, mainly an emergence of algorithms known as Deep

Learning (Machine Learning). Deep learning is a class of machine learning algorithms that enable computers to do unsupervised learning and is a step bringing researchers closer to development of an AI. These algorithms can be defined by having multiple layers of nonlinear processing units and the supervised or unsupervised learning of feature representation in each layer, with the layers forming a hierarchy from low-level to high-level features (Deng & Yu, 2013).

For NLP, use of neural networks and deep learning enabled for machines to be trained to assess sentence similarity and detect paraphrasing (Socher & Manning, Deep Learning for NLP, 2013) and process statistical parsing that allows computers to cope with intuition of natural languages intuitive grammar (Socher, Bauer, Manning, & Ng, 2013). The machines may be also trained to assess sentiment analysis that gives machines the ability to systematically identify, extract, quantify and study affective states and subjective information (Socher, et al., 2013), information retrieval (Shen, He, Gao, Deng, & Mesnil, 2014),or spoken language understanding (Mesnil, et al., 2015).

The research by Socher, et al. (2013) is deemed especially groundbreaking. Over time, machines proved that they can understand the meaning of a complicated sentence. Sentence "This movie doesn't care about cleverness, wit or any other kind of intelligent humor" is for a human reader clearly of negative sentiment; however, a traditional algorithm would most likely detect it as positive, for words *care*, *cleverness*, *wit*, *kind, intelligent* and *humor* are of a positive sentiment. The deep learning algorithm in question understood that the sentence is negative.

We can speculate that further development of neural networks and deep learning algorithms will bring machines much closer to a full and better understanding of natural languages. When or if this goal is achieved, we may expect a shift in NLP as well, because as for today, most available linguistic analysis tools do not fully benefit from recent achievements in machine learning.

CHAPTER III

METHODOLOGY

This chapter presents the nature of the research, the selection of the participants, the instruments, the data collection procedure, and the data analysis procedure. Quantitative approaches were used for the analysis of the data obtained from the participants.

3.1. The Research Design

The study aims to explore capabilities of currently available algorithms used for linguistic analysis. To answer the research questions, authentic written texts to be analysed had to be collected first. After the collection, Discriminant Analysis was conducted to determine the relationship between human scoring and that of the linguistic analysis.

The current study was inspired by two preceding studies (McNamara, Crossley, & McCarthy, 2010 and Perin & Lauterbach, 2016). They used Linear Discriminant Analysis to target Coh-Metrix scores that may have prediction value considering relationships between human scores and those of the Coh-Metrix system. The first research tested the hypothesis on high-school grade students and in the study McNamara et al.(2010) were able to find three strong prediction scores that correlated with human scores. The second study by Perin & Lauterbach (2016) replicated the research finding ten different predictors and scoring 70% of successful prediction using prediction values they pinpointed as having strongest prediction values. The current study replicates these research designs on certain level, this study will evaluate 84 Coh-Metrix scores, instead of 52 that were analysed in original studies. Also, this study will test whether it is possible to correlate the Coh-Metrix scores with standardised scores that are currently in use (CERMAT), that have more diverse scores than ‘better‘ and ‘worse‘ as was the case in original studies and they also require to grade different qualities of texts separately (e.a. vocabulary, cohesion, content). However, the study will replicate the procedure, which means that it will systematically try every group of Coh-Metrix scores using ANOVAs and

Linear Discriminant Analysis to find correlations between the scores provided by the Coh-Metrix system and grades as produced by human evaluators.

The texts analysed come from three sources in total. First and the largest sample of sixty written texts were acquired from a pool of written final exams of students attending Muğla Sıtkı Koçman University's School of Foreign Languages, English language preparatory classes during the school year 2015/16. These final written exams were selected randomly.

Later, the second sample of seventeen writings was collected by the researcher from volunteers of the first year students of the English Language Teaching department of the same university, approximating length and topics of the first set of samples. The students also undertook the language preparatory year that is in the question of the first set of samples and received a year of training in English language writing and multiple other courses in order to improve their overall English language level. These samples will be used to validate the parts of analysis that will correlate with the human scores, as their writing level is expected to be of overall better quality due to their training.

The third part of samples comes from a publicly available set of 19 essays of University of Padua's English language classes. The third set of samples' authenticity cannot be guaranteed. However, the samples are related to the two main sets of samples as to the topic, as to extent and scope of the essays, and are corrected with precisely determined levels of B1 and B2 according to CEFR. These samples were used merely to validate the importance of variables that were deemed statistically significant for the prediction model created.

Before the machine grading itself, it was important to grade the samples by a human examiner. For this part of the research, two experienced English language teachers were approached to provide human grading for the samples, and the researcher graded the works as well. For accurate and synchronal grading, the standardised grading scale of written texts in a foreign language by CERMAT (Centre for Ascertainment of Outcomes of Education) was used, as its classification suits purposes of the study; moreover, both evaluators and the researcher have experience in using it. The first evaluator (referred to as 'Hscr1') followed analytical approach in grading. The second evaluator (Hscr2) was asked to grade the samples holistically and the researcher graded the essays analytically as well, in order to compare the outcomes (Note: The first two evaluators were absolutely independent in their grading, as they did not know origins of texts in question and probably never encountered area specifics of EFL in Turkish environment). The CERMAT grading is split into four separate categories (described in detail in 3.3.2), and their different relationships were tested.

All the samples were transcribed to an electronic form by the researcher and then graded by the evaluators and analysed by the Coh-Metrix system. To be able to distinguish which variables are provided by the Coh-Metrix are relevant to the human grading, discriminant analysis using SPSS was run in a similar fashion to previous research done on a similar topic (McNamara, Crossley, & McCarthy, 2010, Crossley & McNamara, 2012, Perin & Lauterbach, 2016). The scores in which a statistically significant relationship to human grades could be found were recorded and later, the benchmark samples were used to verify whether the variables can be used to build a robust enough prediction model (in previous researches, the hit ratio achieved was as high as 70%).

3.2. Participants

Participants of the study were 60 students from the School of Foreign Languages, preparatory classes of English, in year 2015/16, who were selected randomly. Also, 17 student volunteers attending the freshman year at the ELT department in 2016/17 academic year participated in the study and they had also attended the preparatory classes in 2015/16 academic year at Muğla Sıtkı Koçman University. Participants were both female and male.

3.2.1. Background of the Participants

The 60 participants are students whose skills in the English language are deemed to be insufficient for following their studies and have to recieve a year of intensive language training to enhance their language skills for academic purposes. In the school year of their preparatory classes, the students are expected to achieve a B2 level in English, according to CEFR, via a proficiency test, which is mandatory to pass by the end of their preparatory classes. Apart of the examination is a test in writing, which was used for this study.

The other part of samples, the seventeen freshmen students of ELT, had already passed the preparatory class via an exam and attended ELT classes for a year. The ELT training also includes obligatory courses such as Contextual Grammar I, Advanced Reading and Writing I in their first semester and Contextual Grammar II and Advanced Reading and Writing II. These subjects are studied together with other courses aiming at developing students' English skills, such courses as Verbal Communication Skills I, Listening and Pronunciation in the first semester and Lexicology or Verbal Communication Skills II in the second. The language level of these students may thus be expected to exceed that of the first group.

3.3. Instruments

The study will use quantitative research tools for answering the research questions. For collecting, machine evaluating and categorising data, the Coh-Metrix analysis system has been used. For human scores, CERMAT scale has been used. The data were then analysed using the Discriminant Analysis using IBM SPSS v.22.

3.3.1. Coh-Metrix

Coh-Metrix is a computational tool that has been designed to provide a broad range of language and discourse measures. It can be utilized by teachers, students, researchers and authors to gain information on numerous levels of language about their texts. Coh-Metrix was developed, refined and tested at the University of Memphis between years 2002 and 2011. The tool is accessible for free for research purposes (McNamara, Graesser, McCarthy, & Cai, 2014). In this part, some linguistic features measured by the Coh-Metrix and their interpretation will be briefly discussed.

3.3.1.1. Word information

Knowledge of vocabulary is ever shifting with ages of individuals and their development, which also has a substantial impact on reading time and comprehension. With the development, the words individuals encounter are changing. From an early age to adulthood, words in textbooks and the texts increase their complexity and number of unfamiliar words increases. The Coh-Metrix tool was therefore designed to provide an abundance of word measures that have relevance to reading development and the construction of meaning in a text. (McNamara, Graesser, McCarthy, & Cai, 2014)

The Coh-Metrix tool provides multiple information on words in the text, using extensive linguistic and psycholinguistic corpus, namely the MRC Psycholinguistic database (Coltheart, 1981). The database consists of 150,837 words and information is given about 26 different linguistic properties of these words, as Coltheart (1981) points out. However, the information about every property is not available for every one of the 150,837 words, as nobody has yet collected imagery ratings on such large set of words. The imagery ratings are thus available for only 9,240 words. The Coh-Metrix tool analyses the six MRC properties of words and values them on a range of 100 to 700. (Greaser, McNamara, Louwerse, & Cai, 2004).

The MRC properties are as follows:

A) *Familiarity* – (WRDFAMc) Frequency in which a word appears in print (Greaser, McNamara, Louwerse, & Cai, 2004), reflects on how are words familiar to an adult. MRC provides ratings for 3,488 unique words and an average rating for content words in a text are reported by Coh-Metrix. Words that are very familiar can be seen almost daily, have high scores (e.g. mother = 632, water = 641, milk = 588) whereas unfamiliar words, that are very rarely used get low ones (e.g. calix = 124, witan = 110, manus 113). (McNamara, Graesser, McCarthy, & Cai, 2014)

B) *Concreteness* – (WRDCNCc) Based on human ratings measures how abstract or concrete a word is

C) *Imageability* – (WRDIMGc) Based on human ratings reports how easy it is to construct a mental image in one's mind

D) *Colorado meaningfulness* – (WRDMEAc) Uses meaningfulness ratings from a corpus developed by Toglia and Bating (1978), multiplied by 100

E) *Paivio meaningfulness* – Meaningfulness rates based on Paivio, Yuille, and Madigan (1968) norms including 925 nouns that were recently extended (Clark & Paivio, 2004) and Bristol norms by Gilhooly and Logie (1980), multiplied by 100 – excluded in Metrix 3,0 that is used for the study

F) *Age of acquisition* – (WRDAOAc) Score based on Bristol norms (Gilhooly & Logie, 1980), multiplied by 100 (Greaser, McNamara, Louwerse, & Cai, 2004). The norms were compiled for 1 903 unique words. The Coh-Metrix reports the average ratings for content words. The fact that some words are acquired earlier (e.g. milk, smile or pony have age-of-acquisition scores 202) and some later (cortex, dogma, matrix have age-of-acquisition scores 700) is reflected here. (McNamara, Graesser, McCarthy, & Cai, 2014)

3.3.1.2. Parts of speech

Each word is assigned a syntactic part-of-speech category; those are sorted to content words (e.g. nouns [WRDNOUN], adjectives [WRDADJ], verbs [WRDVERB], adverbs [WRDADV]) and function words (e.g. determiners, pronouns, prepositions). As King, J (2015) points out, some words, can be assigned multiple syntactic categories. As an example, word "bank" can serve as a noun, such as in connection with 'river bank', or as the 'monetary bank', a verb as in 'They want to bank with the Chase around the corner' or an adjective 'bank shot'. Each word is assigned to only one part-of-speech category by the Coh-Metrix, based on its syntactic context. Coh-Metrix counts the number of instances of the category per 1,000 words of text to compute the relative frequency of it.

There are over 50 Part-of-speech's categories that are adopted form the Penn Treebank (Marcus, Santorini, & Marcinkiewicz, 1993) and the Brill POS tagger (Brill, 1992). The words that can be assigned to multiple part-of-speech categories are assigned the most likely one. The Brill POS tagger is a self-learning mechanism. Thus it can assign the most likely part-of-speech category of words it does not know.

The incidence of nouns, verbs, adjectives, adverbs and pronouns, which are segregated into the first-person singular (WRDPRP1s), first-person plural (WRDPRP1p), second-person (WRDPRP2), third-person singular (WRDPRP3s) and plural (WRDPRP3p) are counted. The distinction between different types of pronouns has important repercussions on other levels of meaning (McNamara, Graesser, McCarthy, & Cai, 2014).

3.3.1.3. Word frequency

Word frequency measurement indicates how often particular words are used in the English language. This is an essential measure because for how frequently is the term used in English indicates how quickly can be the text read, as well as how easy it can be understood. Words that are not used frequently in the language may cause the reader to read on a slower pace, and it may be harder to understand the meaning of the text. Word processing time tends to decrease linearly with the logarithm of word frequency, rather than with raw word frequency (Haberlandt & Graesser, 1985), because some words have extremely high frequencies (articles, is or and), whereas other words may be common, but not nearly as frequent. For that reason, logarithmic transformation is employed as they better fit a normal distribution and have a linear fit with reading times. (Greaser, McNamara, Louwerse, & Cai, 2004)

The word frequency is computed using four corpus-based standards. The first being CELEX database developed by the Dutch Centre for Lexical Information corpora, (Baayen, Piepenbrock, & Gulikers, 1995) that consists of frequencies taken from the 1991 version COBUILD corpus of 17.9 million words. The corpus includes written sources such as newspapers and books and spoken sources that include taped telephone conversations and BBC World Service. Second frequency count comes from the norms of Thorndike & Lorge (1944). The third corpus is taken from the Kučera-Francis norm (Francis & Kučera, 1982) and the fourth is the frequency count of spoken English analysed by Brown (Brown, 1984). Separate measures are computed for both raw and logarithm values in each of these. Content words (adverbs, lexical verbs, adjectives, nouns), function words (prepositions or determiners, etc.) and all words are computed separately as well (Greaser, McNamara, Louwerse, & Cai, 2004). In these metrics raw word frequency is displayed as WRDFRQc, the logarithm of word frequency of all words is shown as WRDFRQa, and minimum log word frequency is represented by WRDFRQmc output mark.

3.3.1.4. Lexical diversity

The Coh-Metrix system provides various scores on lexical diversity. The most well-know is the type-token ratio (TTR) (Templin, 1957). The TTR score represents the number of unique words in text divided by the overall number of words in the text. Because this method is sensitive on the total count of words, the Coh-Metrix also provides the Measure of Textual Lexical Diversity (MTLD) and *vocd,* which is a computational program taking different approach to show metrics of vocabulary diversity used within the text. These two are considered to be particularly reliable, as McCarthy and Jarvis (2010) summarize: *"We conclude by advising researchers to consider using MTLD, vocd-D (or HD-D), and Maas in their studies, rather than any single index, noting that lexical diversity can be assessed in many ways and each approach may be informative as to the construct under investigation."* (Abstract, p.381)

Coh-Metrix provides these indexes as LDTTRc for the TTR content word caption, LDTTRa for TTR of all words, LDMTLD for MTLD and LEXDIVVD for *vocd*.

3.3.1.5. Text readability levels

The Coh-Metrix tool provides three text readability level indexes that report how easy it is to read a text. Two traditional metrics are employed and one new index that has been developed by the Coh-Metrix team to report on second-language texts' readability.

The first conventional and popular metric provided is the Flesch-Kincaid Grade Level (RDFKGL) (Kincaid, Fishburne, Rogers, & Chissom, 1975). The Flesch-Kincaid Grade Level was developed for purposes of the U.S. Navy, that computes the readability levels. The formula for determination of the grade level is [(0.39*mean number of words per sentence {DESSL})+(11.8*mean number of syllables per word{DESWLsy}) – 15.59].

The second traditional index is the Flesch Reading Ease (RDFRE) (Flesch, 1948 as cited in McNamara, Graesser, McCarthy, & Cai, 2014) (Klare, 1974-1975) is calculated as [206.835-

(1.015* mean number of words per sentence{DESSL})-(84.6* mean number of syllables per word{DESWLsy})]

These measures are designed to provide predictions that can accurately predict the amount of time it will take to read a passage and sentence-level understanding of the measured text. These metrics are however rather simplified, as they provide only a single dimension of text difficulty (McNamara, Graesser, McCarthy, & Cai, 2014). They may, however, offer an easy way to pre-estimate students' level readiness, as we can estimate how difficult the text a student produced is, relative to texts of others in the same group.

The last metric on text readability metric has been developed by the Coh-Metrix team to provide readability formula based on psycholinguistic and cognitive models of reading. The metric reports on readability levels with respect to L2 learners' comprehension levels (beginner, intermediate, advanced) (RDL2) significantly better than traditional readability formulas (Crossley, Allen, & McNamara, 2011). This metric is counted by content word overlap, sentence syntactic similarity and word frequency. The formula takes into account not only sentence level and word level, but also the cohesion of the sentences in the text (McNamara, Graesser, McCarthy, & Cai, 2014).

3.3.1.6. Text readability and easability

The matrix which provides readability and easability comes from the deeper understanding of measures of the Coh-Metrix development team. It was introduced in version 3.0 of the system and categorizes eight statistics and reflect, as the name implies, how easy it would be for a reader to read the text measured (McNamara, Graesser, McCarthy, & Cai, 2014)

1. Narrativity (PCNARz, PCNARp). The texts gain scores depending on their narrativity. According to the authors, narrative texts tell stories, with places, events, characters and things that are familiar to the reader. This metric is highly affiliated with word familiarity, word knowledge and oral language.
2. Syntactic Simplicity (PCSYNz, PCSYNp). Texts gain scores in this metric depending on a number of words in sentences and on how they use familiar syntactic structures.
3. Word Concreteness (PCCNCz, PCCNCp). The score of PCCNC depends on content words that are concrete and meaningful. The metric also reflects abstractness of words, as abstract words are harder to process.

4. Referential Cohesion (PCREFz, PCREFp). Texts with higher referential cohesion will be typically easier to process.
5. Deep Cohesion (PCDCz, PCDCp). This statistic reports the degree to which the text contains causal and intentional connectives when there are causal and logical relationships with the text.
6. Verb Cohesion (PCVERBz, PCVERBp). This dimension reports the degree to which there are overleaping verbs in the text. A text with higher verb overlap will be likely more relevant for younger readers and narrative texts
7. Connectivity (PCCONNz, PCCONNp). The component reflects the number of logical relations in the text that are explicitly conveyed.
8. Temporality (PCTEMPz, PCTEMPp). Temporal cohesion reports a number of cues about temporality (i.e., aspect, tense).

3.3.1.7. Referential cohesion

The Coh-Metrix computes different types of cohesion relation, or coreference, between sentences by text base analysis where it identifies clauses. When a noun, pronoun or noun-phrase argument refers to another constituent in the text, we call this 'Referential cohesion'; if the content word in a sentence does not connect to another constituent in the text, the phenomena is called cohesion gap. (Halliday & Hassan, 1976)

The Coh-Metrix system tracks five major types of lexical coreference by computing overlap in morpheme units, content words, arguments, pronouns and nouns. (McNamara, Graesser, McCarthy, & Cai, 2014)

1) Morpheme unit overlap/Stem overlap – One sentence has a noun with same semantic morpheme in common with a word of any grammatical category in another sentence (e.g. to run – a runner). These are measured locally (CRFSO1), meaning between adjacent sentences and globally (CRFSOa0) in the whole text.
2) Content words overlap – More content words in different sentences overlap. This measures proportion of explicit content words that overlap between a pair of sentences. Locally measured content word overlap (CRFCWO1) and global measurement (CRFCWOa) indicate for example a sentence, that has fewer words that overlap in a shorter sentence, the proportion will be larger, than if the pair has many words and two words overlap. Indices also include their standard deviations (CRFCWO1d, CRFCWOad)

3) Argument overlap – Same nouns and pronouns are shared between sentences (a car/a car, it/it). The local argument overlap (CRFAO1) indicates how nouns and pronouns overlap in neighbouring sentences and global overlap (CRFAOa) reports on how nouns and pronouns overlap in the text as a whole.
4) Pronoun overlap – Sentences share at least one pronoun with same gender and number. This overlap is included in the argument overlap
5) Noun overlap – Sentences share at least one common noun. The Coh-Metrix gives a number that represents the average number of sentences that overlap locally (CRFNO1), which means they have to be adjacent to each other. In this measurement, to be counted, the nouns in different sentences must be exactly same. The second measurement of noun overlap (CRFNOa) computes the cohesion globally, how every sentence in the text overlaps with every other sentence in the text.

3.3.1.8. Coh-Metrix data table

The table that follows shows all categories of Coh-Metrix analysis, with their acronyms as used in the program and a short description.

Table 3.3
Coh-Metrix scores table, all variables

Category	Label	Name
Descriptive	DESPC	Paragraph count, number of paragraphs
	DESSC	Sentence count, number of sentences
	DESWC	Word count, number of words
	DESPL	Paragraph length, number of sentences in a paragraph, mean
	DESPLd	Paragraph length, number of sentences in a paragraph, standard deviation
	DESSL	Sentence length, number of words, mean
	DESSLd	Sentence length, number of words, standard deviation
	DESWLsy	Word length, number of syllables, mean
	DESWLsyd	Word length, number of syllables, standard deviation
	DESWLlt	Word length, number of letters, mean
	DESWLltd	Word length, number of letters, standard deviation
Text Easability Principle Component (PC) Scores	PCNARz	Text Easability PC Narrativity, z score
	PCNARp	Text Easability PC Narrativity, percentile
	PCSYNz	Text Easability PC Syntactic simplicity, z score
	PCSYNp	Text Easability PC Syntactic simplicity, percentile
	PCCNCz	Text Easability PC Word concreteness, z score
	PCCNCp	Text Easability PC Word concreteness, percentile
	PCREFz	Text Easability PC Referential cohesion, z score
	PCREFp	Text Easability PC Referential cohesion, percentile
	PCDCz	Text Easability PC Deep cohesion, z score
	PCDCp	Text Easability PC Deep cohesion, percentile
	PCVERBz	Text Easability PC Verb cohesion, z score
	PCVERBp	Text Easability PC Verb cohesion, percentile
	PCCONNz	Text Easability PC Connectivity, z score
	PCCONNp	Text Easability PC Connectivity, percentile

	PCTEMPz	Text Easability PC Temporality, z score
	PCTEMPp	Text Easability PC Temporality, percentile
Referential Cohesion	CRFNO1	Noun overlap, adjacent sentences, binary, mean
	CRFAO1	Argument overlap, adjacent sentences, binary, mean
	CRFSO1	Stem overlap, adjacent sentences, binary, mean
	CRFNOa	Noun overlap, all sentences, binary, mean
	CRFAOa	Argument overlap, all sentences, binary, mean
	CRFSOa	Stem overlap, all sentences, binary, mean
	CRFCWO1	Content word overlap, adjacent sentences, proportional, mean
	CRFCWO1d	Content word overlap, adjacent sentences, proportional, standard deviation
	CRFCWOa	Content word overlap, all sentences, proportional, mean
	CRFCWOad	Content word overlap, all sentences, proportional, standard deviation
LSA	LSASS1	LSA overlap, adjacent sentences, mean
	LSASS1d	LSA overlap, adjacent sentences, standard deviation
	LSASSp	LSA overlap, all sentences in paragraph, mean
	LSASSpd	LSA overlap, all sentences in paragraph, standard deviation
	LSAPP1	LSA overlap, adjacent paragraphs, mean
	LSAPP1d	LSA overlap, adjacent paragraphs, standard deviation
	LSAGN	LSA given/new, sentences, mean
	LSAGNd	LSA given/new, sentences, standard deviation
Lexical Diversity	LDTTRc	Lexical diversity, type-token ratio, content word lemmas
	LDTTRa	Lexical diversity, type-token ratio, all words
	LDMTLD	Lexical diversity, MTLD, all words
	LDVOCD	Lexical diversity, VOCD, all words
Connectives	CNCAll	All connectives incidence
	CNCCaus	Causal connectives incidence
	CNCLogic	Logical connectives incidence
	CNCADC	Adversative and contrastive connectives incidence
	CNCTemp	Temporal connectives incidence
	CNCTempx	Expanded temporal connectives incidence

	CNCAdd	Additive connectives incidence
	CNCPos	Positive connectives incidence
	CNCNeg	Negative connectives incidence
Situation Model	SMCAUSv	Causal verb incidence
	SMCAUSvp	Causal verbs and causal particles incidence
	SMINTEp	Intentional verbs incidence
	SMCAUSr	Ratio of casual particles to causal verbs
	SMINTEr	Ratio of intentional particles to intentional verbs
	SMCAUSlsa	LSA verb overlap
	SMCAUSwn	WordNet verb overlap
	SMTEMP	Temporal cohesion, tense and aspect repetition, mean
Syntactic Complexity	SYNLE	Left embeddedness, words before main verb, mean
	SYNNP	Number of modifiers per noun phrase, mean
	SYNMEDpos	Minimal Edit Distance, part of speech
	SYNMEDwrd	Minimal Edit Distance, all words
	SYNMEDlem	Minimal Edit Distance, lemmas
	SYNSTRUTa	Sentence syntax similarity, adjacent sentences, mean
	SYNSTRUTt	Sentence syntax similarity, all combinations, across paragraphs, mean
Syntactic Pattern Density	DRNP	Noun phrase density, incidence
	DRVP	Verb phrase density, incidence
	DRAP	Adverbial phrase density, incidence
	DRPP	Preposition phrase density, incidence
	DRPVAL	Agentless passive voice density, incidence
	DRNEG	Negation density, incidence
	DRGERUND	Gerund density, incidence
	DRINF	Infinitive density, incidence
Word Information	WRDNOUN	Noun incidence
	WRDVERB	Verb incidence
	WRDADJ	Adjective incidence
	WRDADV	Adverb incidence
	WRDPRO	Pronoun incidence
	WRDPRP1s	First person singular pronoun incidence

	WRDPRP1p	First person plural pronoun incidence
	WRDPRP2	Second person pronoun incidence
	WRDPRP3s	Third person singular pronoun incidence
	WRDPRP3p	Third person plural pronoun incidence
	WRDFRQc	CELEX word frequency for content words, mean
	WRDFRQa	CELEX Log frequency for all words, mean
	WRDFRQmc	CELEX Log minimum frequency for content words, mean
	WRDAOAc	Age of acquisition for content words, mean
	WRDFAMc	Familiarity for content words, mean
	WRDCNCc	Concreteness for content words, mean
	WRDIMGc	Imagability for content words, mean
	WRDMEAc	Meaningfulness, Colorado norms, content words, mean
	WRDPOLc	Polysemy for content words, mean
	WRDHYPn	Hypernymy for nouns, mean
	WRDHYPv	Hypernymy for verbs, mean
	WRDHYPnv	Hypernymy for nouns and verbs, mean
Readbility	RDFRE	Flesch Reading Ease
	RDFKGL	Flesch-Kincaid Grade level
	RDL2	Coh-Metrix L2 Readability

3.3.2. CERMAT Grading Scale

CERMAT stands for *Centrum pro zjišťování výsledků vzdělávání* or *Centre for Ascertainment of Outcomes of Education* in English. It is a public organisation that has been established by The Ministry of Education, Youth and Sports of the Czech Republic to provide testing and measurement of outcomes of the Czech education system (CERMAT, 2010). The organisation is in charge of national standardised testing of all graduating high school students (Maturita) by law (Ministry of Education, Youth and Sports of the Czech Republic, 2006). The organisation hence created, tested and uses a scale system for grading written texts in English, as it is a mandatory part of the final exam (Maturita).

The scale consists of four basic criteria as presented in the table below.

Table 3.4

CERMAT scale for grading

Criteria	Points
I. Content of written work a. Task fulfilment b. Scope, content of text	0-1-2-3
II. Organisation and text cohesion a. Organisation of text b. Text Cohesion and linking of the text	0-1-2-3
III. Vocabulary and spelling a. Precision of used vocabulary b. Scope of vocabulary used	0-1-2-3
IV. Use of grammatical devices a. Precision of used grammatical devices b. Range of grammatical devices used	0-1-2-3

For the purposes of this study, criteria IV. have been omitted, as grammar check is not part of the Coh-Metrix system, providing human scores on it is thus irrelevant.

The point system is fundamentally simple. If the criteria were not met at all, the student receives 0 points. For the criteria I., for example, the graded work would have to be off topic, having an insufficient number of words, or exceed it, and its content would have to be also other specified by entry points. If either of the criteria is met, the graded work receives 1 or 2 points. Depending on the extent to which the criteria is met (for example, criteria II., the text is cohesive, and linked well, but lacks paragraphs entirely, thus receives 1 point. When the text is cohesive, linked fairly, but the text is organised poorly, yet can be understood easily despite having weak structure, and receives 2 points). When the examined text fulfils all criteria, it receives 3 points. The overall score is the sum of all points received. (Ministry of Education, Youth and Sports of the Czech Republic, 2006)

In the CERMAT scale, each component is graded on scale 0-1-2-3, maximum of points that can be achieved in the written exam from a foreign language is 24 (8×3). Three points being highest achievement possible and 0 the smallest. As a standardised evaluation method, there are standardised guidelines to how an essay should be evaluated. Full description is shown in the table below.

Table 3.5
CERMAT grading criteria

	I - Content of written work	**II - Organisation and text cohesion**	**III - Vocabulary and spelling**	**IV - Use of grammatical devices**
	a - Task fulfilment	*a- Organisation of text*	*a- Precision of used vocabulary*	*a- Precision of used grammatical devices*
3	* Text characteristics are maintained * Points are mentioned clearly and are understandable	* The text is coherent with the linear sequence of ideas * The text is appropriately structured and organised	* Errors in vocabulary and morphology do not prevent text from being understood * Vocabulary and spelling are almost always used correctly	* Errors in grammar do not prevent understanding of the text * Grammatical devices are mostly used correctly
2	* Most characteristics are maintained * Most points are mentioned - The scale of the text is not entirely maintained (by one interval shorter/longer)	* Most text is coherent with the linear sequence of ideas. * Most text is appropriately structured and organised	* Errors in vocabulary and morphology usually do not prevent the understanding of the text / part of the text * Vocabulary and spelling are mostly used correctly	* Most errors in grammar do not prevent the text / part of the text from being understood * Grammatical devices are mostly used correctly
1	* Most text characteristics are not maintained * Most points are not mentioned and some of them are not understandable - The scale of the text is not entirely maintained (by two intervals shorter/longer)	* Most text is not coherent with the linear sequence of ideas * Most text is not appropriately structured and/or organised	* Vocabulary and morphologic mistakes largely prevent the understanding of the text / part of the text * Vocabulary and spelling are not used correctly to a greater extent	* Most errors in grammar prevent the text / part of the text from being understood * Grammatical devices are mostly used incorrectly

0	! Text characteristics are not maintained ! Points are not mentioned and are not understandable ! Scale of the text is not maintained (by three intervals and shorter) - Scale of the text is not maintained (by three intervals and longer)	* Most text is not coherent and does not contain a linear sequence of ideas * Most text is not appropriately structured and/or organised	* Vocabulary and morphologic mistakes prevent most of the text from being understood * Vocabulary and spelling are incorrectly used in most of the text	* Errors in grammar prevent understanding of the text * Grammatical devices are used incorrectly
	b - Scope, content of text	*b - Text Cohesion and linking of the text (CLT)*	*b - Scope of vocabulary used*	*b - Range of grammatical devices used*
3	* Points in the text are handled appropriately and in an appropriate degree of detail *The essence of an idea or a problem is clearly explained in the text	* Range of CLT is wide * Errors in CLT do not prevent understanding of the text * CLT is almost always used correctly and appropriately	* The vocabulary is wide	* Range of grammatical devices is wide
2	* Most points in the text are handled appropriately and in an appropriate degree of detail *The essence of an idea or a problem is mostly explained in the text	* Range of CLT is relatively wide * Errors in CLT do not prevent understanding of most of the text/part of text * CLT is mostly used correctly and appropriately	* The vocabulary is mostly wide	* Range of grammatical devices is mostly wide
1	* Most points in the text are not handled appropriately and in an inappropriate degree of detail *The essence of an idea or a problem is mostly explained in the text	* Range of CLT is relatively small * Errors in CLT prevent understanding of most of the text/part of text * CLT is mostly used incorrectly and inappropriately	* Vocabulary is mostly limited	* Range of grammatical devices are mostly limited
0	* Entry points are not worked out appropriately and to an appropriate degree of detail	* Range of CLT is limited * Errors in CLT prevent understanding of the text * CLT is used incorrectly and inappropriately	* Vocabulary is limited / insufficient	* Range of grammatical devices are limited/ insufficient

! – If the student earns 0 points in this part, the whole essay is dismissed with 0 points

This grading scale has been chosen, because previous research done on a similar topic simply tried to find a correlation between human scores marking texts as ‘better’ and ‘worse’, and individual scores were provided by the Coh-Metrix. As Coh-Metrix provides data on cohesion and vocabulary, it has been decided that this form of grading would serve the purpose of the study better, moreover, the scale used by CERMAT is used for grading final exams nationwide for years officially and can be therefore deemed validated.

3.3.3. LDA

Fisher's linear discriminant, commonly known as linear discriminant analysis (LDA) is a method that is in statistics, or machine learning used to find and separate two or more classes of events or objects (Fisher, 1936).

In particular, a multiclass LDA was used. This LDA generalisation has been made by Rao, R. (1948) and allows analysing a great portion of data of multiple groups at once and allowed to analyse relationships between portions of the Coh-Metrix data and the human scores.

3.4. Data Collection

The data for this thesis were collected from a set of three different pools and later quantified by instruments of Coh-Metrix linguistic analysis and human scores using human evaluators in order to apply quantitative research methods, namely the multiclass linear discriminant analysis.

The first set of samples of students written texts were retrieved in the form of electronic copies from the pool of archived final exam texts of students attending Muğla Sıtkı Koçman University's, School of Foreign Languages' English language preparatory classes. The volume of works retrieved is 60, and they were selected randomly, across students' future departments, to get writings as diverse as possible as to vocabulary and cohesive devices used.

The second set of samples was collected by the researcher from first-year students of the English Language Teaching department of the same university who volunteered. The samples were collected in the class of Advanced Reading and Writing II course, approximating the conditions in which the first set of samples were made, and the researcher attempted to set similar conditions as to time limitation, and allowed size of text, as well as topic. Seventeen essays were collected and analysed.

The third set of data comes from a small publicly available pool of short essays of the Padua University. Nineteen of these essays were collected over the internet, selected by their size and topic, to match the main two sets of samples. These samples were chosen mainly for their apparently set level according to the Common European Framework of Reference for Languages (CEFR). These essays were corrected and used as representatives of a 'best case results' for group of CEFR (B1 and B2). This portion of samples serves only as an auxiliary source of data and will not have an impact on outcomes of the study, however, they will be used as an independent set of samples for the prediction model that will be created by analysing the first two sets of samples to see whether the prediction model will work with samples other, than those used for the primary analysis.

All data were written by hand, so before any analysis, it was important to transfer all texts into an electronic form. After completing this task, all texts were separately fed to the Coh-Metrix web interface, and all results were saved. The data had to be transferred from an open office formatted output of the Coh-Metrix using Microsoft Excel 2010 to the IBM SPSS v.23 where linear discriminant analysis was conducted.

3.5. Data Analysis

To answer the research questions, the data obtained in the form of hand-written texts had to be quantified in order for quantitative methods to be implemented. This task has been completed through the Coh-Metrix web tool analysis. After the quantification, the data were transferred to the Microsoft Excel 2010 program, and means of all 106 datasets were created and categorised. After categorization, data were transferred to IBM SPSS v.23 together with human CERMAT scores and the Linear Discriminative Analysis (LDA) was used to find relationships between human scores and those of Coh-Metrix.

The data were analysed in relation to each other, as well as in relation to each of three human-made scores to find whether there are more prominent links to each of human scores separately.

To target the relationship accurately, the data were put into the LDA analysis in their separate categories, relative to those of human evaluators' as shown in Table 3.6. That allowed finding relationships between human scores and the machine ones to each other as well as in relation to each evaluator and the Coh-Metrix. The Syntactic Pattern Density values were not analysed, as there is no human score available to compare it with the Coh-Metrix scores. Apart from the hit ratio that the prediction model could achieve when the Coh-Metrix data and grades of human evaluators were entered, results of Wilks' lambda tests (a probability distribution used in multivariate hypothesis testing) were also reported. Wilks' lambda values report levels on which the discrimination model will discriminate the cases. If the *p-v*alue was lower than p= 0,05, the results are recorded, as it is the first indication of the value having an impact on the results and that there could be a relationship between the scores of the Coh-Metrix system and those of human evaluators.

The scores obtained were then compared to the benchmarks set by the benchmarking samples.

Table 3.6
Coh-Metrix and CERMAT scores attribution

Coh-Metrix 3.0 Analysis	CERMAT scores
Descriptive *(DESPC, DESSC, DESWC, DESPL, DESPLd, DESSL, DESSLd, DESWLsy, DESWLsyd, DESWLlt, DESWLltd)*	I
Text Easability Principle Component Scores *(PCNARz, PCNARp, PCSYNz, PCSYNp, PCCNCz, PCCNCp, PCREFz, PCREFp, PCDCz, PCDCp, PCVERBz, PCVERBp, PCCONNz, PCCONNp, PCTEMPz, PCTEMPp)*	I, II & III
Referential Cohesion *(CRFNO1, CRFAO1, CRFSO1, CRFNOa, CRFAOa, CRFSOa, CRFCWO1, CRFCWO1d, CRFCWOa, CRFCWOad)*	II
LSA *(LSASS1, LSASS1d, LSASSp, LSASSpd, LSAPP1, LSAPP1d, LSAGN, LSAGNd)*	II
Lexical Diversity *(LDTTRc, LDTTRa, LDMTLD, LDVOCD)*	III
Connectives *(CNCAll, CNCCaus, CNCLogic, CNCADC, CNCTemp, CNCTempx, CNCAdd, CNCPos, CNCNeg*	II
Situation Model *(SMCAUSv, SMCAUSvp, SMINTEp, SMCAUSr, SMINTEr, SMCAUSlsa, SMCAUSwn, SMTEMP)*	I & II
Syntactic Complexity *(SYNLE, SYNNP, SYNMEDpos, SYNMEDwrd, SYNMEDlem, SYNSTRUTa, SYNSTRUTt)*	I & II
Syntactic Pattern Density *(DRNP, DRVP, DRAP, DRPP, DRPVAL, DRNEG, DRGERUND, DRINF)*	X
Word Information *(WRDNOUN, WRDVERB, WRDADJ, WRDADV, WRDPRO, WRDPRP1s, WRDPRP1p, WRDPRP2, WRDPRP3s, WRDPRP3p, WRDFRQc, WRDFRQa, WRDFRQmc, WRDAOAc, WRDFAMc, WRDCNCc, WRDIMGc, WRDMEAc, WRDPOLc, WRDHYPn, WRDHYPv, WRDHYPnv)*	I & III
Readibility *(RDFRE, RDFKGL, RDL2)*	I, II & III

CHAPTER IV

RESULTS

To answer the research questions, we will concentrate on detailed results of the research in this chapter. Firstly, the correlation of human scores and those of the linguistic analysis will be explored through Linear Discriminative Analysis. The analysis will be firstly reported on the first set of samples, the preparatory classes' students (marked numerically 1 to 60), secondly on the first year ELT students' sample (marked numerically as A1-A17) in combination with the first sample set.

The manner of the report follows the order of data, as provided by the Coh-Metrix results, as shown in Table 3.5, from top to bottom (starting with Descriptive analysis, ending with Readability scores). Results are thus categorised into nine parts (Syntactic pattern density scores were left out, and scores on cohesion are in one sub-chapter), which are then separated into two parts each, reporting on results of the first group of samples and then on both of them together. Each part reports the results in relation to each human evaluator separately.

The second part of results tries to follow similar manner, using auxiliary data of clearly separated B1 and B2 samples and will try to verify results of the main body of samples. Lastly, after the verifying analysis, benchmarking results comparing the main body of data with the auxiliary ones will follow briefly.

4.1. Coh-Metrix Scores and Human Scores

This part of results chapter reports the relationship between Coh-Metrix analysis scores and those provided by human evaluators. As reports are detailed, a sum-up of each part is provided.

4.1.1. Descriptive scores

The descriptive part of the Coh-Metrix analysis provides descriptive scores, such as number of words, the number of paragraphs, number of sentences or mean of syllables etc. As creators of Coh-Metrix web tool claim, these results mainly serve for checking whether the system works correctly (Greaser, McNamara, Louwerse, & Cai, 2004). Relation between human scores and descriptive scores was thus expected to be minimal. The test was conducted comparing the content ratings (CERMAT I) of each of three human evaluators. Because descriptive functions of Coh-Metrix are not expected to have a real impact on any of scores, the results are described briefly.

4.1.1.1. Results of descriptive statistics of the first sample set

Firstly, Tests of Equality of Group Means showed that there is a statistically significant relationship (Sig.,019) between DESSC (Sentence count, number of sentences) and the scores of the first human evaluator (Hscr1). The standard deviation of the DESSC metric is 50 for those graded 1, 55 for those graded 2 and 77 for those being awarded 3 points. The first evaluator did not use 0 point scores for any of the students' writings. This simply suggests that students whose works were deemed as better by the first human evaluator were longer and had more sentences. It should come as little surprise then, that the impact of DESWC (Word count, number of words) scores on the discrimination had the statistical significance p=,000.

There was also strong relationship (Sig.,013) between DESSL (Sentence length, number of words, mean) scores and the first evaluator's scores. To put simply, texts with a higher mean of sentence length were deemed of better quality. Lastly, DESWLsyd (Word length, number of syllables, standard deviation) score has, according to the analysis, had also statistically significant (Sig., 027) relationship with the scores of the first evaluator.

Interestingly, the classification model created by the Linear Discriminant analysis reported a hit ratio of 78,3% and 65% of these were cross-validated, suggesting that given the nature of samples, it could be relatively easy to predict the scores of the first evaluator purely on descriptive data.

The scores of the second human evaluator (Hscr2) also showed a statistically significant relationship between DESWC (Word count, number of words) and overall scores, but in all other cases, the p-value was not higher than 0.05, accepting the zero theory. These results suggest that the second evaluator also deemed longer texts of better quality, but not in line with other descriptive statistics.

The hit ratio of the prediction model was same as in case of the first human evaluator, 78,3%. However only 50% hit ratio could be cross-validated.

The third evaluator's (Hscr3) scores confirmed results of the first one, having a statistically significant relationship between DESSC, DESWC and DESSL *p*-value smaller than .05 in Wilks' Lambda test. The scores of DESWLsyd were, however, exceeding the p>,05 value as well as in case of the second evaluator.

The hit ratio of the classification model was even higher than in case of first two evaluators, having 83,3% of predicted group membership and 66,7% of cross-validated grouped cases correctly classified, given that these scores are merely descriptive. The model calculated that even after cross-validation, it could predict the group membership of 70% of those graded 1 and 2 based on descriptive results only.

Table 4.1

Descriptive scores first sample set

	Human scores 1 Sig.	Human scores 2 Sig.	Human scores 3 Sig.
DESPC	,275	,832	,649
DESSC	,019	,106	,031
DESWC	,000	,000	,000
DESPL	,147	,384	,395
DESPLd	,234	,184	,214
DESSL	,013	,053	,006
DESSLd	,072	,277	,119
DESWLsy	,414	,822	,464
DESWLsyd	,027	,268	,056
DESWLlt	,151	,303	,099
DESWLltd	,109	,282	,157

4.1.1.2. Results of descriptive statistics of both sample sets

When running the Linear Discriminant Analysis using both sets of samples, the results are naturally expected to be different than in case of the first analysis. The second sample set alone narrows standard deviations in the descriptive statistics, as the second sample set had a lower tolerance for deviations in DESWC for example.

The results of the first evaluator's Tests of Equality of Group means confirmed that DESWC (Word count, number of words) still has the *p*-value at ,000 and that the number of words still aligns with their grading. DESSL (Sentence length, number of words, mean) statistic and DESWLsyd (Word length, number of syllables, standard deviation) still maintain statistical significance, as when testing the first set of samples. The most interesting information comes from the results of Wilks' lambda test concerning DESWLlt (Word length, number of letters, mean) and DESWLltd (Word length, number of letters, standard deviation), which were now attributed higher statistical significance. This result can be easily explained, as the samples of the second set just possess longer words and their mean scores are slightly higher than those of the first set of samples.

The hit ratio of predicting group memberships was increased to 81,8% and 66,2% after cross-validation. This change may be attributed to bigger sample size.

In case of the second evaluator, most of the results remained without significant changes when adding the second set of samples. Only the hit ratio decreased to 74% in the prediction of group classifications.

Similarly, the Wilks' Lambda test's results of the third evaluator's scores were similar, as in the case of the first samples, excluding only DESSC scores, which now exceeded the value $p < 0.05$. The hit ratio decreased by 2% in both classifications of original groups and cross-validation.

Table 4.2

Descriptive scores both sample set

	Human scores 1 Sig.	Human scores 2 Sig.	Human scores 3 Sig.
DESPC	,100	,880	,663
DESSC	,084	,191	,052
DESWC	,000	,000	,000
DESPL	,437	,522	,492
DESPLd	,763	,304	,264
DESSL	,034	,056	,019
DESSLd	,146	,192	,119
DESWLsy	,138	,446	,253
DESWLsyd	,010	,174	,110
DESWLlt	,030	,334	,229
DESWLltd	,010	,226	,273

The results of Descriptive statistics' influence on the overall score was not expected to be particularly high from the beginning, and its results may, in fact, tell us more about evaluators and overall fashion of quality of samples. The most interesting outcome is that in all cases, DESWC count had a significant impact on the final scores. As stated above, this result may simply be explained by evaluators agreeing that longer essays were better. Note, however, that in the matrix itself, the DESWC is destined to have the highest deviation, as the DESWC count is the highest (simply number of words).

4.1.2. Text Easability Principle Component (PC) Scores

This part of the data that Coh-Metrix provides is discussed in detail section 3.3.1.5 and reports the text easability. The scores are attributed a z score and a percentile. As both measurements indicate the same value in different numbers, only z scores will be analysed and reported.

4.1.2.1. Results of test easability PC scores of the first data set, z score

While testing results of the first human evaluator's overall scores (CERMAT I), The Test of Equality of Group means reported only one denial of the null hypothesis in the PCSYNz (Text Easability PC Syntactic simplicity, z score) score by *p-value* being ,037. Further, the Box's M Test Results met the *p-value* at exactly *p*=,050. Moreover, comparing the standardised canonical discriminant function coefficients with the structure matrix proved that most of the values provided by the Coh-Metrix do not match with those of the first evaluator.

The hit ratio in predicting group memberships was measured to 68,3% and 60% upon cross-validation, proving that there was not strong enough relationship between the scores of the first human evaluator and the easability PC scores.

When testing relationship between the first evaluator's scores on cohesion (CERMAT II), which were expected to give better results especially in assigning predicted group memberships, the test of equality of group means failed to confirm that any of the eight variables measured would have a statistical significance in the following prediction model. The algorithm could still predict the Group Membership at 86,7% accuracy and 76,7% upon cross-validation, thus we may conclude that there could be some relationship.

The last analysis ran on vocabulary scores of the first evaluator (CERMAT III), the only Coh-Metrix score breaking the null hypothesis was PCREFz (Text Easability PC Referential cohesion, z score) by *p-value* p=,047. The prediction model successfully assigned 71,7% of the samples and 60% upon cross-validation..

First results of the second human evaluator and relationship of their scores of CERMAT I followed similar fashion, as only relation was found between PCSYNz (Text Easability PC Syntactic simplicity, z score) scores, at *p-value* being p=,032 and the CERMAT I scores. The prediction hit ratio was 71,7% and 53,3% upon the cross-validation.

As well as in case of the first evaluator, the second evaluator's cohesion scores didn't align in all Coh-Metrix Easability PC scores, but the prediction model showed 80% of correct classifications and 73,3% of cross-validated grouped cases being correctly classified.

And finally, the third analysis, of vocabulary scores (CERMAT III) and the Easability PC scores were found non-conclusive as in case of the first evaluator, with neither of the Coh-Metrix scores breaking the null hypothesis. The prediction hit ratio was 60% of original grouped cases correctly classified and 51,7% of them being cross-validated, showing that based upon this matrix, the grouping is very random.

The results of the third evaluator confirmed further that the only statistically significant score of the Coh-Metrix Easability PC z scores in comparison to CERMAT I (Content score) has been PCSYNz, (Text Easability PC Syntactic simplicity, z score). The *p*-value is p=,032.

Furthermore, all other score's affiliation was deemed as insignificant. The prediction of grouping has been again 70% upon original test and 61,7% after the cross-validation.

When analysing the relationships between the third evaluator's scores in cohesion (CERMAT II), the PCDCz (Text Easability PC Word concreteness) score had *p* value of *p*=,021 according to Wilks' Lambda test. Also, this analysis further confirmed the results of previous tests. The prediction hit score was 81,7% of original grouped cases correctly classified and 71,7% of cross-validated grouped cases correctly classified.

The relationship between the third evaluator's scores on vocabulary used (CERMAT III) was also found insignificant in all matrixes and hit scores of prediction values were again very low, 65% were correctly classified, and only 46,7% were cross-validated.

Table 4.3

Easability PC scores first sample set

	Hscr1	Hscr2	Hscr3
CERMAT I	PCSYNz (Sig.,037)	PCSYNz (Sig.,032)	PCSYNz (Sig.,032)
CERMAT II	X	X	PCDCz (Sig.,021)
CERMAT III	PCREFz (Sig.,047)	X	X

4.1.2.2. Results of test easability PC scores of both data sets, z score

When comparing the results of the scores of the first human evaluator (CERMAT I) with the set of Easability PC scores, findings that were observed in the first set of data cannot be confirmed, as no Coh-Metrix scores broke the null hypothesis, including the PCSYNz (Text Easability PC Syntactic simplicity, z score) score. The hit ratio is further decreased to 58,4% and 41,6% upon cross-validation.

When comparing the second set of grades, the ones scoring overall Cohesion of texts (CERMAT II), two Coh-Metrix scores PCCONNz (Text Easability PC Connectivity, z score) and PCTEMPz's (Text Easability PC Temporality, z score) statistical significance *p*-value was set to *p*=,043 and *p*=,025. The hit score was 81,8%, and 77,9% of samples would be correctly assigned in cross-validation.

The third human scores (CERMAT III) of the first evaluator again showed a relation to PCREFz (Text Easability PC Referential cohesion, z score) with *p*-value being *p*=,032.

The results of the second evaluator's scores of content (CERMAT I), and their statistical significance to the Coh-Metrix variables showed no occasion, where the null hypothesis could be denied. The hit ratios were 64,9% upon the first group membership prediction, and 55,8% of them were cross-validated.

Comparing the cohesion scores of the second evaluator with the Coh-Metrix data showed, interestingly, that PCDCz (Text Easability PC Deep cohesion) *p-value* is p=,002 and PCCONNz *p-value* is p=,027. The prediction model still showed only 75,3% hit ratio and 70,1% after cross-validation.

The scores of the second evaluator's vocabulary and content grades (CERMAT II) then showed no statistically significant correlation, and again proved very little prediction value (58,4% of original cases correctly classified, 53,2% of cross-validated cases correctly classified.

Moving on to the scores of the third evaluator, results of two preceding analyses were confirmed. Neither of Coh-Metrix scores had statistically significant value p<,050, in CERMAT I scores and hit scores again moving at 71,4% of original group memberships correctly predicted, and only 61% cross-validated grouped cases correctly classified.

The analysis of the Easability PC scores compared to the scores on cohesion CERMAT II aligned with the results on the same topic with the first group of samples, having *p-value* p=,002 in PCDCz (Text Easability PC Deep cohesion) score. The hit scores for the analysis of cohesion related scores was 75,3% of original cases and 68,8% cross-validated cases correctly assigned.

The third score on vocabulary and grammar again showed no data having a statistically significant relationship of the scores to the discriminant analysis, unsurprisingly having 62,3% hit score and 51,9% after cross-validation.

Table 4.4

Easability PC scores both sample sets

	Hscr1	Hscr2	Hscr3
CERMAT I	X	X	X
CERMAT II	PCCONNz (Sig.,043) PCTEMPz (Sig.,025)	PCDCz (Sig.,002) PCCONNz (Sig.,027)	PCDCz (Sig.,002)
CERMAT III	PCREFz (Sig.,032)	X	X

Upon these results, it may be concluded, that none of the Coh-Metrix z scores on text easability correlates with human evaluators' scores. Given the fact that in the first analysis, there has been found a specific connection between PCDCz (Text Easability PC Word concreteness) and scores attributed by human evaluators on content (CERMAT I) may suggest, that, on this level, there is a connection between human scores and those of the Coh-Metrix. After adding the second set of samples to the pool and running the discriminant analysis, some of the scores reporting cohesion were attributed statistically significant scores. This fact may suggest that the differentiation between cohesion of both samples is substantial enough to play a difference when comparing the machine grading and that of human evaluators. The prediction values were also higher with cohesion scores; it may be thus concluded that there is some measurable connection between the two. Given the nature of the samples and number of evaluators, the scores could improve when a number of samples is increased.

4.1.3. Referential Cohesion and Latent Semantic Analysis

In this part, the human scores of CERMAT II that should report on the cohesion in analysed texts will be compared with scores provided by the Coh-Metrix system's scores on referential cohesion and LSA measures of semantic overlap between sentences or between paragraphs. These results are of particular interest, as human evaluators were asked to grade cohesion holistically (as grading cohesion on only analytic level would be too time-consuming given the scope of the current study) and the machine analysis is purely analytical.

4.1.3.1. Results of referential cohesion and lsa analysis of the first data set

Analysing all scores provided by the Coh-Metrix and comparing them to scores of the first human evaluator's scores on cohesion (CERMAT II), interestingly none of the scores showed statistical significance on scores.

Despite scores lacking statistical significance, the model could still predict group membership, in 90% of cases and 70% after cross-validation, suggesting a certain level of connection between both scores. The main issue may lay in the scores the first evaluator assigned to the essays. As can be seen in table 4.5, the first evaluator assigned mark 2 to a majority of students, and the model decided to move five essays scored 1 and one essay scoring 3 to the group that scored 2. We can see, that the distribution of the essays matches mostly just within the group that scored 2 and except one essay that has achieved 3 points according to the evaluator remained in given group.

Table 4.5

Referential Cohesion and LSA a. prediction model

		Total number of essays per grade	Predicted CERMAT scores 1	2	3
Original CERMAT scores	1	8	3	5	0
	2	48	0	48	0
	3	4	0	1	3
Cross-validated					
Original CERMAT scores	1	8	0	8	0
	2	48	5	41	2
	3	4	0	3	1

a. 90,0% of original grouped cases correctly classified.

b. 70,0% of cross-validated grouped cases correctly classified.

The second evaluator's scores showed very similar results to those of the first evaluator, meaning, that none of the scores proved to play a statistically significant role in relation to human scores or the discrimination process. The prediction model also showed a very little increase in the percentage of hit scores, as 88,3% of original grouped cases were correctly classified, and 65% were correctly classified in cross-validation.

The scores of the third evaluator were the only one, where at least one group of scores played, according to group equality of group means, statistically significant role, *p-value* being

p=,007 with LSAPP1 (LSA overlap, adjacent paragraphs, mean). The prediction model, in this case, reported an 86,7% hit ratio and 68,3% after the cross-validation.

4.1.3.2. Results of referential cohesion and LSA analysis of both data sets

The model for prediction of statistical significance has not found any statistically significant connections of Coh-Metrix scores to the scores of the first evaluator. The hit score of prediction model stays however around the same percentage, 83,1% marks being correctly classified and 74% of being correctly classified after cross-validation.

In the case of the second evaluator's scores, the LSAPP1 (LSA overlap, adjacent paragraphs, mean) metric shows statistical significance (Sig.,006), as it has been the case in the analysis of first sample's prediction values of the third evaluator alone. The hit ratio of the prediction model shows 81,8% of original grouped cases being correctly classified and 64,9% after cross-validation

Surprisingly, in the case of the third evaluator, the statistical significance of the LSAPP1 metric stays at *p-value* p=,006 having precisely the same hit ratio and after cross-validation hit ratio percentages as the scores of the second evaluator.

Table 4.6

Referential Cohesion and LSA results

	Hscr1	Hscr2	Hscr3
First sample	X	X	LSAPP1 (Sig., 007)
Both samples	X	LSAPP1 (Sig., 006)	LSAPP1 (Sig., 006)

These results may be disappointing, as the correlation between human scores on cohesion (CERMAT II) and scores of the Coh-Metrix on cohesion was expected to be higher than displayed. The failure of this test can be attributed to multiple issues from the sample size to grading of each evaluator, or that cohesion does not have to play a significant role in grading, in the way as analysed by .the Coh-Metrix.

Since the nature of this Coh-Metrix allows making a mean of all scores on cohesion, as they are recorded all in the same format, a T-Test was conducted to double-check the result. Mean cohesion score was calculated for each evaluated student, and then the means were compared to the scores of human evaluators. The T-Test verified results of the discriminant analysis, as mean scores and even standard deviations were very similar cross-grades.

4.1.4. Lexical Diversity

It was presumed that lexical diversity scores of the Coh-Metrix system could be correlated with human scores on vocabulary and spelling. The Coh-Metrix system provides metrics on TTR, MTLD as well as on *vocd*. The MTLD and *vocd* should be used together for best results (McCarthy & Jarvis, 2010), and together with TTR should provide whole picture, given that the length of all graded texts is similar (McNamara, Graesser, McCarthy, & Cai, 2014). Given the fact that results from the descriptive part suggested that the number of words is reflected in the scores, the TTR scores could reflect the scores as well in this particular case.

4.1.4.1. Results of lexical diversity scores, first data set

Evaluating the first set of scores, as provided by the first evaluator, this metric could reflect the scores provided by the Coh-Metrix. Firstly, the mean of each score was increasing, and standard deviation of the scores also gained consistency.

The test of Equality of Group Means, therefore showed statistically significant results as LDMTLD (MTLD) scored *p-value* p=,014 and LDVOCD (vocd) had *p-value* p=,002.

The group prediction model has assigned 73,3% of original grouped cases correctly, and 65% were cross-validated.

Table 4.7

Lexical diversity scores means and std. deviation

Human CERMAT scores Vocabulary and Grammar First evaluator		Mean	Std. Deviation
0	LDTTRc	,57	,02
	LDTTRa	,48	,04
	LDMTLD	52,37	37,03
	LDVOCD	56,97	35,11
1	LDTTRc	,63	,05
	LDTTRa	,49	,06
	LDMTLD	54,74	17,16
	LDVOCD	60,61	19,53
2	LDTTRc	,64	,04
	LDTTRa	,49	,04
	LDMTLD	67,20	17,41
	LDVOCD	79,62	18,99
3	LDTTRc	,70	,06
	LDTTRa	,52	,06
	LDMTLD	87,78	3,95
	LDVOCD	98,85	5,19
Total	LDTTRc	,63	,05
	LDTTRa	,49	,05
	LDMTLD	59,29	18,87
	LDVOCD	67,15	21,80

To double check the results, the discriminant analysis was conducted again, leaving out LDDTRc and LDTTRa indexes. Upon the re-test, the prediction model was able to predict 68,3% of grades correctly and more importantly, had the same score with same distribution after cross-validation. This result could be anticipated, as both LDMTLD and LDVOCD scores reflect same data in a different manner. The prediction model assigned the two cases, in which the evaluator graded the level of vocabulary as 0 or 3 to 1 and 2 grades. It should be noted that it assigned the incidences of 0 to 1 and of 3 to 2. 10,3% of essays graded 1 point were moved to group of essays given 2 points and 64,7% of those scoring 2 were be moved to the 1 point group by the model.

Table 4.8
Lexical diversity scores prediction model

		Total number of essays per grade	Predicted CERMAT scores 0	1	2	3
Original CERMAT scores	0	2	0	2	0	0
	1	39	0	35	4	0
	2	17	0	11	6	0
	3	2	0	0	2	0
Cross-validated						
Original CERMAT scores	0	2	0	2	0	0
	1	39	0	35	4	0
	2	17	0	11	6	2
	3	2	0	0	2	0

a. 63.3% of original grouped cases correctly classified.
b. 68.3% of cross-validated grouped cases correctly classified.

The results of the second examiner's scores confirmed the results of those of the first one. The LDMTLD and LDVOCD scores of the Coh-Metrix system were given *p-value*s *p*=,007 and *p*=,001. The prediction model has scored even higher values, being 71,7% for both original prediction and cross-validation. Interestingly, in this case, the prediction model would not assign any of the students' writings mark 1. This effect is mainly caused by mean values and standard deviations being very similar for both groups.

Table 4.9
Lexical diversity scores test of equality of group means

	Wilks' Lambda	Sig.
LDTTRc	,881	,067
LDTTRa	,931	,257
LDMTLD	,808	,007
LDVOCD	,754	,001

Testing the third evaluator's scores correlated highly with those of the first one. Both LDMTLD and LDVOCD had *p-value* *p*<,050. However, the prediction model failed to differentiate between grade 1 and 2, as was the case in the previous grading.

4.1.4.2. Lexical diversity scores, both data sets

The second round of analysis confirmed the results of the first set, considering the first evaluator's scores, as they were nearly identical. LDMTLD's *p-value* *p*=,004 and LDVOCD's *p-value p*=,000. The prediction model classified 70,1% of original groups correctly, and 66,2% of cases were classified correctly in cross-validation.

When were the second evaluator's vocabulary scores (CERMAT III) were compared to those of the Coh-Metrix, the statistical significance of LDMTLD and LDVOCD values were further confirmed, but the prediction model achieved only 58,4% correct classifications, and 55,8% upon cross-validation. The interesting effect of adding the second set of samples into the pool was, however, that the model was able to distribute grades more equally than in the case of the first analysis.

As was the case with the previous examiner, the third examiner's results confirmed the results of all previous runs. LDMTLD had the *p*=,048 and LDVOCD *p*=,017. The prediction model was able to assign 63,6% of original grades correctly, and 55,8% of the grades could be cross-validated.

Table 4.10
Lexical diversity results

	Hscr1	Hscr2	Hscr3
First sample	LDMTLD (Sig., 014) LDVOCD (Sig.,002)	LDMTLD (Sig.,007) LDVOCD (Sig.,001)	LDMTLD (Sig.,019) LDVOCD (Sig.,005)
Both samples	LDMTLD (Sig.,004) LDVOCD (Sig.,000)	LDMTLD (Sig.,015) LDVOCD (Sig.,006)	LDMTLD (Sig.,048) LDVOCD (Sig.,017)

It is clear from the data provided that the LDMTLD and LDVOCD values might probably be used for grading vocabulary. It is uncertain why the scores of the first evaluator were a bit better than those of other two evaluators. However, it can be stated so far that the probability of these two indexes having a considerable connection to the real human scores should be examined further.

4.1.5. Connectives

Coh-Metrix provides an incidence score (measured per 1000 words). Because connectives play an important word as cohesive devices (Halliday & Hassan, 1976), it was decided to test whether a relationship can be found between human scores on cohesion and the Coh-Metrix scores on connectives. The Coh-Metrix system provides incidence for all connectives

(CNCAll), causal (CNCCaus), logical (CNCLogic), adversative/contrastive (CNCADC), temporal (CNCTemp, CNCTempx), additive (CNCAdd), positive connectives (CNCPos) and negative (CNCNeg) (McNamara, Graesser, McCarthy, & Cai, 2014). The last two will be left out of the analysis.

4.1.5.1. Result of connectives

The first run of the analysis done on the set of scores on cohesion (CERMAT II) as provided by the first evaluator did not match any prerequisites for being successful in correlating with scores of the Coh-Metrix. At first sight, all mean scores of grouped scores were overlapping as well as standard deviations. None of the scores were able to pass the Test of Equality of Group Means, and the prediction model was highly randomised. The same was followed in both tests of the second and third evaluator. It was quickly concluded, that no relationship can be found between connectives scores and human scores on cohesion.

The only possible implication of these scores for grading students' works may lay in zero scores of some students, for when students fail to use some connectives, it may be implied that they may not know how to use them properly.

4.1.6. Situation Model

The situation model covers what goes beyond referential cohesion. Standard measures of referential cohesion cannot see what lies behind the lines. When the quality of text gets beyond certain level, cohesion begins to decrease (Halliday & Hassan, 1976). The Coh-Metrix system includes algorithms that try to compensate this and includes them in the section of the situation model.

4.1.6.1. Results of situation model, first data set

When running the first analysis of the first evaluator's scores on cohesion (CERMAT II), statistics suggested that there might be evidence that SMCAUSv (Causal verb incidence) and SMCAUSvp (Causal verbs and causal particles incidence) could play a significant role in following prediction model. Their means and standard deviations were significantly higher with increasing mark awarded by the first evaluator. Despite that, SMINTEr (Ratio of intentional particles to intentional verbs) and SMCAUSlsa (LSA verb overlap) were the only two metrics that had *p-value* $p<,05$. The prediction model could correctly predict group memberships of 81,7% of grades, and confirm them at 78,3% ratio. After having a closer look, however, it was discovered, that the prediction model would move all works to the 2 mark (as it had the most significant number of memberships initially) and just keep 50% of those awarded 3 points in its respective category, together with one from the 2 mark group.

A similar outcome was observed in the case of the second examiner's grades. SMINTEr and SMCAUSlsa scores were attributed statistical significance the prediction model calculated 76,7% of group memberships correctly and 70% of group memberships cross-validate. Interestingly, the prediction model followed a similar path, pushing everyone's score towards mark 2 and leaving and even adding one to the mark 3 group.

The third evaluator's scores further confirmed the fashion. Same scores had high statistical significance, the prediction model functioned on 80% and 68,3% after cross-validation, leaving those attributed mark 3 and moving every other score to 2 mark membership.

4.1.6.2. Results of situation model, both data sets

The analysis of the first evaluator's cohesion scores and Coh-Metrix scores further disproved any affiliations, firstly by removing both SMINTEr and SMCAUSlsa form statistically significant category, after that by having even more randomised prediction model, having roughly same percentages.

In case of the second evaluator, SMINTEr and SMCAUSlsa remained with their *p-value* *p*<,050 statistically significant for the discrimination process, while the prediction model scored slightly lower percentage (74% and 61%), showing tendency to move evaluation towards grade 3.

The results of the third evaluator turned out to be a combination of the two preceding. Moving SMINTEr's impact of grouping out of range of statistical significance, having somehow randomised prediction model with hit scores being 72,7% of original groups being classified correctly, and 58,4% cross-validated.

Table 4.11

The Results Situation model

	Hscr1	Hscr2	Hscr3
First sample	SMINTEr (Sig.,005) SMCAUSlsa (Sig.,019)	SMINTEr (Sig.,001) SMCAUSlsa (Sig.,004)	SMINTEr (Sig.,013) SMCAUSlsa (Sig.,005)
Both samples	X	SMINTEr (Sig.,016) SMCAUSlsa (Sig.,006)	SMCAUSlsa (Sig.,018)

To conclude, it seems that SMINTEr and SMCAUSlsa indexes might have some correlation with human scores. The fashion in which the prediction model followed the tendency to move most grades toward grade 2 and switching some cases between 2 and 3 can be easily explained by exploring group means and standard deviations. It seems that the mean scores of Coh-Metrix in most cases increased rapidly witch grade 3. This result may imply that these metrics could indeed be used in future for grading, at some level of students' writing. A much bigger sample would be needed to confirm this hypothesis.

4.1.7. Syntactic Complexity

Scores on syntactic complexity, as the name implies, measures how complicated syntactic structures are in the measured text. The Coh-Metrix produces mean grades on different types of complexity, and in theory, more complex syntaxes could be graded as 'better' by the evaluators.

4.1.7.1. Results of syntactic complexity, first set of samples

The content scores of the first evaluator suggested when put to discriminant analysis, that there is no statistically significant incidence where scores of the evaluator and scores of the Coh-Metrix meet. Furthermore, the prediction model has shown randomised results, as only 55% of original grouped cases were correctly classified and only 48,3% could be cross-validated.

The similar case followed in the case of the second evaluator's scores. The only difference is that in case of the second evaluator the SYNSTRUTa and SYNSTRUTt scores' *p-value* was $p<,05$.

The third evaluator's scores also suggested that there was certain equality in group means, as in the case of the second evaluator, SYNSTRUTa and SYNSTRUTt scores' *p-value* was $p<,05$, moreover, SYNNP metric showed slight statistical significance (Sig.,036). The prediction model had in the case of the third evaluator highest hit rate, 65% and 55% after cross-validation.

4.1.7.2. Results of syntactic complexity, both sets of samples

On the second run of analysis of these Coh-Metrix values, it was hoped that enriching data set with works of the ELT students could make some difference in outcomes of the discriminant analysis. This was however not the case with the first evaluator's scores, as statistical significance *p-value*s again could not breach the value of p<,050. The prediction model showed roughly the same values as in the first analysis.

The analysis of the second evaluator was more in line with the outcomes of the discriminant analysis of the first evaluator, as none of the variables gained statistical significance, both of the preceding from the first set of data lost it. The prediction model was highly randomised, and no conclusion could be drawn.

The third evaluator's scores were the only one that kept qualities of SYNSTRUTa and SYNSTRUTt scores, adding SYNMEDlem (Minimal Edit Distance, lemmas) and SYNMEDpos (Minimal Edit Distance, part of speech). This was probably reflected in the prediction model that could hit prediction slightly more than in the case of two other examiners.

When closely examining mean scores and standard deviations of all scores within groups, it was found that Syntactic Complexity scores failed to show more than highly randomised outcomes, their possible relationship to human scores were therefore dismissed in this research.

4.1.8. Word Information

The Analysis of word information had to be analysed differently than in cases above. Firstly all incidence indexes have been left out, as they cannot be used for grading themselves, only in the case, where the teacher would be looking for a zero, or near-to-zero score to note whether the student does not try to avoid usage of certain parts of speech.

This part is further split to indexes which contain CELEX (the enormous database of word frequencies developed by the Dutch Centre for Lexical Information) of scores means, and parts of speech that measure the age of acquisition, familiarity, imagability and meaningfulness, for content words scores.

4.1.8.1. Results of word information, first data set

The first evaluator's content scores (CERMAT I) indicated that from the CELEX mean scores, WRDFRQc (CELEX word frequency for content words, mean) could have a relationship with the evaluator's scores. The statistical significance of Tests of Equality of

Group Means assigned it the *p-value* to be *p*=,015. The prediction model reported hit score to be 60% and 55% correctly assigned grouped classes to be cross-validated.

Testing the age of acquisition, familiarity, imagability and meaningfulness, for content words scores' relationship to CERMAT I scores, yielded similar results, prediction model scoring 2% lower scores than in case of CELEX means.

When the values tested were changed for scores on vocabulary (CERMAT III) of the first evaluator, we could observe specific improvement. The scores of the CELEX scores and CERMAT III scores were compared, the test could not prove and statistical significance of tested values. The prediction model could however correctly assign 61,7% of original grouped cases and could repeat the result after cross-validation. The model, however, sorted nearly all writings to the 1 mark group and moved 2 of originally marked 1 essays to mark 3. As Linear Discriminant Analysis is sensitive to outliers, the test was repeated, leaving out all 0 and 3 marks. The prediction model's hit rate increased to 67,9%, it, however, moved all cases again to the mark 1 group, confirming results of the first test.

The relationship between the first evaluator's vocabulary scores (CERMAT III) and scores of Coh-Metrix on the age of acquisition, familiarity, imagability and meaningfulness, for content words scores as well as in previous cases failed to link.

The second evaluator's results of the overall scores (CERMAT I) failed to recognise statistical significance on discrimination of any of the Coh-Metrix's CELEX scores, and the prediction model was highly randomised with hit scores being 56,7% in original grouping cases, and 45% of them being cross-validated.

The same followed when the age of acquisition, familiarity, imagability and meaningfulness, for content words scores' relationship to CERMAT I scores were tested, prediction model being able to assign 61,7% of the grouped cases correctly, and 51,7% after cross-validation.

Testing relationships between the vocabulary scores of the second evaluator (CERMAT III) and CELEX scores of the Coh-Metrix further confirmed results of the first evaluator's scores, repeating the prediction model's results, moving all groups to the group of mark 1, insisting on putting two works marked 1 to mark 3.

The relationship between the second evaluator's vocabulary scores (CERMAT III) and Coh-Metrix age of acquisition, familiarity, imagability and meaningfulness, for content words mean scores further confirmed results of previous evaluator's scores. Prediction model being able to assign 66,7% correctly, and 55% could be correctly classified in cross-validation.

In the case of the third evaluator's scores, the results of the second evaluator, as described above, were confirmed in all measured incidences.

While testing the age of acquisition, familiarity, imagability and meaningfulness, for content words scores' relationship to CERMAT I scores of the third evaluator, the Test of Equality of Group means suggested that there could be relationship between the scores, and WRDAOAc (Age of acquisition for content words, mean) (Sig.,004) and WRDCNCc (Concreteness for content words, mean) (Sig.,049). The prediction model, however, could classify only 60% of original grouped cases and 55% cross-validated were correctly classified.

The test of CELEX scores relationships to third evaluator's scores on vocabulary (CERMAT III) correlated to the results of the other two evaluators.

The third evaluator's score on vocabulary (CERMAT III) also failed to show any statistical significance between the Coh-Metrix scores on the age of acquisition, familiarity, imagability and meaningfulness for content words scores. The hit ratio of the prediction model was 68,3%, and only 50% of cases were correctly assigned in cross-validation.

4.1.8.2. Word information, both sets of samples

When the second set of data was added to the analysed set, the first evaluator's scores again failed to match any of the Coh-Metrix, except for the WRDFRQc scores. The p-value is p=,026 in this test. The prediction model achieved hit scores as low as 49,4% and 46,8% after cross-validation.

After testing the relation between the content scores (CERMAT I) and age of acquisition, familiarity, imagability and meaningfulness, for content words scores' relationship, no statistically significant associations have been found. The prediction model was able to classify 55,8% of original grouped cases, and 44,2% in cross-validation.

Adding the second pool to the analysis of the CELEX scores and switching the first evaluator's human scores for those on vocabulary (CERMAT III) to alter the results from the first set of data significantly. The same followed when the vocabulary scores were compared to scores of Coh-Metrix on the age of acquisition, familiarity, imagability and meaningfulness for content words.

The same case followed with the scores of the second evaluator. None of the Coh-Metrix scores would match grouping values, and the prediction model has scored only 59,7% and 57,1% cross-validated.

The results of the age of acquisition, familiarity, imagability and meaningfulness, for content words scores' on both data sets also followed the results from the first set of data, scoring 57,1% of correct hits, and 45,5% in cross-validation.

When the second part of data was added to the pool and relationships were compared between second evaluator's vocabulary scores (CERMAT III) and CELEX scores of the Coh-Metrix, no significant difference in the first results was observed. The hit score has kept as low as 58,4% of original grouped cases were correctly classified, and 57,1% were correctly cross-validated. The correlation between second evaluator's vocabulary scores and Coh-Metrix scores on the age of acquisition, familiarity, imagability and meaningfulness could not be proven for content words either.

The third evaluator's results also followed the results from the first data set, having a hit score of 54,5% and 49,4% cross-validated.

Testing age of acquisition, familiarity, imagability and meaningfulness, for content words scores' relationship on the CERMAT I scores of the third evaluator indicated, that WRDCNCc (Concreteness for content words, mean) scores had statistically significant relationship (Sig.,049), WRDIMGc (Imagability for content words, mean) scores having *p-value* *p*=,019 and WRDMEAc (Meaningfulness, Colorado norms, content words, mean) having *p*=,018. The prediction model reflected this fact in hit score 66,2% and 59,7% cross-validated hits, which is slightly higher than in previous cases.

The test of vocabulary scores of the third evaluator and their correlation with the CELEX Coh-Metrix scores saw the same results as in previous evaluators' tests, proving little to no relationship between human scores and CELEX scores as provided by the Coh-Metrix.

The third evaluator's scores on vocabulary (CERMAT III) and their relationship to the age of acquisition, familiarity, imagability and meaningfulness could not be proven ether for content words scores, confirming results of previous tests.

Table 4.12

Word information results

	Hscr1	Hscr2	Hscr3
First sample	WRDFRQc (Sig.,015)	X	WRDAOAc (Sig.,004) WRDCNCc (Sig.,049)
Both samples	WRDFRQc (Sig.,026)	X	WRDCNCc (Sig.,049) WRDIMGc (Sig.,019) WRDMEAc (Sig.,018)

After examining Coh-Metrix scores on Word Information and their relationship with human scores on Content (CERMAT I) and vocabulary scores (CERMAT III), no significant relationship could be found. Given the statistically significant results of relationships between the scores of the first evaluator and the third one, some correlation can be found. The scores could be probably used on grading if the setting could be changed. It is also possible that their scores were more influenced by the state of vocabulary than in the case of the second evaluator, who could grade the essays on different bases.

4.1.9. Readability

Readability scores of texts, as provided by the Coh-Metrix system are discussed in greater detail in chapter 3.3.1.5, and their relationship on human scores was expected to be significant. The scores will be tested on all human evaluators' scores.

4.1.9.1. Results of readability scores, first set of data

The first evaluator's scoring of content (CERMAT I) score's relationship to the Readability scores of the Coh-Metrix system was inconclusive. None of the tested Coh-Metrix scores on readability showed a statistically significant approximation with the human scores, When the analysis was re-done with first evaluator's scores on cohesion (CERMAT II), the algorithm calculated that there is a statistically significant relationship between group means of RDFRE (Flesch Reading Ease) (Sig.,026) and RDFKGL (Flesch-Kincaid Grade level) (Sig.,014). The result was however refuted by the prediction model, as, even though it reported 78,3% hit score and confirmed the result upon cross-validation, the grouping was nevertheless randomised and inconclusive, and also by mean scores of each of Coh-Metrix indexes, which were near each other. The test redone using the evaluator's scores on vocabulary (CERMAT

III) has also been inconclusive, as well as in the case of the previous analyses, as no relationship between the Coh-Metrix scores and CERMAT III scores exceeded statistical significance *p-value* $p<,050$. The prediction model showed 60% hit ratio and 58,3% cross-validated cases being correctly assigned.

The second evaluator's scores of content (CERMAT I) analysis results were in line with the results of the same score of the first evaluator. The test of the cohesion scores (CERMAT II) relationship with the Coh-Metrix scores confirmed statistically significant relationship in case of RDFKGL scores (Sig.,017), but the RDFRE score could not achieve statistical significance in this case. The prediction model behaved similarly as in case of the first evaluator's scores on cohesion, reporting 73,3% hit score and 71,7% after cross-validation, however, as in the case of the first evaluator's scores, the prediction model moved most of the grades to one and randomly moved few cases to a higher mark. The test of vocabulary scores was similar to results of the first evaluator's scores; simulation model scoring even lower hit ratio.

The third evaluator's scores on content surprisingly confirmed certain correlation with the Coh-Metrix readability scores in all three cases, in values as follows: RDFRE (Sig.,044), RDFKGL (Sig.,005), and RDL2 (Coh-Metrix L2 Readability) (Sig.,012). The incidence of all three values matching somehow the Coh-Metrix scores was further supported by the test results, where the Box's M value for statistical significance reported *p-value* to be $p=,000$. Despite positive values, the prediction model still had hit ratio of 61,7% and 56,7% after cross-validation. The issue with the scores was discovered when examining the Classification Function Coefficients. The RDFKGL mean scores' means were not increasing with the scores enough in the standard deviation. The values were only 9,6 and 2,2 on 0-100 scales, and 2,2 on a 1-10 scale. The results of cohesion scores relationship to readability scores were mostly inconclusive as well as in case of the first evaluator, the only difference being in the test of equality of group means, where RDFKGL score was assigned (Sig.,033). The third score did not make much difference; only further confirming the results of previous two evaluators.

4.1.9.2. Results of readability scores, both data sets

Adding the second set of data to the tested pool, the test of equality of group means suggested that the group means of RDFKGL (Sig.,026) and RDL2 (Sig.034) are important for the discrimination model. The prediction model's hit ratio was as low as 50,6%. Testing the cohesion scores (CERMAT II), the model reported that only RDFKGL (Sig.,040) would play a role in following prediction model. The prediction model hit scores were low as well. The same case followed when changing the human scores for scores on vocabulary.

The results of the second evaluator were virtually same as in the first case, and the results were the same, apart from not verifying the RDFKGL level of importance on following prediction model.

The third evaluator’s sores virtually copied the results from the first data set and confirmed the importance of RDFRE (Sig.,014), RDFKGL (Sig.,002) and RDL2 (Sig.,016) on following discriminant model when compared to the content scores (CERMAT I). The prediction models in all three cases had unsatisfying results.

Table 4.13

Readability scores results

	Hscr1	Hscr2	Hscr3
First sample	RDFRE (Sig.,026) RDFKGL (Sig.,014)	RDFKGL (Sig.,017)	RDFRE (Sig.,044) RDFKGL (Sig.,005) RDL2 (Sig.,012)
Both samples	RDFKGL (Sig.,026) RDL2 (Sig.034)	X	RDFRE (Sig.,014) RDFKGL (Sig.,002) RDL2 (Sig.,016)

The results of the analyses suggest that there is no substantial relationship between human scores and readability scores of the Coh-Metrix system. However, the scores that were statistically significant for the discriminant model imply that they too could play a significant role in grading students’ essays in a separate set of conditions,

4.2. Benchmarking the Results

From the analysis of all Coh-Metrix data and their relationship to the discriminant model as defined by the Wilks’ Lambda test, the variables that may play role correlating them with the human scores, the results were obtained, as shown in the table below:

Table 4.14

All results of LDA

	Hscr1*	Hscr2*	Hscr3*
First sample	WRDFRQc (Sig.,015) [2] SMINTEr (Sig.,005)[1] SMCAUSlsa (Sig.,019) [1] LDMTLD (Sig., 014) [1a] LDVOCD (Sig.,002) [1a] RDFRE (Sig.,026) [2] RDFKGL (Sig.,014) [1]	SMINTEr (Sig.,001) [1] SMCAUSlsa (Sig.,004) [1] LDMTLD (Sig.,007) [1a] LDVOCD (Sig.,001)[1a] PCDCz (Sig.,002)[2] PCCONNz (Sig.,027) RDFKGL (Sig.,017) [1]	WRDAOAc[2] (Sig.,004) WRDCNCc (Sig.,049) SMINTEr (Sig.,013) [1] SMCAUSlsa (Sig.,005) [1] LDMTLD (Sig.,019) [1a] LDVOCD (Sig.,005) [1a] LSAPP1 (Sig., 007) [2] PCDCz (Sig.,002) [2] RDFRE (Sig.,044) [2] RDFKGL (Sig.,005) [1] RDL2 (Sig.,012) [2]
Both samples	WRDFRQc (Sig.,026) [2] LDMTLD (Sig.,004) [1a] LDVOCD (Sig.,000) [1a] PCCONNz (Sig.,043) PCTEMPz (Sig.,025) PCREFz (Sig.,032) PCSYNz (Sig.,037) [1] PCREFz (Sig.,047) RDFKGL (Sig.,026) [2] RDL2 (Sig.034) [2]	SMINTEr (Sig.,016) SMCAUSlsa (Sig.,006) [2] LDMTLD (Sig.,015) [1a] LDVOCD (Sig.,006) [1a] LSAPP1 (Sig., 006) [2] PCSYNz (Sig.,032) [1]	WRDCNCc (Sig.,049) WRDIMGc (Sig.,019) WRDMEAc[2] (Sig.,018) SMCAUSlsa (Sig.,018)[2] LDMTLD (Sig.,048) [1a] LDVOCD (Sig.,017) [1a] LSAPP1 (Sig., 006) [2] PCSYNz (Sig.,032) [1] PCDCz (Sig.,021) [2] RDFRE (Sig.,014) [2] RDFKGL (Sig.,002) [2] RDL2 (Sig.,016) [2]

* Hscr1 & Hscr3 used analytical approach, Hscr2 used holistic approach

[1a] Incidence of the score in all cases

[1] Incidence of the score in case of one sample set, all evaluators

[2] Incidence of the score in case of one sample set, two evaluators

As we can see in the table above, the LDMTLD (Lexical diversity, MTLD, all words) and LDVOCD (Lexical diversity, VOCD, all words) were the strongest predictors, based on the test, and they appeared statistically significant in all tests across samples and evaluators' scores. The SMINTEr (Ratio of intentional particles to intentional verbs), SMCAUSlsa (LSA verb overlap), RDFKGL (Flesch-Kincaid Grade level), and PCSYN (Text Easability PC Syntactic simplicity) scores also played a statistically significant role in either of samples, as they were recorded in scores of all evaluators.

The strongest predictors were then put to benchmarking tests. First benchmarking test was conducted on all MSKU samples and mean scores of all evaluators were used as the grouping variable. The means and standard deviations of each Coh-Metrix can be seen in the table below.

Table 4.15

Selected discriminator's means and std. deviations

		Mean	Std. Deviation
1,00	LDMTLD	56,57	21,17
	LDVOCD	61,69	24,52
	PCSYNp	69,57	25,23
	SMCAUSlsa	,12	,04
	RDFKGL	5,61	2,44
	SMINTEr	1,07	,94
2,00	LDMTLD	61,55	14,82
	LDVOCD	71,61	18,18
	PCSYNp	63,53	25,97
	SMCAUSlsa	,11	,04
	RDFKGL	6,27	2,34
	SMINTEr	1,26	,93
3,00	LDMTLD	78,44	18,81
	LDVOCD	88,27	20,90
	PCSYNp	52,40	33,31
	SMCAUSlsa	,08	,02
	RDFKGL	8,86	2,04
	SMINTEr	2,17	2,21
Total	LDMTLD	60,77	18,24
	LDVOCD	68,95	21,81
	PCSYNp	65,09	26,18
	SMCAUSlsa	,11	,04
	RDFKGL	6,19	2,46
	SMINTEr	1,25	1,06

The prediction model building on these scores achieved hit score 68,8%, and 61,0% could be cross-validated. The model used LDMTLD and LDVOCD scores as strongest values for discrimination. Therefore it may be concluded that LDMTLD and LDVOCD scores are the strongest discrimination values.

Table 4.16

Prediction model based on all analyses

		Total number of essays per grade	Predicted CERMAT scores 1	2	3
Original CERMAT scores	1	29	11	17	1
	2	43	3	40	0
	3	5	0	3	2
Cross-validated					
Original CERMAT scores	1	29	10	18	1
	2	43	5	37	1
	3	5	0	5	0

a. 68,8% of original grouped cases correctly classified.

b. 61,0% of cross-validated grouped cases correctly classified.

When the 'outliers' (essays awarded either 3 or 0 points, that are rare in the current data) are left out, as the Linear Discriminant Analysis is sensitive to outliers, the hit ratio improves for 4% after the cross-validation.

To verify the data obtained, the metrics of the Coh-Metrix system, which according to the Wilks' Lambda tests had statistically significant impact on discriminant analyses conducted, had been used again on a different set of samples. The second set of samples consists of 17 essays of Padua University's students at B1 and B2 levels. The system was fed the data, and it was supposed to assign the essays to their B1 and B2 groups correctly.

The resulting prediction model was able to classify 88,2% of original grouped cases correctly and had scored 76,5% after the cross-validation.

Table 4.17

Prediction model based on all analyses, auxiliary samples

		Total number of essays per grade	Predicted CEFR scores B1	B2
Original CEFR scores	B1	8	7	1
	B2	9	1	8
Cross-validated				
Original CEFR scores	B1	8	6	2
	B2	9	2	7

a. 88,2% of original grouped cases correctly classified.

b. 76,5% of cross-validated grouped cases correctly classified.

CHAPTER V

DISCUSSION, CONCLUSION AND IMPLICATIONS

5.1. Discussion

5.1.1. The Relationship between Coh-Metrix Scores and Human Scores

To answer the question, whether there is a measurable relationship between any of scores provided by the Coh-Metrix system and possible real-life human scores, a series of analyses was conducted. Firstly, let us remember, where we are coming from, as this topic has been previously debated, for instance McNamara, Crossley, & McCarthy in 2010 hypothesised that such a system as Coh-Metrix could, thanks to the range of metrics it provides, serve as a useful tool for grading written utterances of various texts. They conducted a study with a wide range of high-school level students, the purpose of which was to validate this theory. They have been successful, and they could efficiently predict students' levels of writing, and the program has attributed the ability to provide scores that may serve to predict human grading. They pinpointed three strongest predictors such as Syntactic Complexity scores (SYNLE, SYNNP, SYNMED, SYNMED, SYNSTRUT), Lexical Diversity (MTLD) scores and Word Frequency scores (CELEX).

The theory was put to trial in 2016 by Perin & Lauterbach who tried it on native-speaking adults. They replicated the study on two sets of samples and tried to verify findings of McNamara, Crossley & McCarthy. Their results could not replicate the success of McNamara, Crossley, & McCarthy in 2010, however, they tageted ten scores, as provided by the Coh-Metrix, which had a relationship to the scores of low-proficiency and high-proficiency adults. They found that two scores supplied by the Coh-Metrix had statistically significant differences between the high and low proficiency groups using the first set of samples, Argument overlap (CRFAOa) and Lexical diversity (LDTTRa). For the second set of samples, they were able to target content word overlap (CRFCWO1d), *BVOCD, all words* and familiarity for content words, (WRDFAMc) to have statistically signifficant relationships with human scores. They tested their prediction model, and they achieved a 70% and 68% hit ratio in their two samples. They concluded that their sample of college developmental education students' scores had a relationship with two of the Coh-Metrix variables, that could in prediction model predict human holistic scores on persuasive essays, and three could predict human analytic scores on written summaries. They concluded that there is no overlap between their measures and those found in McNamara, Crossley & McCarthy (2016).

In the current study, the two strongest predictors were the MTLD scores and VOCD Lexical Diversity scores. This fact overlaps with the original research of McNamara et al. (2010) where MTLD has also been found to be one of the most influential predictors. Another relationship similar to the McNamara et al. (2010) study were with the CELEX scores, where some have been found to have a relationship to the scores in the current study. No overlap in the results has been seen with the Perin & Lauterbach (2016) study.

The prediction model could in benchmarks score 60-70% hit rate after cross-validation, which is comparable to the results of the two previous studies.

The incidence, under which the Coh-Metrix system's scores have indicated statistical significance to the human scores, suggests that the system's scores correlate better with analytical human grades than holistic ones.

When we return to the beginning and the first analysis of the descriptive data, we observed a considerably high correlation between the descriptive data and the scores. These results do not tell much about capabilities of the Coh-Metrix and the relationship of its scores to the human scores, however, it served as a good reminder that the extent of evaluated essays correlated well with the human scores. The essays came from various students who applied to a variety of departments, and the number of words in the essays and quality varied considerably depending on what faculty students applied to. Having positive results thus confirmed that the evaluators did a fairly good job in determining students' levels. More extended essays could be attributed to students who applied for courses, where the English language was the aim of the course (i.e. ELT, ELL) and their language level is expected to be higher..

5.1.2. Grading using Coh-Metrix System

The results of the current study cannot suggest using the system in its current version (3.0) for the grading of the FL students. The fact that benchmarking model could achieve around 60% accuracy can be deemed as sufficient for purposes of this study; however, it is not near the percentages that would be required for a successful replacement of human evaluators. The Coh-Metrix could be however used for different purposes. It could be used, for example, by students or teachers, to track their development in writing as analysed by the system.

We must also remind ourselves of the fact mentioned earlier. The Coh-Metrix is a system for linguistic analysis that uses databases created as early as in 1970's and with deep-learning algorithms emerging, we may observe these databases being used by modern AI systems. With emergence of deep learning processes and neural networks, we may see these databases coming 'to life' as the algorithms will be able to learn from human evaluators themselves not just by being programmed. The current approach to machine analysis may very soon seem obsolete and clumsy.

Another way to approach this issue is to build sufficient database that could be used by teachers themselves, who could have, given they would possess a broad enough pool of data, make their prediction models which would reflect the way they grade. This could be an extraordinary contribution to their development, for the teachers could learn much about how they grade. This research showed that there were re-occurring themes in human grades and their relationship to Coh-Metrix values. It has been, for example, observed that all evaluators considered more extended essays as better. The word information scores of the Coh-Metrix correlated only with one evaluator's scores, which may be, hypothetically speaking, explained either by the way that the evaluator pays too much attention to single words, or in turn, that other evaluators did not pay enough attention to this dimension of the text.

5.2. Conclusion and Implications

The following chapter presents the reader with the summary of the study together with some implications for teachers and students who might be interested in machine grading of written texts. The limitations of the study and recommendations for further research are discussed in this chapter as well.

5.2.1. Summary and Conclusion of the Study

This study was concerned with the problem of machine grading and possible implications of currently available free-to-use technology for linguistic analysis for evaluating natural language of FL students. The study was concerned with Coh-Metrix system in particular, and it tried to find relationships between scores of linguistic analysis, as provided by the system, and scores of human evaluators on different dimensions (i.e. vocabulary, cohesion, content). To find answers to these questions, the researcher presented the following research questions:

1) What, if any, relation is there between linguistic features and usage of cohesive devices in final essays as rated by Coh-Metrix and grading provided by human evaluators?

a. Is there a relationship between different parts of human grading (i.e. Cohesion, Vocabulary, Content score) and those of Coh-Metrix relevant to this grading?

2) Can the Coh-Metrix system be used for grading final essays of Turkish FL students?

a. Which parts of Coh-Metrix analysis system can be used for grading and to what extent?

b. Can the system be used for targeting exact level of essays according to CEFR?

To answer the research questions presented, the researcher has collected two samples, 60 randomly selected final essays produced by language preparatory classes' students of the MSKU and 17 essays produced by volunteers from English Language Teaching department, who were first-year students. These essays were then transcribed to an electronic form (as they were, with grammatical and other errors) and were fed to the Coh-Metrix system, which can run linguistic analysis and reported data of students' essays in numbers.

The essays were then in an electronic form distributed to two experienced teachers who were given the task of grading the essays, one holistically and the other analytically in line with CERMAT scale for grading foreign language essays. The essays were attributed scores 0 to 3 for their content, vocabulary and coherence. The two extremes 0 and 3 are attributed to only exceptionally bad or good essays (the evaluator who graded the essays analytically did not grade any essay with 0). The essays were also graded by the examiner for the validity of grades given by the other two.

The data provided by the Coh-Metrix system were categorised and transferred to IBM SPSS v.23 program, where relationships between the data and real human scores were searched for using the Linear Discriminant Analysis, several MANOVAs and ANOVA tests were also conducted when detailed information on individual scores was needed. The method was vastly inspired by the original research of the topic by McNamara et al. (2010) and the analysis methodology used is the same.

The Coh-Metrix scores were analysed by the groups as provided by the system, the scores which had a statistically significant impact in following discriminant model were searched for and recorded. The results of the discriminant models were recorded in percentage, depending on accuracy in which they could predict group memberships, in short, whether they were able to predict the grade of the text correctly using the Coh-Metrix data.

The Coh-Metrix indexes that proved to have statistical significance on discriminant models that followed were then checked for their in-group (group = grade) means and standard deviation. When the means and standard deviation values correlated with the grouping, the prediction model was then examined, as the model would sometimes report considerably high hit-ratio, which would be a result of nature of grading. The analytically graded essays were scored 1 and 2 points the most, very few if any were attributed marks 0 or 3 and the Linear Discriminant Analysis is known to be sensitive to outliers. The prediction model would then change the most of the grades to either 1 or 2 having high hit-ratio by merely narrowing the grades.

The strongest predictors of human grading were identified to be Coh-Metrix scores on Lexical Diversity, namely the MTLD and VOCD scores. These were followed by Ratio of intentional particles to intentional verbs, LSA verb overlap, Flesch-Kincaid Grade level, and Text Easability PC Syntactic simplicity scores. These had re-occurring role to play in the discriminant analysis, mostly with grades given analytically.

The first research question was thus answered. Indeed, there is a measurable relationship between human and algorithmically obtained scores, namely the ones named in the preceding paragraph. The secondary question asking whether there is a relationship between different parts of human grading (i.e. Cohesion, Vocabulary, Content score) and those of Coh-Metrix relevant to this grading was answered as well. The difference was observed several times, and in some cases, for example, MTLD and VOCD scores, the correlation with the scores on vocabulary was stronger than in the case of the score on content. This may come as little surprise; nevertheless, this result may serve as a proof that it should be considered by future researchers assessing this topic, to provide more layers to the human scores. The prediction capabilities of the Coh-Metrix on predicting whether the text is ‘better’ or ‘worse’ was addressed multiple times.

The answer to the second research question, whether the Coh-Metrix system could be used for grading essays of Turkish FL students, the answer to this question should stay negative, as in best case scenario, the results of the prediction model could not exceed 70% success rate. However, we may imply and hypothesise that there may be situations and occasions, where the grading by such algorithmic system could serve as a valuable tool to anyone grading, and it could likely serve as a tool for tracking personal development in writing, although, more research on this topic is needed.

The first of secondary questions here was which values of the Coh-Metrix could be used. As this study could not overlap with results of preceding studies significantly, it should be concluded that none can be used as they are. In any case, the person who would like to use the system for aid in grading or grade themselves should first find scores relevant to their grading and graded texts. If anything, this study can prove that there are connections, but which scores could be used depends on occasion and people involved.

The second secondary question was whether the system could be used for targeting specific levels of essays. As a result of benchmarking, where the system was asked merely to categorise B1 and B2 level essays, and it could do so in about 76% of cases tested. This result may imply that it could be possible. More research will be needed to prove this, but the results of the current study may mean that this could be the case in future.

To sum up, despite having almost all research question answered negatively, it seems, in the shadow of the evolution of machine learning, namely deep learning and neural networks' development that the World is just a step away from having machines aiding teachers with grading writing in everyday life. This method of predicting levels is still sensitive to almost every tiny change and current, despite the fact that, we may soon find ourselves looking at these programs like we do now on computers that occupied whole rooms and warehouses.

5.2.2. Implications of the Study

The study aimed at examining possibilities of currently available linguistic analysis tool called Coh-Metrix, in its current version 3.0. The study tried to find the most relevant scores provided by the system and how they correlate with holistic and analytic approaches in standardised grading.

Firstly, the method the study implemented has been used before for the same goal, and it has been verified times over. Thus, one may conclude that the method may be used by those who would be interested in building their personal scale for grading and evaluation.

The study's outcomes may serve for preparation an automated interface that would serve as an aid for teachers and students. Given the origin of the samples, Turkish freshmen students, the measures provided by the Coh-Metrix could be used for enrichment of corpora that would be needed for grading both FL students and Turkish students as well via such interface. The analysis could be used in for programming the tool and for building a simple program. Firstly the Coh-Metrix output that is in .xls format would have to be transferred to an up-to-date MS Excel format, then, simply using Excel programming capabilities, the score could be calculated from the variety of Coh-Metrix output margins. This simple program could serve for testing purposes and could build corpora on its own if the Coh-Metrix outputs would be collected.

The follow-up to a simple Excel program would be a web interface. There a student would copy paste his or her essay to a text box and upon hitting a button, the text would be automatically given a Coh-Metrix analysis. Upon scoring, the interface could produce pre-programmed commentary on the text submitted by the student. This step would have to be done in cooperation with the Coh-Metrix developers, as the system is protected against bot use by a Captcha system designed to prevent an automated extraction of data from websites.

Other than corpora and program building, this study may serve for purposes of those concerned with the accuracy of standardised grading. The study implies that there are considerable differences between individual evaluators and their grading. Using such a system as Coh-Metrix may be used as a supplement by a number of evaluators who grade English written texts on standardised bases, as the outcome of the Coh-Metrix system is purely analytical, it could be used to determine a grade of a student, when evaluators could not agree upon a mark.

Thanks to its brief description of Coh-Metrix scores and what they mean and imply, this study may also serve anyone wishing to use the system for purposes of their personal research on their writing, and the scores of the samples examined could be used for comparison.

5.2.3. Limitations of the Study and Recommendations for Further Research

This study has several limitations that can have an impact on outcomes of the study. The sample size was limited to N=60 in the case of the first set of samples and N=17 in case of the second one. Increasing number of samples dramatically would very likely have an impact on the outcome of the study. Conducting a similar study on a more significant sample can be recommended for further research.

The samples were used including all their short-comings, miss-types, grammatical errors, syntactic errors etc. This was intentional as the capabilities of the Coh-Metrix interacting with authentic FL texts were one of the points of the current study. A similar study using texts already produced in electronic form, free of fundamental mistakes that are usually corrected by an automatic error-checking add-on, such as Thesaurus or Grammarly, would probably yield more accurate data on metrics such as cohesion or readability level readiness of students. This field of research in Turkish or any other area-specific is thus recommended.

The texts used were in most cases produced by Turkish EFL students. There were texts produced by other nationalities educated in Turkey. Their essays could not be traced due to randomised samples. Extending the area of study and comparing two or more nationalities' level of writing in English, as measured by Coh-Metrix, could probably tell a lot about L1's impact on writing and more.

The scores compared with those of the Coh-Metrix were provided by Czech evaluators using standardised grading scale that is designed for grading foreign language essays of Czech high school graduates. Using different scale could yield different outcomes, as shown when the current study is compared with preceding ones. Research, which would attempt to find grading scale that could correlate more with scores of the Coh-Metrix, could help advancement in the field of automated grading.

Increasing number of evaluators, having more dimensions to scores and more samples could provide a team of researchers with deeper insight on the topic.

The possibility of development of an automated grading system was proposed several times in this study. Extensive research that would follow a similar approach with a higher number of participants, more evaluators, more dimensions to grading implementing neural networks and machine learning could produce a comprehensive program, building on the research of the Coh-Metrix development team, which is in theory able to grade students' essays in future.

REFERENCES

Aşici, M. (2015). Literacy in the World and Turkey: "A general assessment of the situation". *Reading Improvement, 52*(1), 1-8. Retrieved from http://www.projectinnovation.com/reading-improvement.html

Baayen, R. H., Piepenbrock, R., & Gulikers, L. (1995). The CELEX lexical database. PA: Philadelphia: University of Pennsylvania, Linguistic Data Consortium. Retrieved from https://catalog.ldc.upenn.edu/ldc96l14

Bello, T. (1997, March). Writing topics for adult ESL students. Paper presented at the 31st Annual Teachers of English to Speakers of Other Speakers of Other Languages Convention. FL: Orlando.

Brill, E. (1992). Proceedings of the third conference on Applied natural language processing: *A simple rule-based part of speech tagger.* PA: Stroudsburg. doi:10.3115/974499.974526

Brown, G. D. (1984). A frequency count of 190,000 words in theLondon-Lund Corpus of English Conversation. *Behavior Research Methods, Instruments, & Computers, 16*(6), 502-532. doi:10.3758/bf03200836

CERMAT ::. S.r.o, M. (2010). Centrum pro zjišťování výsledků vzdělávání. Retrieved August 17 2017 http://www.cermat.cz/

Chastain, K. (1998). *Developing second - language skills: Theory to practice.* Orlando: Harcourt Brace Jovanovich.

Chomsky, N. (1957). *Syntactic structures.* The Hague, The Netherlands: Mouton.

Chowdhury, G. G. (2005). Natural language processing. *Annual Review of Information Science and Technology, 37*(1), 51-89. doi:10.1002/aris.1440370103

Chris, T. (1997). *Writing: Language teaching : A scheme for teacher education.* Oxford, The United Kingdom: Oxford University Press.

Clark, J. M., & Paivio, A. (2004). Extensions of the Paivio, Yuille, and Madigan (1968) norms. *Behavior Research Methods, Instruments, & Computers, 36*(3), 371-383. doi:10.3758/bf03195584

Coltheart, M. (1981). The MRC psycholinguistic database. *The Quarterly Journal of Experimental Psychology Section A, 33*(4), 497-505. doi:10.1080/14640748108400805

Connor, U. (1990). Linguistic/Rhetorical measures for international persuasive student writing. *Research in the Teaching of English,24*(1), 67-87. Retrieved from http://www.jstor.org/stable/40171446

Crossley, S. A., & McNamara, D. S. (2012). Predictings second language writing proficiency: the roles of cohesion and linguistic sophistication. *Jouran of Research in Reading*, *35*(2) 115-135. doi:10.1111/j.1467-9817.2010.01449.x

Crossley, S. A., Allen, D. B., & McNamara, D. S. (2011, April). Text readability and intuitive simplification: A comparison of readability formulas. *Reading in a Foreign Language, 23*(1), pp. 84-101. Retrieved May 15, 2017, from https://eric.ed.gov/?id=EJ926371.

Çubukçu, F. (2012). *Teaching skills from theory to practice.* Ankara, Turkey: Nobel Academic press.

Deng, L. (2014). Deep learning: methods and applications. *Foundations and Trends® in Signal Processing,7*(3-4), 197-387. doi:10.1561/2000000039

Engber, C. A. (1995). The relationship of lexical proficiency to the quality of ESL compositions. *Journal of Second Language Writing, 4*(2), 139-155. doi:10.1016/1060-3743(95)90004-7

Ferris, D. R. (4. June 1994). Lexical and syntactic features of ESL writing by students at different levels of L2 proficiency. *TESOL Quarterly*, *28*(2), pp. 414-420. doi:10.2307/3587446

Fisher, R. A. (1936). The use of multiple measurements in taxonomic problems. *Annals of Eugenics, 7*(2), 179-188. doi:10.1111/j.1469-1809.1936.tb02137.x

Flesch, R. (1948). A new readability yardstick. *Journal of Applied Psychology, 32*(3), 221-233. doi:10.1037/h0057532

Francis, W. N., & Kučera, J. (1982). *Frequency analysis of english usage: Lexicon and Grammar.* Boston, MA: Houghton-Mifflin.

Gebhard, J. G. (2006). *Teaching English as a foreign or second language, second edition.* Ann Arbor, MI: University of Michigan Press.

Gilhooly, K. J., & Logie, R. H. (1980, July). Age-of-acquisition, imagery, concreteness, familiarity, and ambiguity measures for 1,944 words. *Behavior Research Methods & Instrumentation , 12*(4), pp. 395 - 427. doi:10.3758/BF03201693

Greaser, A. C., McNamara, D. S., Louwerse, M. M., & Cai, Z. (2004, March 15). Coh-Metrix: Analysis of text on cohesion and language. *Behavior Research Methods, Instruments & Computers*, pp. 193-202. doi:10.3758/bf03201693

Haberlandt, K. F., & Graesser, A. C. (1985). Component processes in text comprehension and some of their interactions. *Journal of Experimental Psychology: General, 114*(3), 357-374. doi:10.1037/0096-3445.114.3.357

Halliday, M. A., & Hassan, R. (1976). *Cohesion in English.* London, The United Kingdom: Longman Group Limmited.

Harmer, J. (2004). *How to teach writing.* Harlow, The United Kingdom: Pearson Education Limited.

Hochreiter, S., & Schmidhubner, J. (1997). Long short-term memory. *Neural computation*, *9*(8), 1735–1780. doi:10.1162/neco.1997.9.8.1735

Hutchins, J. (1993, May 5). The first MT patents. *MT News International*, pp. 14-15. Retrieved June 5, 2017, from http://www.hutchinsweb.me.uk/MTNI-5-1993.pdf

Hutchins, J. (1997, February 16). Fifty years of the computer and translation. *MT News International*, pp. 14-15. Retrieved June 5, 2017, from http://www.hutchinsweb.me.uk/MTNI-16-1997.pdf

Jameson, D. A. (2007, September 1). Literacy in decline: Untangling the evidence. *Bussiness Communication Quarterly*, *70*(1), 16-33. doi:10.1177/1080569906297923

Jarvis, S. (2002). Short texts, best-fitting curves and new measures of lexical diversity. *Language Testing,19*(1), 57-84. doi:10.1191/0265532202lt220oa

Kincaid, J. P., Fishburne, J. R., Rogers, R. L., & Chissom, B. S. (1975, February). Derivation of new readability formulas (automated readability index, fog count and flesch reading ease formula) for navy enlisted personnel. *Research Branch Report, 51*(1), 8–75. Millington, Tennessee, USA: U. S. Naval Air Station, Memphis, TN. Retrieved July 1, 2017, from http://www.dtic.mil/dtic/tr/fulltext/u2/a006655.pdf

King, J., Linzen, T., & Marantz, A. (2015, April). *Syntactic categories as lexical features or syntactic heads: An MEG approach.* Retrieved April 12, 2017, from: http://ling.auf.net/lingbuzz/002477/

Kirkgoz, Y , Ünaldi, İ . (2014). Coh-Metrix: Introduction and validation of an online tool for text analysis. *Çukurova Üniversitesi Eğitim Fakültesi Dergisi, 41*(2), 1-17. Retrieved April 21, 2017 from http://dergipark.gov.tr/cuefd/issue/4134/54264

Klare, G. R. (1974). Assessing readability. *Reading Research Quarterly, 10*(1), 62-102. doi:10.2307/747086

Li, J. (2013). The application and significance of discourse cohesion and analysis in practical teaching of foreign language. *Theory and Practice in Language Studies, 3*(8). doi:10.4304/tpls.3.8.1393-1398

Liddy, E. D. (2001). Natural language processing. *Encyclopedia of Library and Information*. New York City, NY. Marcel Decker, Inc.

Marcus, M., Santorini, B., & Marcinkiewicz, M. (1993). *Building a large annotated corpus of English: The penn treebank.* Philadelphia, PA: University of Pennsylvania Scholary Commons.

Mccarthy, P. M., & 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. doi:10.3758/brm.42.2.381

McDonald, R., & Russel, C. (2002). *Teaching writing.* Southern Illionis University Press.

McNamara, D. S., Crossley, S. A., & McCarthy, P. M. (1. January 2010). Linguistic Features of Writing Quality. *Written Communication*, 57-86. doi:10.1177/0741088309351547

McNamara, D. S., Graesser, A. C., McCarthy, P. M., & Cai, Z. (2014). *Automated Evaluation of Text and Discourse with Coh-Metrix.* New York City, NY: Cambridge University Pres.

McNamara, D. S., Crossley, S. A., & McCarthy, P. M. (2010). The linguistic features of quality writing. *Written Communication, 27(1)*, 57–86. doi:10.1177/0741088309351547

McNamara, D.S., Louwerse, M., Cai, Z., & Graesser, A. (2005, January 1). *Coh-Metrix version 1.4*. Retrieved May 4, 2017, Retrieved April 18, 2017, from http://cohmetrix.memphis.edu

Mesnil, G., Dauphin, Y., Yao, K., Bengio, Y., Deng, L., Hakkani-Tur, D., et al. (2015). Using recurrent neural networks for slot filling in spoken language understanding. *IEEE Transactions on Audio, Speech, and Language Processing*, *23*(3) 530-539. doi:10.1109/TASLP.2014.2383614

Ministry of Education, Youth and Sports of the Czech Republic. (10. June 2006). Education Act. *Preschool, Primary, Secondary, Higher Vocational and other education*, § 81 et seq. 11, § 91 et seq. 1 and § 113a et seq. 2 B. 561/2004 Col

Paivio, A., Yuille, J. C., & Madigan, S. A. (1968, January). Concreteness, imagery, and meaningfulness values for 925 nouns. *Journal of Experimental Psychology, 76(1, Pt.2)* 1-25. doi:10.1037/h0025327

Perin, D., & Lauterbach, M. (2016, November 8). Assessing text-based writing of low-skilled college students. *International Journal of Artificial Intelligence in Education*, 1-23. doi:10.1007/s40593-016-0122-z

Pirce, J. R., Carroll, J. B., Hamp, E. P., Hays, D. G., Hockett, C. F., Oettinger, A. G., et al. (1966). Computers in translation and linguistics. *Language and Machines*. Wahington, D.C.: Printing and Publishing Office, National Academy of Sciences. doi:10.17226/9547

Rao, R. C. (1948). The Utilization of Multiple Measurements in Problems of Biological Classification. *Journal of the Royal Statistical Society. Series B (Methodological) 10*(2), 159-203. ISSN 1369-7412

Reid, J. (1992). A computer text analysis of four cohesion devices in english discourse by native and nonnative writers. *Journal of Second Language Writing, 1*(2), 79-107. doi:10.1016/1060-3743(92)90010-m

Richards, J. C., & Schmidt, R. W. (2010). *Longman dictionary of language teaching and applied linguistics.* Harlow, The United Kingdom: Pearson Education Limited.

Russ, T. L. (2016, November 23). The status of the business communication course at U.S. colleges and universities. *Business and Professional Communication Quarterly*, *72*(4) 395-413. doi:10.1177/1080569909349524

Shen, Y., He, X., Gao, J., Deng, L., & Mesnil, G. (2014). *A latent semantic model with convolutional-pooling structure for information retrieval.* New York City, NY: Microsoft Research & University of Montréal ACM

Socher, R., & Manning, C. (2013). *Deep learning for NLP*. Atlanta, GA: Stanford University.

Socher, R., Bauer, J., Manning, C., & Ng, A. (2013, August 4-9). Parsing with compositional vector grammars. *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics*, 455–465. Retrieved April 1, 2017 http://www.aclweb.org/anthology/P13-1045

Socher, R., Perelygin, A., Wu, J. Y., Chuang, J., Manning, C. D., Ng, A. Y., at al. (2013). *Recursive deep models for semantic compositionality over a sentiment treebank*. Stanford, GA: Stanford University.

Templin, M. C. (1957). *Certain language skills in children: their development and interrelationships.* Minneapolis, MN: University of Minnesota Press.

Terk, N. (2014). *Professional writing skills: five simple steps to write anything to anyone.* Oakland, CA: Advanced Communication Designs, Inc.

Thorndike, E. K., & Lorge, I. (1944). *Teacher's word book of 30,000 words.* New York City, NY: Berau of Publications, Teachers College, Columbia University.

Toglia, M. P., & Battig, W. F. (1978). *Handbook of semantic word norms.* Oxford, The United Kingdom: Hillsdale, N.J. : Erlbaum [u.a.].

Ur, P. (1996). *A course in language teaching.* Cambridge, The united Kingdom: Cambridge University Press.

Witte, S. P., & Faigley, L. (1981, May). Coherence, cohesion, and writing quality. *College Composition and Communication*, *32*(2), 189-204. doi:10.2307/356693

APPENDICES

Appendix 1. Coh-Metrix scores, order

Number of the sample (1-60;A1-A17;B1.1-B1.8;B2.1-B2.9)
DESPC 'Paragraph count, number of paragraphs'
DESSC 'Sentence count, number of sentences'
DESWC 'Word count, number of words'
DESPL 'Paragraph length, number of sentences in a paragraph, mean'
DESPLd 'Paragraph length, number of sentences in a pragraph, standard deviation'
DESSL 'Sentence length, number of words, mean'
DESSLd 'Sentence length, number of words, standard deviation'
DESWLsy 'Word length, number of syllables, mean'
DESWLsyd 'Word length, number of syllables, standard deviation'
DESWLlt 'Word length, number of letters, mean'
DESWLltd 'Word length, number of letters, standard deviation'
PCNARz 'Text Easability PC Narrativity, z score'
PCNARp 'Text Easability PC Narrativity, percentile'
PCSYNz 'Text Easability PC Syntactic simplicity, z score'
PCSYNp 'Text Easability PC Syntactic simplicity, percentile'
PCCNCz 'Text Easability PC Word concreteness, z score'
PCCNCp 'Text Easability PC Word concreteness, percentile'
PCREFz 'Text Easability PC Referential cohesion, z score'
PCREFp 'Text Easability PC Referential cohesion, percentile'
PCDCz 'Text Easability PC Deep cohesion, z score'
PCDCp 'Text Easability PC Deep cohesion, percentile'
PCVERBz 'Text Easability PC Verb cohesion, z score'
PCVERBp 'Text Easability PC Verb cohesion, percentile'
PCCONNz 'Text Easability PC Connectivity, z score'
PCCONNp 'Text Easability PC Connectivity, percentile'
PCTEMPz 'Text Easability PC Temporality, z score'
PCTEMPp 'Text Easability PC Temporality, percentile'
CRFNO1 'Noun overlap, adjacent sentences, binary, mean'
CRFAO1 'Argument overlap, adjacent sentences, binary, mean'
CRFSO1 'Stem overlap, adjacent sentences, binary, mean'
CRFNOa 'Noun overlap, all sentences, binary, mean'
CRFAOa 'Argument overlap, all sentences, binary, mean'
CRFSOa 'Stem overlap, all sentences, binary, mean'
CRFCWO1 'Content word overlap, adjacent sentences, proportional, mean'

CRFCWO1d 'Content word overlap, adjacent sentences, proportional, standard deviation'
CRFCWOa 'Content word overlap, all sentences, proportional, mean'
CRFCWOad 'Content word overlap, all sentences, proportional, standard deviation'
LSASS1 'LSA overlap, adjacent sentences, mean'
LSASS1d 'LSA overlap, adjacent sentences, standard deviation'
LSASSp 'LSA overlap, all sentences in paragraph, mean'
LSASSpd 'LSA overlap, all sentences in paragraph, standard deviation'
LSAPP1 'LSA overlap, adjacent paragraphs, mean'
LSAPP1d 'LSA overlap, adjacent paragraphs, standard deviation'
LSAGN 'LSA given/new, sentences, mean'
LSAGNd 'LSA given/new, sentences, standard deviation'
LDTTRc 'Lexical diversity, type-token ratio, content word lemmas'
LDTTRa 'Lexical diversity, type-token ratio, all words'
LDMTLD 'Lexical diversity, MTLD, all words'
LDVOCD 'Lexical diversity, VOCD, all words'
CNCAll 'All connectives incidence'
CNCCaus 'Causal connectives incidence'
CNCLogic 'Logical connectives incidence'
CNCADC 'Adversative and contrastive connectives incidence'
CNCTemp 'Temporal connectives incidence'
CNCTempx 'Expanded temporal connectives incidence'
CNCAdd 'Additive connectives incidence'
CNCPos 'Positive connectives incidence'
CNCNeg 'Negative connectives incidence'
SMCAUSv 'Causal verb incidence'
SMCAUSvp 'Causal verbs and causal particles incidence'
SMINTEp 'Intentional verbs incidence'
SMCAUSr 'Ratio of casual particles to causal verbs'
SMINTEr 'Ratio of intentional particles to intentional verbs'
SMCAUSlsa 'LSA verb overlap'
SMCAUSwn 'WordNet verb overlap'
SMTEMP 'Temporal cohesion, tense and aspect repetition, mean'
SYNLE 'Left embeddedness, words before main verb, mean'
SYNNP 'Number of modifiers per noun phrase, mean'
SYNMEDpos 'Minimal Edit Distance, part of speech'
SYNMEDwrd 'Minimal Edit Distance, all words'
SYNMEDlem 'Minimal Edit Distance, lemmas'
SYNSTRUTa 'Sentence syntax similarity, adjacent sentences, mean'
SYNSTRUTt 'Sentence syntax similarity, all combinations, across paragraphs, mean'
DRNP 'Noun phrase density, incidence'
DRVP 'Verb phrase density, incidence'
DRAP 'Adverbial phrase density, incidence'
DRPP 'Preposition phrase density, incidence'

DRPVAL 'Agentless passive voice density, incidence'
DRNEG 'Negation density, incidence'
DRGERUND 'Gerund density, incidence'
DRINF 'Infinitive density, incidence'
WRDNOUN 'Noun incidence'
WRDVERB 'Verb incidence'
WRDADJ 'Adjective incidence'
WRDADV 'Adverb incidence'
WRDPRO 'Pronoun incidence'
WRDPRP1s 'First person singular pronoun incidence'
WRDPRP1p 'First person plural pronoun incidence'
WRDPRP2 'Second person pronoun incidence'
WRDPRP3s 'Third person singular pronoun incidence'
WRDPRP3p 'Third person plural pronoun incidence'
WRDFRQc 'CELEX word frequency for content words, mean'
WRDFRQa 'CELEX Log frequency for all words, mean'
WRDFRQmc 'CELEX Log minimum frequency for content words, mean'
WRDAOAc 'Age of acquisition for content words, mean'
WRDFAMc 'Familiarity for content words, mean'
WRDCNCc 'Concreteness for content words, mean'
WRDIMGc 'Imagability for content words, mean'
WRDMEAc 'Meaningfulness, Colorado norms, content words, mean'
WRDPOLc 'Polysemy for content words, mean'
WRDHYPn 'Hypernymy for nouns, mean'
WRDHYPv 'Hypernymy for verbs, mean'
WRDHYPnv 'Hypernymy for nouns and verbs, mean'
RDFRE 'Flesch Reading Ease'
RDFKGL 'Flesch-Kincaid Grade level'
RDL2 'Coh-Metrix L2 Readability'

Appendix 2. Scores 1-10

1	2	3	4	5	6	7	8	9	10
5	4	4	5	5	4	5	8	5	6
26	14	15	27	29	11	27	27	22	36
246	207	206	243	330	203	227	282	167	303
5,2	3,5	3,75	5,4	5,8	2,75	5,4	3,375	4,4	6
3,271	1,291	0,957	5,128	3,633	2,872	5,505	1,506	2,608	3,286
9,462	14,786	13,733	9	11,379	18,455	8,407	10,444	7,591	8,417
5,171	9,423	7,116	3,99	4,499	10,386	3,285	5,827	4,067	4,031
1,309	1,372	1,374	1,498	1,388	1,468	1,251	1,532	1,425	1,317
0,678	0,593	0,594	0,84	0,672	0,779	0,605	0,84	0,74	0,665
3,947	4,058	4,083	4,65	4,306	4,3	3,599	4,603	4,341	3,855
2,072	1,957	1,97	2,434	2,207	2,237	2,003	2,455	2,364	2,255
1,727	0,365	0,27	-0,017	0,739	0,438	2,189	0,708	0,798	2,178
95,73	64,06	60,64	49,6	76,73	66,64	98,54	75,8	78,52	98,5
0,183	-0,198	-0,015	0,862	0,763	0,09	0,282	0,813	0,604	0,77
57,14	42,47	49,6	80,51	77,64	53,19	61,03	79,1	72,57	77,94
-0,739	-1,485	-1,61	-0,537	-1,112	-1,245	-0,723	-0,891	-1,185	-0,348
23,27	6,94	5,37	29,81	13,35	10,75	23,58	18,67	11,9	36,69
0,775	-0,453	-0,235	-0,46	-0,611	-0,328	0,969	0,789	-0,047	0,395
77,94	32,64	40,9	32,28	27,9	37,45	83,15	78,23	48,4	65,17
-0,151	-0,087	-0,144	-0,506	0,615	-0,8	0,798	0,264	1,413	0,444
44,04	46,81	44,43	30,85	72,91	21,19	78,52	60,26	92,07	67
1,158	1,079	1,587	0,928	0,33	-0,777	0,628	-0,323	1,746	-0,146
87,49	85,77	94,29	82,12	62,55	22,6	73,24	37,45	95,91	44,43
-1,912	-4,192	-4,313	-4,893	-3,46	-4,997	0,754	-2,911	-2,645	-1,748
2,81	0	0	0	0,03	0	77,34	0,18	0,41	4,9
0,517	0,765	0,504	-0,118	0,253	-0,421	-0,251	-0,265	-0,036	-0,076
69,5	77,64	69,15	45,62	59,87	33,72	40,13	39,74	48,8	47,21
0,32	0,231	0,286	0,154	0,214	0,2	0,077	0,423	0,286	0,143
0,56	0,462	0,5	0,462	0,464	0,5	0,538	0,538	0,381	0,543
0,32	0,462	0,429	0,231	0,214	0,3	0,115	0,423	0,333	0,2
0,102	0,282	0,326	0,205	0,221	0,291	0,098	0,395	0,176	0,131
0,527	0,447	0,442	0,353	0,379	0,527	0,465	0,53	0,273	0,393
0,122	0,424	0,389	0,265	0,238	0,345	0,121	0,409	0,194	0,151
0,138	0,071	0,099	0,11	0,111	0,117	0,199	0,168	0,145	0,141
0,147	0,115	0,144	0,164	0,128	0,13	0,209	0,184	0,218	0,146
0,134	0,087	0,098	0,081	0,081	0,099	0,14	0,146	0,092	0,11
0,15	0,112	0,13	0,123	0,112	0,109	0,183	0,158	0,148	0,156
0,183	0,189	0,18	0,104	0,124	0,146	0,208	0,176	0,176	0,206
0,173	0,206	0,196	0,154	0,198	0,184	0,17	0,243	0,236	0,187
0,152	0,19	0,205	0,08	0,082	0,172	0,117	0,183	0,108	0,179
0,162	0,18	0,204	0,136	0,166	0,218	0,151	0,242	0,138	0,187

0,408	0,411	0,414	0,446	0,347	0,204	0,53	0,555	0,276	0,481
0,103	0,145	0,131	0,101	0,14	0,13	0,085	0,113	0,166	0,203
0,334	0,302	0,299	0,284	0,292	0,3	0,356	0,384	0,296	0,363
0,136	0,168	0,184	0,152	0,166	0,184	0,17	0,198	0,155	0,141
0,585	0,598	0,593	0,674	0,64	0,764	0,651	0,599	0,622	0,62
0,451	0,517	0,51	0,473	0,494	0,576	0,454	0,418	0,515	0,465
56,994	78,566	71,251	56,477	92,903	80,67	44,36	43,908	54,036	63,404
62,233	81,801	74,402	63,328	110,722	95,386	54,786	53,507	59,097	61,954
73,171	91,787	92,233	98,765	93,939	88,67	79,295	85,106	101,796	72,607
28,455	14,493	14,563	12,346	36,364	14,778	22,026	28,369	29,94	23,102
36,585	57,971	58,252	49,383	42,424	44,335	48,458	49,645	65,868	59,406
24,39	28,986	29,126	32,922	24,242	49,261	8,811	24,823	29,94	26,403
8,13	9,662	9,709	4,115	0	9,852	39,648	10,638	17,964	16,502
0	4,831	4,854	12,346	30,303	24,631	13,216	10,638	23,952	13,201
44,715	67,633	67,961	78,189	57,576	68,966	17,621	60,284	47,904	39,604
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
32,52	14,493	19,417	24,691	36,364	19,704	26,432	28,369	11,976	46,205
52,846	19,324	24,272	28,807	51,515	19,704	39,648	42,553	29,94	52,805
36,585	9,662	9,709	24,691	21,212	29,557	35,242	28,369	17,964	33,003
0,556	0,25	0,2	0,143	0,385	0	0,429	0,444	1	0,133
0,5	0,667	0,667	0,429	1,375	0,429	0,333	0,667	1,25	0,455
0,131	0,187	0,225	0,146	0,111	0,091	0,119	0,096	0,203	0,093
0,5	0,342	0,413	0,342	0,396	0,436	0,424	0,346	0,448	0,397
0,9	0,923	0,893	0,846	0,875	0,8	0,827	0,827	0,81	0,843
2,808	5,5	4,733	2,667	3,448	3,545	2,667	2,556	2,455	2,722
0,547	0,574	0,64	0,635	0,532	0,542	0,494	0,457	0,603	0,52
0,61	0,736	0,726	0,652	0,712	0,661	0,595	0,702	0,648	0,704
0,876	0,922	0,914	0,904	0,926	0,906	0,838	0,886	0,874	0,916
0,86	0,825	0,853	0,856	0,913	0,888	0,805	0,876	0,868	0,888
0,184	0,086	0,119	0,169	0,133	0,085	0,248	0,147	0,197	0,157
0,166	0,08	0,097	0,142	0,11	0,07	0,17	0,118	0,171	0,16
418,699	314,01	291,262	399,177	381,818	389,163	414,097	404,255	383,234	343,234
239,837	275,362	262,136	242,798	248,485	236,453	229,075	258,865	233,533	247,525
28,455	33,816	33,981	16,461	24,242	39,409	30,837	17,73	29,94	46,205
77,236	86,957	87,379	65,844	103,03	103,448	92,511	120,567	65,868	59,406
0	4,831	0	0	9,091	0	4,405	3,546	0	3,3
0	4,831	4,854	4,115	15,152	4,926	0	0	0	6,601
12,195	28,986	29,126	4,115	15,152	9,852	0	0	5,988	23,102
20,325	38,647	38,835	8,23	12,121	9,852	4,405	35,461	0	19,802
186,991	207,73	199,029	312,757	254,546	246,305	176,211	283,688	239,521	171,618
77,235	101,45	97,087	106,995	136,363	113,3	110,132	117,021	101,796	132,014
69,106	115,942	121,359	74,073	66,667	64,039	79,295	42,553	119,76	89,109
36,585	67,633	67,961	24,691	57,576	54,187	52,863	28,369	47,904	69,307

207,317	82,126	77,67	115,226	109,091	93,596	237,886	120,567	161,677	234,323
117,886	9,662	9,709	4,115	9,091	0	140,969	0	65,868	132,013
4,065	0	0	8,23	0	0	52,863	28,369	47,904	23,102
0	14,493	14,563	74,074	36,364	64,039	4,405	78,014	0	3,3
4,065	0	0	0	3,3	4,926	39,648	0	29,94	62,706
81,301	43,478	43,689	20,576	39,394	4,926	0	10,638	17,964	13,201
2,806	2,609	2,652	2,52	2,677	2,33	2,807	2,476	2,725	2,751
3,215	3,112	3,121	3,159	3,149	3,038	3,206	3,132	3,172	3,136
1,925	1,418	1,303	1,541	1,504	1,11	2,135	1,406	1,833	1,939
294,043	369,773	369,773	371,714	353,074	354	306	315,867	312,5	269,25
594,063	588,181	588,372	582,4	588,964	568,863	590,519	584,286	593,149	593,248
343,483	339	336,025	342,667	346,238	324,5	377,093	338,786	356,55	364,297
398,094	373,651	370,907	393,438	379,784	367,588	426,519	389,079	398,119	419,232
437,291	412,429	408,985	434,719	422,716	393,302	429,149	419,18	425,269	446,319
3,971	4,113	4,388	3,491	4,4	3,143	3,634	3,101	3,887	4,054
6,229	6,624	6,557	5,366	5,385	4,907	6,26	4,916	4,095	6,395
1,524	1,262	1,316	1,662	1,476	1,824	1,41	1,597	1,426	1,38
1,316	1,474	1,434	1,789	1,498	1,48	1,225	1,501	1,124	1,231
86,49	75,756	76,656	70,969	77,861	63,91	92,467	66,627	78,575	86,874
3,546	6,366	5,979	5,596	5,226	8,93	2,451	6,561	4,185	3,233
35,757	21,889	26,314	27,021	28,367	18,021	42,889	27,724	35,126	33,053

Appendix 3. Scores 11-20

11	12	13	14	15	16	17	18	19	20
6	5	5	5	6	7	5	4	5	4
24	18	33	34	26	30	15	12	26	25
225	398	317	309	268	352	131	172	243	192
4	3,6	6,6	6,8	4,333	4,286	3	3	5,2	6,25
2,828	1,673	5,177	5,357	3,445	2,138	1,871	2,16	3,633	5,377
9,375	22,111	9,606	9,088	10,308	11,733	8,733	14,333	9,346	7,68
4,799	12,704	4,943	5,444	5,136	5,431	4,22	5,929	5,036	3,889
1,316	1,382	1,42	1,243	1,351	1,449	1,344	1,419	1,362	1,401
0,607	0,803	0,687	0,531	0,651	0,734	0,552	0,764	0,756	0,793
3,938	4,324	4,495	3,741	3,843	4,463	3,893	4,256	4,263	4,198
2,172	2,315	2,101	1,908	2,194	2,308	1,746	2,238	2,27	2,595
2,296	1,083	0,795	1,971	1,352	0,009	1,137	0,443	1,19	1,568
98,9	85,99	78,52	97,56	91,15	50	87,08	67	88,1	94,06
0,452	-1,593	1,625	0,377	1,329	0,572	1,058	0,115	0,184	0,866
67,36	5,59	94,74	64,43	90,66	71,57	85,31	54,38	57,14	80,51
0,527	-0,207	0,809	0,047	-1,665	-0,774	-0,044	-1,063	-0,829	0,616
69,85	42,07	78,81	51,6	4,85	22,6	48,4	14,46	20,33	72,91
2,059	-0,053	-0,173	0,16	-0,494	-0,483	-1,331	0,083	-0,372	0,603
97,98	48,01	43,25	55,96	31,21	31,56	9,18	53,19	35,57	72,57
1,422	2,564	1,544	2,144	3,537	2,069	0,753	3,267	-0,128	0,462
92,22	99,48	93,82	98,38	100	98,03	77,34	99,94	45,22	67,72
0,089	1,091	-0,548	-0,12	-0,944	0,609	-0,701	1,155	0,356	0,553
53,19	86,21	29,46	45,22	17,36	72,57	24,2	87,49	63,68	70,88
-3,536	-2,728	-4,519	-2,595	-5,098	-2,001	-0,868	-1,221	-1,825	-5,505
0	0,33	0	0,48	0	2,28	19,49	11,12	3,44	0
1,128	0,803	-0,26	1,658	0,479	-1,345	-1,348	1,673	0,943	0,299
86,86	78,81	39,74	95,05	68,08	9,1	9,1	95,25	82,64	61,41
0,304	0,412	0,313	0,091	0,08	0,31	0,071	0,364	0,16	0,167
0,652	0,529	0,438	0,515	0,44	0,379	0,214	0,455	0,4	0,667
0,478	0,412	0,344	0,121	0,2	0,448	0,071	0,545	0,2	0,167
0,195	0,288	0,215	0,088	0,063	0,302	0,126	0,477	0,098	0,072
0,605	0,544	0,367	0,495	0,288	0,404	0,221	0,523	0,317	0,431
0,351	0,328	0,262	0,123	0,141	0,384	0,126	0,538	0,107	0,097
0,22	0,112	0,102	0,133	0,125	0,091	0,054	0,129	0,143	0,259
0,173	0,109	0,133	0,16	0,141	0,109	0,095	0,121	0,164	0,215
0,195	0,108	0,087	0,129	0,085	0,09	0,059	0,129	0,088	0,128
0,198	0,105	0,131	0,152	0,128	0,115	0,107	0,121	0,137	0,152
0,317	0,161	0,118	0,143	0,141	0,149	0,126	0,183	0,167	0,143
0,176	0,18	0,142	0,187	0,134	0,216	0,184	0,216	0,2	0,144
0,293	0,183	0,087	0,151	0,128	0,178	0,108	0,261	0,135	0,09
0,22	0,151	0,153	0,143	0,147	0,226	0,153	0,311	0,169	0,112

0,558	0,335	0,335	0,465	0,272	0,451	0,295	0,4	0,292	0,351
0,134	0,092	0,089	0,27	0,179	0,205	0,317	0,058	0,196	0,25
0,439	0,317	0,301	0,334	0,299	0,338	0,269	0,325	0,315	0,291
0,15	0,131	0,179	0,136	0,128	0,181	0,138	0,216	0,16	0,12
0,604	0,66	0,642	0,637	0,669	0,633	0,8	0,679	0,669	0,634
0,378	0,482	0,451	0,429	0,507	0,462	0,685	0,581	0,508	0,544
31,947	90,576	58,182	59,761	79,089	84,643	96,746	60,671	57,437	50,422
34,378	102,53	73,353	61,582	88,83	94,91	106,599	83,008	72,59	68,137
88,889	110,553	107,256	103,56	141,791	105,114	61,069	122,093	78,189	119,792
53,333	35,176	25,237	51,78	63,433	45,455	30,534	63,953	16,461	15,625
62,222	47,739	69,401	74,434	100,746	51,136	53,435	75,581	37,037	62,5
48,889	17,588	31,546	22,654	41,045	19,886	15,267	5,814	12,346	20,833
13,333	22,613	25,237	6,472	11,194	19,886	0	11,628	24,691	15,625
4,444	10,5	6,309	16,181	33,582	17,045	22,901	34,884	4,115	20,833
53,333	60,302	72,555	55,016	74,627	48,295	22,901	40,698	49,383	93,75
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
48,889	12,563	34,7	35,599	26,119	17,045	38,168	23,256	28,807	20,833
80	37,688	53,628	58,252	63,433	45,455	61,069	40,698	32,922	26,042
26,667	2,513	47,319	19,417	41,045	19,886	22,901	17,442	8,23	31,25
0,583	1,667	0,5	0,583	1,25	1,429	0,5	0,6	0,125	0,2
0,857	5,5	0,313	1,714	1	1,375	0,75	2,75	1,333	0,286
0,141	0,078	0,092	0,067	0,11	0,111	0,077	0,118	0,061	0,112
0,427	0,493	0,409	0,445	0,331	0,525	0,396	0,465	0,415	0,406
0,957	0,882	0,828	0,97	0,84	0,69	0,714	1	0,92	0,854
2,958	6,722	2,273	2,971	2,731	3,2	2,333	4	3,115	2,56
0,614	0,664	0,495	0,59	0,425	0,681	0,512	0,851	0,566	0,644
0,617	0,675	0,694	0,728	0,694	0,704	0,763	0,647	0,691	0,67
0,823	0,902	0,901	0,942	0,919	0,919	1	0,914	0,921	0,832
0,81	0,896	0,877	0,913	0,889	0,877	0,974	0,892	0,885	0,782
0,161	0,064	0,159	0,171	0,125	0,125	0,11	0,076	0,187	0,266
0,15	0,073	0,154	0,168	0,111	0,112	0,107	0,077	0,171	0,238
355,556	417,085	397,476	375,405	373,134	400,568	335,878	343,023	407,407	359,375
191,111	208,543	252,366	223,301	253,731	221,591	259,542	186,047	213,992	203,125
44,444	30,151	34,7	45,307	63,433	8,523	38,168	46,512	37,037	62,5
137,778	113,065	69,401	84,142	67,164	110,795	68,702	93,023	90,535	72,917
0	0	3,155	0	0	8,523	0	5,814	0	0
0	5,025	3,155	0	14,925	2,841	7,634	5,814	8,23	0
0	25,126	3,155	3,236	3,731	11,364	53,435	5,814	4,115	20,833
22,222	22,613	6,309	25,89	11,194	8,523	0	11,628	24,691	10,417
182,222	213,567	252,367	158,577	223,881	312,501	221,374	255,814	209,877	208,334
102,222	115,579	135,648	119,741	74,626	119,319	152,672	110,466	94,649	145,833
53,333	103,016	53,628	103,56	67,164	62,5	68,702	110,465	106,996	125
62,222	60,302	50,473	61,488	100,746	25,568	76,336	81,395	53,498	83,333

257,778	140,704	135,647	223,301	134,328	82,386	152,672	69,767	176,955	187,5
182,222	35,176	6,309	87,379	70,896	0	0	17,442	74,074	78,125
17,778	15,075	3,155	6,472	11,194	8,523	45,802	5,814	0	0
0	15,075	47,319	12,945	11,194	31,25	61,069	17,442	0	5,208
48,889	10,5	6,309	93,851	0	0	0	0	16,461	98,958
8,889	57,789	63,091	12,945	33,582	19,886	30,534	11,628	78,189	10,417
2,702	2,602	2,415	2,668	2,562	2,44	2,508	2,739	2,584	2,704
3,166	3,139	3,036	3,132	3,032	3,111	2,997	3,142	3,128	3,153
2,186	1,164	1,593	1,669	1,218	1,513	1,84	1,98	1,635	1,828
305,222	345,912	298,5	304,034	351,333	342,375	317,444	337,333	325,125	291,15
593,598	590,779	576,777	591,033	585,887	583,189	588,296	586,242	587,854	589,476
384,195	345,052	388,478	366,207	337,2	342,068	401,094	343,317	349,313	380,24
439,817	392,123	416,165	421,458	375,457	382,811	440,037	392,409	402,517	426,451
472,054	431,173	446,179	450,38	393,464	422,797	447,128	425,571	427,792	437,892
3,681	4,526	3,997	3,624	4,029	3,59	4,37	4,414	3,269	3,848
5,799	6,303	6,024	6,501	6,005	5,543	5,872	5,328	5,756	5,817
1,953	1,364	1,751	1,482	1,348	1,424	1,88	1,408	1,356	1,131
1,236	1,472	1,725	1,149	1,396	1,779	1,496	1,455	1,293	1,189
85,986	67,475	76,953	92,453	82,078	72,341	84,269	72,24	82,124	80,515
3,595	9,341	4,912	2,622	4,372	6,084	3,675	6,744	4,127	3,937
36,328	22,522	23,675	31,639	26,047	21,564	20,229	27,187	31,289	44,839

Appendix 4. Scores 21-30

21	22	23	24	25	26	27	28	29	30
5	5	5	5	5	2	4	6	5	4
25	18	29	41	16	11	12	18	13	22
269	170	339	322	197	87	122	172	189	265
5	3,6	5,8	8,2	3,2	5,5	3	3	2,6	5,5
3,536	2,191	4,266	5,63	1,924	0,707	1,414	2,608	1,949	7,047
10,76	9,444	11,69	7,854	12,313	7,909	10,167	9,556	14,538	12,045
5,562	3,838	5,306	4,624	5,896	3,145	6,191	5,102	12,129	6,671
1,413	1,535	1,451	1,301	1,386	1,425	1,32	1,413	1,228	1,468
0,678	0,731	0,692	0,558	0,765	0,757	0,607	0,647	0,501	0,844
4,312	4,518	4,593	3,913	4,203	3,862	3,762	4,227	3,878	4,498
2,4	2,28	2,5	1,915	2,169	2,436	2,025	1,968	1,807	2,556
0,678	0,707	1,271	1,429	1,725	1,088	0,566	1,022	1,324	1,606
74,86	75,8	89,8	92,22	95,73	85,99	71,23	84,61	90,66	94,52
1,013	1,64	0,542	1,339	-0,572	-0,13	1,263	0,461	-1,007	0,03
84,38	94,84	70,54	90,82	28,43	44,83	89,62	67,72	15,87	51,2
-0,128	-1,518	-0,767	-0,405	-0,397	-0,327	-2,052	-0,173	1,553	-1,038
45,22	6,55	22,36	34,46	34,83	37,45	2,2	43,25	93,94	15,15
0,79	0,663	0,046	-0,564	0,807	0,828	0,722	-0,065	1,804	-0,481
78,23	74,54	51,6	28,77	78,81	79,39	76,42	47,61	96,41	31,56
1,605	0,995	1,527	1,545	0,748	-1,327	1,986	1,172	2,191	-0,6
94,52	83,89	93,57	93,82	77,04	9,34	97,61	87,9	98,57	27,43
0,585	0,19	-0,038	0,675	2,108	0,42	0,263	1,295	1,043	-0,556
71,9	57,14	48,8	74,86	98,21	65,91	60,26	90,15	85,08	29,12
-3,687	-2,132	-3,109	-4,199	-6,197	-2,579	-2,451	-4,55	-3,122	-1,298
0	1,66	0,09	0	0	0,49	0,71	0	0,09	9,85
-0,004	-0,54	-0,051	0,542	1,445	0,259	-0,595	1,489	0,672	0,27
50	29,46	48,01	70,54	92,51	59,87	27,76	93,06	74,86	60,64
0,375	0,294	0,321	0,075	0,2	0,3	0,364	0,176	0,583	0,19
0,667	0,529	0,536	0,375	0,667	0,7	0,545	0,529	0,667	0,524
0,5	0,294	0,393	0,125	0,2	0,6	0,364	0,176	0,583	0,238
0,174	0,2	0,298	0,073	0,152	0,218	0,292	0,04	0,467	0,176
0,344	0,456	0,498	0,242	0,505	0,455	0,338	0,48	0,64	0,43
0,221	0,224	0,311	0,115	0,171	0,527	0,292	0,072	0,48	0,218
0,23	0,156	0,152	0,107	0,199	0,216	0,186	0,174	0,218	0,113
0,164	0,163	0,171	0,139	0,201	0,086	0,19	0,153	0,184	0,131
0,104	0,123	0,109	0,078	0,126	0,164	0,151	0,137	0,158	0,093
0,146	0,145	0,127	0,135	0,149	0,152	0,211	0,132	0,141	0,121
0,307	0,235	0,162	0,118	0,168	0,152	0,257	0,164	0,267	0,147
0,271	0,21	0,18	0,131	0,145	0,251	0,325	0,241	0,254	0,177
0,176	0,243	0,099	0,101	0,138	0,181	0,217	0,108	0,163	0,087
0,202	0,202	0,132	0,151	0,12	0,264	0,262	0,116	0,224	0,124

0,507	0,272	0,371	0,44	0,41	0,094	0,613	0,283	0,459	0,348
0,236	0,246	0,049	0,126	0,2	0	0,109	0,084	0,155	0,217
0,386	0,362	0,302	0,308	0,299	0,272	0,397	0,264	0,382	0,287
0,167	0,187	0,145	0,136	0,105	0,193	0,27	0,15	0,163	0,122
0,592	0,612	0,599	0,62	0,638	0,72	0,544	0,688	0,546	0,753
0,424	0,488	0,472	0,435	0,472	0,54	0,475	0,581	0,402	0,573
51,074	42,287	71,157	64,61	48,715	29	41,173	60,568	31,717	84,984
68,395	53,628	81,882	82,377	52,506	0	36,132	71,079	37,177	95,178
118,959	100	109,145	108,696	137,056	80,46	90,164	104,651	84,656	60,377
37,175	29,412	35,398	27,95	30,457	11,494	49,18	34,884	37,037	18,868
66,915	70,588	56,047	80,745	55,838	22,989	65,574	63,953	68,783	26,415
26,022	17,647	20,649	40,373	25,381	11,494	40,984	23,256	31,746	15,094
14,87	0	17,699	12,422	5,076	11,494	24,59	0	0	7,547
44,61	5,882	5,9	12,422	10,152	11,494	8,197	23,256	10,582	7,547
63,197	52,941	58,997	62,112	101,523	57,471	40,984	75,581	52,91	41,509
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
22,305	29,412	20,649	46,584	15,228	22,989	32,787	17,442	5,291	33,962
37,175	47,059	41,298	59,006	40,609	34,483	49,18	46,512	37,037	45,283
37,175	58,824	32,448	24,845	15,228	22,989	8,197	29,7	15,873	26,415
0,571	0,5	0,875	0,25	1,25	0,333	0,4	1,25	3	0,3
0,818	0,273	0,917	0,889	0,75	0,333	2,5	0,5	1,5	0,375
0,111	0,135	0,095	0,116	0,175	0,167	0,138	0,2	0,157	0,037
0,509	0,504	0,464	0,403	0,46	0,5	0,213	0,585	0,364	0,439
0,833	0,794	0,821	0,888	0,933	0,85	0,773	0,941	0,875	0,881
2,56	2,944	2,724	2,39	2,125	2,455	3,25	2,444	5,154	2,409
0,578	0,492	0,5	0,443	0,507	0,667	0,556	0,804	0,597	0,658
0,599	0,636	0,69	0,732	0,598	0,821	0,677	0,705	0,706	0,748
0,837	0,898	0,937	0,926	0,834	0,883	0,901	0,926	0,827	0,94
0,807	0,839	0,892	0,9	0,824	0,845	0,842	0,901	0,816	0,911
0,15	0,163	0,134	0,163	0,167	0,09	0,159	0,232	0,091	0,13
0,143	0,167	0,12	0,159	0,123	0,131	0,127	0,192	0,087	0,114
375,465	411,765	368,732	372,671	401,015	448,276	319,672	354,651	502,646	339,623
223,048	264,706	265,487	251,553	182,741	160,92	303,279	168,605	174,603	252,83
44,61	11,765	50,147	46,584	55,838	45,977	32,787	46,512	10,582	26,415
100,372	58,824	82,596	80,745	76,142	80,46	90,164	46,512	116,402	83,019
0	0	0	0	0	0	0	0	0	3,774
3,717	0	5,9	0	0	0	0	0	0	0
3,717	5,882	14,749	0	0	0	32,787	0	0	0
3,717	11,765	26,549	6,211	0	11,494	8,197	0	0	26,415
252,788	300,001	212,389	186,336	187,817	252,874	204,917	244,187	312,169	200,001
89,219	117,646	123,894	136,646	76,142	126,437	131,148	145,349	95,238	184,906
74,35	76,47	76,696	108,697	96,446	80,46	98,361	145,349	79,365	94,339
66,914	41,176	85,546	59,006	91,37	91,954	40,984	87,209	26,455	75,472

130,112	100	123,894	192,547	187,817	137,931	106,557	151,163	148,148	173,585
0	5,882	2,95	46,584	86,294	68,966	24,59	63,953	84,656	71,698
18,587	76,471	0	9,317	10,152	0	0	0	0	15,094
55,762	0	14,749	40,373	0	0	65,574	5,814	0	0
0	0	0	80,745	35,533	68,966	0	81,395	5,291	45,283
40,892	5,882	91,445	6,211	55,838	0	0	0	58,201	33,962
2,584	2,694	2,619	2,789	2,795	2,268	2,645	2,628	2,619	2,576
3,162	3,151	3,047	3,197	3,265	2,973	3,1	3,3	2,992	3,3
1,876	1,674	1,695	2,042	2,053	1,59	1,589	1,751	1,41	1,536
344,5	322,625	313,758	290	285,647	344,333	396,889	274,133	234,143	372,08
586,637	581,375	585,748	595,892	594,838	584,769	593,604	591,886	595,107	584,118
362,264	323,717	334,831	349,037	350,786	370,542	324,953	373,338	416,963	339,865
401,245	379,393	378,339	400,858	400,149	416,769	360,646	417,5	461,452	389,909
444,571	421,4	430,991	441,452	416,952	439,524	407,457	435,908	463,785	423,2
3,715	4,357	3,906	4,033	4,168	2,773	3,837	3,859	4,231	4,523
5,568	3,958	5,261	5,275	4,639	5,173	6,473	5,415	6,248	6,341
1,751	1,574	1,538	1,572	1,172	0,868	1,091	1,087	1,162	1,608
1,6	1,395	1,307	1,115	0,961	1,137	1,388	1,283	1,951	1,395
76,374	67,388	72,215	88,799	77,082	78,252	84,844	77,596	88,19	70,417
5,28	6,206	6,091	2,825	5,567	4,31	3,951	4,81	4,57	6,43
33,547	32,925	29,272	32,473	37,665	22,123	33,166	36,64	30,099	26,036

Appendix 5. Scores 31-40

31	32	33	34	35	36	37	38	39	40
5	7	5	5	5	4	5	5	5	4
31	11	20	23	17	10	28	27	26	23
209	132	242	227	138	208	189	205	199	235
6,2	1,571	4	4,6	3,4	2,5	5,6	5,4	5,2	5,75
3,633	1,134	1,871	3,362	2,191	1,915	4,722	2,881	6,648	4,646
6,742	12	12,1	9,87	8,118	20,8	6,75	7,593	7,654	10,217
3,444	8,567	5,486	4,948	2,176	9,647	3,738	4,244	3,31	5,289
1,445	1,333	1,364	1,291	1,572	1,548	1,455	1,532	1,367	1,579
0,678	0,626	0,694	0,56	0,887	0,772	0,71	0,795	0,66	0,885
4,048	4,189	4,025	3,969	4,63	4,76	4,333	4,673	3,955	4,855
2,068	2,186	2,123	2,179	2,34	2,32	2,397	2,408	2,246	2,927
1,275	1,062	0,702	1,872	0,382	0,161	1,821	0,069	1,901	0,82
89,8	85,54	75,8	96,93	64,8	56,36	96,56	52,39	97,13	79,39
1,379	-0,211	-0,118	0,078	1,556	-0,956	1,649	1,977	0,992	1,33
91,47	41,68	45,62	52,79	93,94	17,11	94,95	97,56	83,89	90,82
-0,968	-0,507	-0,147	-0,697	-0,193	0,727	-0,287	-0,089	0,012	-1,106
16,85	30,85	44,43	24,51	42,47	76,42	38,97	46,81	50,4	13,57
0,066	0,321	-0,977	0,757	1,424	2,302	0,153	-0,568	1,017	-0,515
52,39	62,55	16,6	77,34	92,22	98,93	55,96	28,77	84,38	30,5
-0,086	-1,19	1,094	0,806	0,507	-0,303	1,439	-0,165	-0,241	0,047
46,81	11,7	86,21	78,81	69,15	38,21	92,36	43,64	40,52	51,6
-0,041	1,123	0,589	1,116	0,402	-0,327	-0,816	-0,043	0,637	-0,701
48,4	86,86	71,9	86,65	65,54	37,45	20,9	48,4	73,57	24,2
-1,835	-3,397	-2,233	-4,902	-6,139	-6,943	-2,477	-2,322	-3,994	-4,425
3,36	0,03	1,29	0	0	0	0,68	1,2	0	0
1,373	1,865	1,37	1,281	-0,019	-1,4	0,994	-1,088	-0,31	-1,436
91,47	96,86	91,31	89,97	49,6	8,8	83,89	14,1	37,83	7,64
0,2	0,1	0,105	0,273	0,5	0,889	0,037	0,192	0,24	0,318
0,367	0,3	0,421	0,591	0,75	0,889	0,519	0,462	0,68	0,409
0,233	0,2	0,158	0,273	0,563	0,889	0,074	0,192	0,24	0,318
0,157	0,273	0,103	0,12	0,435	0,978	0,036	0,121	0,151	0,269
0,29	0,382	0,317	0,446	0,565	0,978	0,444	0,307	0,429	0,343
0,165	0,345	0,145	0,137	0,452	0,978	0,049	0,195	0,171	0,286
0,15	0,081	0,074	0,191	0,202	0,241	0,143	0,113	0,216	0,085
0,203	0,148	0,096	0,18	0,174	0,096	0,127	0,12	0,176	0,102
0,116	0,136	0,066	0,143	0,198	0,199	0,118	0,082	0,138	0,074
0,165	0,199	0,102	0,164	0,205	0,099	0,131	0,12	0,177	0,11
0,19	0,135	0,137	0,171	0,276	0,396	0,161	0,109	0,203	0,17
0,239	0,179	0,168	0,191	0,366	0,158	0,161	0,219	0,206	0,244
0,159	0,124	0,128	0,126	0,247	0,374	0,117	0,097	0,153	0,222
0,229	0,053	0,166	0,134	0,36	0,154	0,125	0,205	0,173	0,288

0,309	0,167	0,215	0,454	0,746	0,498	0,349	0,297	0,395	0,342
0,201	0,223	0,067	0,213	0,058	0,252	0,177	0,257	0,166	0,363
0,339	0,361	0,281	0,308	0,422	0,364	0,316	0,275	0,365	0,347
0,181	0,199	0,111	0,14	0,267	0,187	0,127	0,184	0,164	0,204
0,614	0,598	0,729	0,59	0,69	0,702	0,667	0,628	0,663	0,669
0,512	0,447	0,554	0,458	0,522	0,529	0,566	0,5	0,492	0,531
44,268	37,059	85,746	41,247	40,25	50,473	81,628	59,265	40,255	80,404
57,854	34,062	98,341	50,765	35,099	56,92	74,637	70,991	48,873	90,402
66,986	75,758	82,645	118,943	144,928	120,192	84,656	58,537	95,477	114,894
28,708	7,576	20,661	30,837	36,232	4,808	37,037	24,39	25,126	17,021
33,493	30,303	61,983	52,863	50,725	67,308	63,492	43,902	45,226	42,553
14,354	15,152	24,793	26,432	14,493	38,462	26,455	24,39	30,151	17,021
4,785	7,576	24,793	8,811	0	19,231	5,291	0	5,025	17,021
19,139	15,152	28,926	8,811	14,493	9,615	21,164	24,39	15,075	0
43,062	68,182	41,322	88,106	108,696	105,769	47,619	43,902	70,352	80,851
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
47,847	15,152	37,19	22,026	28,986	4,808	42,328	68,293	50,251	46,809
66,986	22,727	45,455	44,053	28,986	4,808	63,492	78,049	65,327	51,064
28,708	15,152	4,132	17,621	50,725	9,615	42,328	29,268	50,251	29,787
0,364	0,333	0,2	0,833	0	0	0,444	0,133	0,273	0,083
0,429	0,333	2,5	1,2	0,625	0,333	0,778	0,429	0,273	0,375
0,108	0,198	0,135	0,109	0,106	0,125	0,138	0,097	0,194	0,072
0,327	0,326	0,361	0,428	0,555	0,563	0,303	0,457	0,567	0,389
0,967	1	0,974	0,932	0,844	0,722	0,926	0,769	0,8	0,705
1,258	2,364	3,5	1,565	2,824	3,5	1,929	1,852	1,577	3,217
0,486	0,765	0,728	0,415	0,681	0,812	0,5	0,442	0,437	0,551
0,675	0,709	0,659	0,689	0,662	0,605	0,6	0,742	0,695	0,763
0,908	0,844	0,927	0,855	0,873	0,804	0,891	0,933	0,862	0,92
0,897	0,844	0,906	0,84	0,853	0,771	0,854	0,896	0,83	0,906
0,189	0,123	0,148	0,171	0,207	0,076	0,288	0,166	0,19	0,136
0,186	0,095	0,104	0,134	0,189	0,06	0,253	0,175	0,197	0,118
368,421	401,515	425,62	383,26	376,812	432,692	365,079	448,78	386,935	374,468
220,096	234,848	152,893	246,696	253,623	182,692	232,804	239,024	190,955	251,064
38,278	30,303	37,19	30,837	14,493	33,654	47,619	34,146	45,226	55,319
43,062	53,03	119,835	52,863	65,217	125	47,619	87,805	85,427	68,085
0	7,576	0	0	7,246	0	0	0	0	12,766
0	0	4,132	0	7,246	0	0	0	0	4,255
4,785	53,03	8,264	13,216	0	9,615	26,455	4,878	10,5	17,021
4,785	0	8,264	8,811	14,493	4,808	5,291	0	0	21,277
263,158	249,999	227,273	185,023	318,841	331,731	238,095	331,707	180,904	238,297
138,756	159,092	99,173	140,969	94,202	96,154	190,476	136,585	135,678	148,936
95,694	90,909	99,173	110,132	79,71	81,731	84,656	82,927	95,477	80,851
100,478	60,606	66,115	44,053	36,232	48,077	84,656	39,024	55,276	89,361

162,679	174,242	136,364	202,643	101,449	81,731	222,222	121,951	251,256	114,894
52,632	45,455	66,116	66,079	7,246	0	58,201	24,39	140,704	34,043
9,569	0	12,397	8,811	0	9,615	37,037	9,756	20,101	12,766
0	0	0	8,811	79,71	62,5	0	73,171	0	17,021
62,201	106,061	16,529	88,106	0	0	105,82	0	75,377	0
38,278	22,727	41,322	22,026	14,493	4,808	15,873	0	15,075	38,298
2,595	2,683	2,548	2,817	2,598	2,423	2,647	2,424	2,664	2,574
2,963	3,128	3,091	3,208	3,211	3,073	3,008	2,95	3,161	3,084
2,055	1,322	1,426	1,692	1,49	0	1,922	1,471	1,867	1,625
292,12	354,071	297,781	326,692	274,556	304,952	305,906	312,2	330,467	305
596,716	586,508	584,472	593,8	592,439	578,951	591,01	585,758	589,816	592,145
351,319	366,542	379,653	325,781	327,868	344,304	365,185	369,831	377,507	322,016
406,096	410,328	427,287	378,718	375,073	380,967	419,424	414,968	429,526	376,478
433,736	438,196	430,859	428,377	429,667	418,44	444,593	452,286	445,277	417,732
3,971	3,959	3,532	4,152	3,81	3,341	4,129	3,578	3,762	3,526
4,607	6,33	6,65	3,988	4,485	4,483	6,682	5,462	5,509	4,438
1,193	1,103	1,233	1,377	1,486	1,64	1,207	1,586	1,488	1,432
1,313	1,698	1,471	0,936	1,541	1,604	1,602	1,804	1,129	1,216
77,745	81,883	79,159	87,598	65,604	54,762	76,891	69,521	83,418	62,881
4,9	4,819	5,224	3,493	6,126	10,788	4,212	5,449	3,526	7,027
32,018	26,308	24,495	37,975	35,903	26,026	38,867	24,864	37,061	24,902

Appendix 6. Scores 41-50

41	42	43	44	45	46	47	48	49	50
4	5	8	4	5	6	5	5	5	4
19	13	26	30	34	18	12	28	22	6
151	135	257	238	335	233	265	252	258	205
4,75	2,6	3,25	7,5	6,8	3	2,4	5,6	4,4	1,5
4,113	1,517	2,252	4,509	4,087	1,265	1,673	3,362	1,949	1
7,947	10,385	9,885	7,933	9,853	12,944	22,083	9	11,727	34,167
3,808	5,347	2,971	4,996	4,554	6,126	12,717	4,023	5,435	31,218
1,437	1,578	1,502	1,391	1,331	1,506	1,592	1,492	1,481	1,454
0,669	0,728	0,735	0,702	0,62	0,841	0,9	0,806	0,739	0,75
4,225	4,793	4,276	4,126	4,012	4,562	4,777	4,504	4,496	4,322
2,298	2,116	2,115	2,273	2,015	2,508	2,51	2,406	2,202	2,282
1,362	-0,34	0,765	0,726	0,766	0,567	0,367	0,007	1,9	1,408
91,31	36,69	77,64	76,42	77,64	71,23	64,06	50	86,21	91,92
1,348	0,937	0,814	1,804	1,254	0,253	-0,333	0,945	0,468	-2,057
90,99	82,38	79,1	96,41	89,44	59,87	37,07	82,64	67,72	2,2
-1,117	1,159	-1,653	-0,478	-1,207	-2,435	0,003	-0,992	-1,642	-1,091
13,35	87,49	4,95	31,92	11,51	0,75	50	16,11	5,5	13,79
2,057	0,959	-0,206	-0,061	0,292	-0,13	-0,137	-0,015	1,58	0,957
97,98	82,89	42,07	47,61	61,41	44,83	44,83	49,6	94,18	82,89
0,545	-0,253	2,9	0,153	1,663	0,408	2,221	0,007	0,427	1,414
70,54	40,13	98,17	55,96	95,15	65,54	98,68	50	66,28	92,07
0,069	0,631	0,955	1,288	0,244	1,64	-0,618	1,87	-0,468	0,428
52,39	73,57	82,89	89,97	59,48	94,95	27,9	96,86	32,28	66,28
-1,046	-2,839	-2,386	-4,778	-2,35	-1,339	-2,454	-5,346	-1,933	-3,911
14,92	0,23	0,87	0	0,94	9,18	0,71	0	2,68	0
-0,539	-0,397	-0,414	1,327	-0,097	-0,305	-1,678	0,212	0,038	1,737
29,81	34,83	34,09	90,66	46,41	38,21	4,75	58,32	51,2	95,82
0,611	0,583	0,2	0,207	0,121	0,235	0,455	0,259	0,429	0,6
0,778	0,667	0,4	0,379	0,485	0,353	0,636	0,444	0,619	1
0,611	0,583	0,24	0,276	0,152	0,235	0,455	0,259	0,476	0,8
0,437	0,387	0,166	0,155	0,049	0,44	0,554	0,271	0,327	0,6
0,511	0,427	0,317	0,286	0,421	0,512	0,6	0,427	0,545	0,8
0,489	0,453	0,195	0,184	0,084	0,456	0,6	0,293	0,388	0,667
0,303	0,213	0,115	0,12	0,144	0,06	0,113	0,146	0,223	0,179
0,232	0,201	0,143	0,135	0,153	0,088	0,11	0,16	0,201	0,044
0,173	0,129	0,096	0,095	0,135	0,145	0,098	0,114	0,188	0,169
0,188	0,173	0,125	0,14	0,15	0,169	0,097	0,144	0,188	0,119
0,285	0,251	0,173	0,172	0,234	0,07	0,19	0,156	0,254	0,22
0,366	0,243	0,23	0,211	0,198	0,102	0,238	0,231	0,293	0,186
0,234	0,196	0,249	0,162	0,188	0,156	0,082	0,131	0,334	0,115
0,335	0,249	0,288	0,199	0,169	0,23	0,138	0,192	0,317	0,14

0,498	0,4	0,176	0,314	0,423	0,517	0,452	0,468	0,472	0,485
0,226	0,294	0,107	0,025	0,131	0,071	0,128	0,24	0,285	0,039
0,354	0,39	0,333	0,356	0,368	0,389	0,293	0,313	0,432	0,29
0,251	0,275	0,191	0,156	0,13	0,211	0,18	0,153	0,217	0,187
0,556	0,578	0,61	0,524	0,598	0,612	0,713	0,619	0,541	0,704
0,457	0,496	0,475	0,42	0,421	0,479	0,581	0,466	0,422	0,566
26,188	40,517	69,625	54,064	56,422	57,848	98,946	52,277	33,966	90,109
32,142	44,21	78,04	61,781	66,49	64,652	110,566	61,801	47,209	85,896
79,47	74,074	108,949	75,63	107,463	77,253	109,434	111,111	85,271	112,195
26,49	29,63	54,475	25,21	32,836	38,627	45,283	11,905	34,884	29,268
46,358	37,037	62,257	54,622	77,612	25,751	71,698	59,524	54,264	68,293
19,868	14,815	23,346	50,42	17,91	30,043	18,868	31,746	19,38	14,634
19,868	0	11,673	4,202	8,955	17,167	11,321	15,873	3,876	4,878
13,245	7,407	19,455	21,008	32,836	17,167	37,736	3,968	19,38	9,756
33,113	59,259	50,584	58,824	44,776	30,043	52,83	83,333	50,388	78,049
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
39,735	22,222	23,346	58,824	29,851	21,459	22,642	19,841	19,38	9,756
59,603	22,222	50,584	75,63	41,791	34,335	30,189	23,81	34,884	24,39
52,98	37,037	19,455	33,613	20,896	8,584	7,547	23,81	42,636	0
0,429	0	1	0,267	0,364	0,5	0,286	0,167	0,667	1
0,222	0,667	1,5	0,444	1,125	2,333	3,333	0,286	0,5	3
0,117	0,158	0,232	0,131	0,149	0,192	0,082	0,166	0,117	0,082
0,314	0,497	0,481	0,731	0,306	0,45	0,297	0,506	0,392	0,507
0,833	0,833	0,76	0,966	0,818	0,794	0,682	0,87	0,833	1
1,368	2,231	3,385	1,9	2,706	4	2,667	2,5	2,864	6
0,44	0,596	0,395	0,671	0,495	0,569	0,851	0,726	0,517	0,88
0,625	0,633	0,712	0,707	0,673	0,792	0,64	0,67	0,668	0,696
0,793	0,868	0,891	0,912	0,903	0,923	0,9	0,882	0,848	0,896
0,76	0,814	0,878	0,878	0,85	0,897	0,9	0,848	0,822	0,888
0,222	0,136	0,135	0,181	0,138	0,128	0,049	0,203	0,084	0,044
0,16	0,12	0,129	0,208	0,138	0,123	0,041	0,152	0,101	0,04
337,748	437,037	381,323	386,555	334,328	356,223	366,038	384,921	387,597	385,366
284,768	274,074	260,7	184,874	289,552	266,094	200	202,381	251,938	190,244
19,868	14,815	31,128	84,034	35,821	47,21	26,415	23,81	34,884	58,537
86,093	59,259	93,385	75,63	80,597	103,004	135,849	91,27	85,271	102,439
0	0	3,891	0	2,985	12,876	3,774	0	7,752	0
6,623	0	3,891	8,403	2,985	0	3,774	0	3,876	19,512
13,245	29,63	35,019	8,403	35,821	17,167	26,415	0	11,628	9,756
13,245	7,407	19,455	4,202	20,896	34,335	7,547	7,937	7,752	9,756
264,901	385,185	206,226	264,706	191,045	218,884	264,151	281,746	244,186	224,39
158,941	133,333	159,533	88,236	113,433	141,631	128,302	75,397	108,528	97,56
59,603	66,666	101,167	67,227	113,433	85,837	90,567	99,206	54,264	102,439
52,981	14,815	46,692	109,244	47,761	55,794	37,736	51,588	54,264	78,049

105,96	51,852	136,187	138,656	128,358	107,296	94,34	91,27	158,915	146,342
52,98	0	11,673	58,824	8,955	12,876	3,774	3,968	7,752	4,878
6,623	0	70,039	25,21	5,97	0	3,774	3,968	34,884	0
39,735	0	3,891	0	101,493	47,21	22,642	75,397	93,023	58,537
0	0	0	21,008	0	0	33,962	0	0	0
6,623	44,444	27,237	25,21	5,97	30,043	15,094	7,937	11,628	53,659
2,644	2,356	2,564	2,631	2,587	2,765	2,451	2,678	2,476	2,563
3,114	2,955	3,117	3,084	3,065	3,214	3,109	3,21	3,084	3,7
1,91	1,297	1,73	1,755	1,472	2,025	1,447	1,721	1,166	2,1
305,917	281,1	312,35	327,103	300,433	328,733	323,448	310,643	356,667	338,769
602,378	583,756	588,462	586,593	591,785	589,5	576,169	587,75	587,561	575,985
333,7	375,756	340	371,644	352,188	312,42	352,973	332,27	328,882	313,754
376,644	441,683	380,925	413,565	386,37	360,432	392,518	376,679	371,509	367,441
433,703	466,243	424,853	433,512	415,509	409,313	438,45	414,217	414,83	402,517
3,624	3,697	3,589	4,674	4,625	3,364	3,19	3,925	3,331	3,91
4,695	4,358	5,088	5,469	6,781	4,843	5,035	4,985	4,815	5,823
1,255	1,606	1,33	1,401	1,466	1,215	1,781	1,39	1,56	1,409
1,433	1,885	1,243	1,445	1,452	1,162	1,58	1,458	1,378	1,441
77,199	62,795	69,733	81,104	84,232	66,289	49,738	71,477	69,64	49,147
4,466	7,081	5,989	3,918	3,958	7,229	11,808	5,526	6,459	14,892
43,109	26,748	26,189	30,757	28,4	27,351	18,293	34,501	26,738	23,918

Appendix 7. Scores 51-60

51	52	53	54	55	56	57	58	59	60
5	5	4	5	4	4	5	5	4	5
19	14	16	18	25	22	41	17	15	26
244	185	227	148	188	278	378	240	193	222
3,8	2,8	4	3,6	6,25	5,5	8,2	3,4	3,75	5,2
2,775	1,483	4,761	3,286	4,573	4,796	11,145	1,949	2,754	4,025
12,842	13,214	14,188	8,222	7,52	12,636	9,22	14,118	12,867	8,538
5,597	4,282	8,742	4,138	3,776	6,244	5,008	6,489	7,782	3,723
1,615	1,265	1,401	1,378	1,41	1,324	1,389	1,363	1,254	1,464
0,884	0,542	0,693	0,723	0,793	0,633	0,671	0,677	0,543	0,709
4,84	3,741	4,15	4,182	4,027	4,158	4,14	4,1	3,834	4,392
2,438	1,922	2,123	2,195	2,316	2,346	2,216	2,252	1,956	2,362
-0,848	0,993	0,64	0,85	2,213	1,825	1,195	1,535	1,628	1,464
20,5	83,89	73,57	80,23	98,64	96,56	88,3	93,7	94,74	92,79
0,437	0,24	0,123	1,355	0,829	-0,121	1,147	-0,832	-1,068	0,622
66,64	59,1	54,78	91,15	79,39	45,22	87,29	20,33	14,46	73,24
-0,925	-1,134	-0,486	0,341	-0,391	0,026	-0,764	-1,231	-1,162	-2,619
17,88	12,92	31,56	63,31	34,83	50,8	22,36	10,93	12,3	0,44
0,335	0,438	-0,371	-0,091	1,857	1,464	-0,362	0,598	0,446	-0,343
62,93	66,64	35,57	46,41	96,78	92,79	35,94	72,24	67	36,69
-0,448	2,308	0,863	0,948	1,295	0,473	0,342	2,615	0,26	1,581
33	98,93	80,51	82,64	90,15	68,08	63,31	99,55	60,26	94,29
1,306	0,661	1,124	0,741	0,729	0,119	0,004	1,892	1,14	0,553
90,32	74,54	86,86	77,04	76,42	54,38	50	97,06	87,08	70,88
-4,444	-2,266	-3,368	-2,112	0,167	-3,397	-5,051	-1,941	-1,4	-2,453
0	1,19	0,04	1,74	56,36	0,03	0	2,62	8,8	0,71
-0,231	0,97	-1,074	0,821	0,246	0,417	0,961	0,897	1,274	2,075
40,9	83,15	14,23	79,39	59,48	65,91	83,15	81,33	89,8	98,08
0,556	0,231	0,267	0,118	0,167	0,524	0,225	0,125	0,143	0,12
0,611	0,615	0,467	0,412	0,792	0,667	0,475	0,625	0,714	0,28
0,556	0,385	0,333	0,118	0,208	0,524	0,25	0,313	0,143	0,16
0,474	0,224	0,2	0,192	0,087	0,436	0,079	0,217	0,084	0,078
0,526	0,424	0,39	0,312	0,487	0,515	0,318	0,461	0,558	0,346
0,519	0,282	0,238	0,192	0,103	0,473	0,11	0,348	0,105	0,122
0,161	0,141	0,091	0,107	0,292	0,219	0,14	0,159	0,184	0,103
0,179	0,117	0,12	0,142	0,187	0,183	0,131	0,14	0,189	0,143
0,137	0,115	0,082	0,098	0,166	0,182	0,082	0,141	0,122	0,123
0,174	0,145	0,123	0,166	0,177	0,174	0,12	0,159	0,138	0,15
0,263	0,204	0,161	0,177	0,23	0,286	0,172	0,215	0,217	0,165
0,245	0,195	0,157	0,201	0,205	0,25	0,154	0,159	0,21	0,162
0,202	0,156	0,124	0,162	0,183	0,219	0,104	0,174	0,16	0,11
0,213	0,16	0,152	0,184	0,16	0,235	0,133	0,168	0,161	0,119

0,664	0,46	0,43	0,376	0,449	0,422	0,29	0,547	0,318	0,481
0,108	0,199	0,104	0,193	0,242	0,273	0,108	0,084	0,21	0,251
0,387	0,355	0,322	0,352	0,354	0,369	0,318	0,343	0,313	0,319
0,231	0,174	0,144	0,164	0,156	0,137	0,114	0,145	0,152	0,126
0,6	0,65	0,72	0,651	0,589	0,608	0,592	0,609	0,67	0,605
0,498	0,497	0,485	0,561	0,455	0,473	0,45	0,429	0,508	0,514
73,671	61,667	73,328	45,362	27,937	40,074	79,334	54,432	53,238	70,896
75,601	63,961	74,526	60,027	50,321	58,07	104,568	56,961	64,919	68,008
94,262	108,108	110,132	74,324	74,468	100,719	95,238	120,833	93,264	99,099
16,393	43,243	26,432	27,027	53,191	10,791	15,873	50	20,725	45,045
45,082	70,27	48,458	67,568	47,872	57,554	63,492	70,833	41,451	58,559
24,59	16,216	22,026	33,784	5,319	32,374	39,683	20,833	15,544	13,514
8,197	16,216	17,621	13,514	0	32,374	2,646	20,833	25,907	9,009
4,098	21,622	0	20,27	21,277	3,597	5,291	20,833	0	22,523
73,771	48,649	57,269	33,784	26,596	57,554	68,783	50	41,451	58,559
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
28,689	37,838	30,837	60,811	42,553	14,388	60,847	12,5	20,725	27,027
36,885	64,865	52,863	74,324	79,787	25,18	71,429	41,667	41,451	54,054
20,492	5,405	17,621	33,784	42,553	21,583	7,937	16,667	10,363	18,018
0,25	0,625	0,625	0,2	0,778	0,6	0,167	1,75	0,8	0,857
0,667	2,5	1	0,333	0,778	0	1,25	1,2	1	1,8
0,239	0,113	0,142	0,137	0,148	0,111	0,112	0,167	0,13	0,177
0,518	0,338	0,45	0,57	0,459	0,436	0,511	0,603	0,412	0,42
0,833	0,923	0,733	0,941	0,854	0,905	0,925	0,875	0,964	0,98
5,526	4,214	4,75	3	1,72	2,682	2,561	2,882	4,667	2,654
0,973	0,339	0,548	0,509	0,397	0,655	0,623	0,595	0,6	0,316
0,685	0,611	0,739	0,689	0,563	0,655	0,699	0,645	0,689	0,668
0,879	0,863	0,91	0,911	0,791	0,871	0,914	0,897	0,894	0,922
0,857	0,813	0,878	0,901	0,765	0,852	0,881	0,85	0,83	0,922
0,121	0,102	0,076	0,151	0,244	0,125	0,161	0,093	0,081	0,16
0,129	0,096	0,077	0,15	0,208	0,116	0,163	0,081	0,091	0,158
381,148	345,946	334,802	432,432	430,851	374,101	335,979	387,5	445,596	423,423
172,131	291,892	277,533	195,946	223,404	169,065	248,677	183,333	165,803	207,207
20,492	21,622	48,458	47,297	42,553	43,165	37,037	25	20,725	76,577
98,361	86,486	88,106	87,838	79,787	93,525	66,138	108,333	93,264	81,081
0	5,405	4,405	6,757	5,319	0	0	0	0	0
4,098	10,811	4,405	0	0	3,597	13,228	4,167	5,181	4,505
12,295	27,027	30,837	0	10,638	3,597	5,291	8,333	10,363	4,505
12,295	37,838	17,621	6,757	10,638	3,597	15,873	12,5	0	13,514
327,869	194,595	185,021	263,514	196,808	230,216	195,767	170,834	165,804	202,703
69,673	118,918	132,159	101,351	143,616	107,913	132,276	95,833	72,539	121,623
127,049	97,297	83,701	87,838	74,468	93,525	111,111	104,167	129,533	90,091
36,885	54,054	74,89	67,568	53,191	71,942	79,366	58,334	46,632	117,117

40,984	108,108	92,511	189,189	239,362	201,439	171,958	179,167	191,71	180,18
0	27,027	0	108,108	143,617	143,885	63,492	104,167	113,99	49,55
0	0	0	6,757	0	10,791	39,683	4,167	20,725	9,009
12,295	27,027	57,269	47,297	0	7,194	0	0	15,544	0
0	0	0	20,27	74,468	17,986	50,265	20,833	0	85,586
28,689	32,432	17,621	6,757	21,277	10,791	13,228	45,833	36,269	27,027
2,407	2,625	2,584	2,58	2,791	2,67	2,542	2,705	2,674	2,693
2,988	3,115	3,21	2,952	3,179	3,035	2,972	3,191	3,167	3,107
1,2	1,175	1,338	1,9	2,114	1,807	1,356	1,557	1,851	1,671
309,789	329,429	343,188	286,35	300,25	294,364	294,452	364,889	341,909	371,444
579,556	597,899	587,573	592,2	594,753	589,364	590,13	589,837	584,375	586,233
329,447	336,277	356,921	404,556	372,069	360,781	347,219	347,217	345,743	321,568
375,395	375,565	397,476	437,185	427,286	414,791	390,379	385,02	398,875	361,767
421,514	416,317	424,125	451,566	450,5	446,151	427	410,597	425,725	398,933
3,738	4,485	4,81	4,206	4,266	4,045	4,35	4,701	3,783	3,798
5,658	6,37	5,695	5,901	5,777	5,541	5,821	6,517	6,53	5,787
1,562	1,183	1,616	1,335	1,436	1,38	1,369	1,247	1,418	1,174
1,841	1,366	1,33	1,514	1,295	1,394	1,251	1,225	1,164	1,244
57,171	86,404	73,91	81,911	79,916	81,999	79,967	77,195	87,687	74,315
8,475	4,49	6,475	3,877	3,981	4,961	4,396	5,999	4,225	5,015
24,24	26,864	21,767	27,103	47,156	33,362	28,599	29,034	28,917	29,951

Appendix 8. Scores A1-A10

A1	A2	A3	A4	A5	A6	A7	A8	A9	A10
5	3	6	5	6	3	5	3	6	5
18	8	19	9	16	14	26	10	14	16
236	148	180	152	146	189	304	108	263	212
3,6	2,667	3,167	1,8	2,667	4,667	5,2	3,333	2,333	3,2
2,074	1,528	1,602	0,837	1,033	3,512	2,49	4,041	1,751	0,837
13,111	18,5	9,474	16,889	9,125	13,5	11,692	10,8	18,786	13,25
8,396	11,174	2,632	6,827	3,364	6,466	6,276	4,917	10,214	3,838
1,525	1,696	1,544	1,651	1,733	1,492	1,497	1,62	1,548	1,665
0,882	0,987	0,867	0,93	0,963	0,891	0,84	0,993	0,867	1,1
4,525	5,297	4,544	4,632	5,185	4,519	4,477	4,806	4,726	5,024
2,708	2,82	2,446	2,761	2,694	2,526	2,561	2,756	2,556	2,877
-0,031	-0,716	0,516	1,338	-0,609	0,53	0,957	0,339	0,595	0,404
48,8	23,89	69,5	90,82	27,9	70,19	82,89	62,93	72,24	65,54
0,945	-0,79	1,431	-0,928	1,455	0,218	1,076	1,303	-1,259	1,067
82,64	21,48	92,36	17,88	92,65	58,32	85,77	90,32	10,56	85,54
-0,866	-0,317	-0,862	-2,057	-0,281	-0,419	-1,106	-0,836	-0,502	-1,089
19,49	37,83	19,49	2,2	38,97	33,72	13,57	20,33	30,85	14,1
-0,334	-0,352	-0,121	-0,583	-0,317	-0,483	-0,709	-0,632	0,636	1,1
37,07	36,32	45,22	28,1	37,83	31,56	23,89	26,43	73,57	84,13
0,85	-1,127	2,95	1,514	1,277	1,892	2,495	4,346	0,965	0,204
79,95	13,14	99,84	93,45	89,8	97,06	99,36	100	83,15	57,93
0,12	-1,011	0,185	-0,281	-0,255	0,833	-0,25	-0,896	-0,13	-1,186
54,38	15,62	57,14	38,97	40,13	79,67	40,13	18,67	44,83	11,9
-1,383	-1,4	-1,744	-2,743	-0,162	-2,398	-1,96	-0,598	-1,744	0,209
8,38	8,8	4,9	0,31	43,64	0,84	2,5	27,76	4,9	57,93
-1,813	-1,016	-0,992	0,118	0,891	-0,109	-0,541	1,1	0,78	-1,445
3,51	15,62	16,11	54,38	81,33	46,02	29,46	84,38	77,94	7,49
0,412	0,714	0,222	0,125	0,4	0,231	0,16	0,222	0,538	0,4
0,412	0,714	0,389	0,5	0,4	0,462	0,36	0,444	0,769	0,6
0,412	0,857	0,278	0,25	0,6	0,385	0,24	0,222	0,769	0,6
0,272	0,464	0,274	0,25	0,257	0,329	0,137	0,244	0,435	0,295
0,32	0,464	0,415	0,583	0,267	0,482	0,41	0,333	0,6	0,61
0,32	0,5	0,296	0,361	0,448	0,412	0,161	0,289	0,576	0,467
0,119	0,106	0,153	0,112	0,132	0,075	0,079	0,116	0,167	0,181
0,145	0,092	0,22	0,118	0,131	0,097	0,132	0,17	0,166	0,152
0,068	0,083	0,103	0,095	0,103	0,078	0,088	0,083	0,137	0,155
0,107	0,115	0,145	0,105	0,12	0,087	0,112	0,136	0,141	0,169
0,164	0,319	0,192	0,093	0,275	0,099	0,096	0,127	0,295	0,325
0,17	0,119	0,214	0,072	0,144	0,092	0,174	0,212	0,155	0,195
0,104	0,257	0,172	0,145	0,232	0,109	0,116	0,083	0,219	0,349
0,141	0,148	0,187	0,054	0,149	0,124	0,195	0,189	0,112	0,22

0,441	0,595	0,404	0,21	0,434	0,503	0,357	0,43	0,463	0,331
0,13	0,157	0,26	0,189	0,139	0,187	0,117	0,347	0,278	0,226
0,297	0,329	0,332	0,237	0,316	0,305	0,319	0,302	0,337	0,388
0,139	0,143	0,145	0,152	0,129	0,114	0,148	0,151	0,155	0,168
0,672	0,794	0,713	0,824	0,704	0,709	0,644	0,804	0,667	0,579
0,511	0,676	0,567	0,625	0,568	0,577	0,493	0,667	0,51	0,458
63,245	91,126	60	80,879	54,817	73,856	76,961	72,818	80,057	54,467
71,883	112,611	78,024	95,517	62,639	90,913	104,617	80,064	80,838	61,24
72,034	60,811	127,778	105,263	75,342	95,238	118,421	120,37	83,65	75,472
33,898	13,514	66,667	39,474	27,397	37,037	42,763	74,074	15,209	23,585
46,61	20,27	77,778	59,211	54,795	52,91	85,526	92,593	45,627	42,453
21,186	0	5,556	19,737	0	26,455	16,447	18,519	15,209	0
8,475	0	0	13,158	27,397	15,873	16,447	18,519	22,814	23,585
12,712	0	50	19,737	13,699	10,582	23,026	27,778	19,011	18,868
29,661	47,297	50	59,211	34,247	42,328	46,053	27,778	45,627	33,019
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
29,661	40,541	50	13,158	47,945	31,746	32,895	46,296	3,802	33,019
55,085	54,054	72,222	32,895	61,644	58,201	42,763	74,074	11,407	33,019
16,949	0	27,778	6,579	47,945	5,291	23,026	18,519	0	47,17
0,75	0,286	0,4	1	0,25	0,714	0,273	0,5	1	0
1,2	1	2	2,5	0,5	3	1,5	2,333	3	0,455
0,102	0,037	0,087	0,059	0,049	0,083	0,11	0,024	0,049	0,104
0,308	0,316	0,478	0,447	0,632	0,332	0,458	0,276	0,386	0,46
0,676	0,786	0,722	0,813	0,967	0,846	0,78	0,944	0,923	0,733
3,778	8,875	2,947	4,222	2,938	3,929	2,462	1,9	5	4,688
0,773	1,244	0,632	0,571	1	0,453	0,552	0,622	0,649	0,75
0,716	0,689	0,646	0,701	0,601	0,687	0,734	0,609	0,649	0,676
0,883	0,908	0,898	0,911	0,915	0,915	0,93	0,907	0,893	0,875
0,875	0,893	0,869	0,89	0,917	0,89	0,891	0,882	0,887	0,87
0,115	0,079	0,171	0,042	0,137	0,103	0,139	0,067	0,05	0,126
0,106	0,076	0,164	0,039	0,147	0,087	0,12	0,091	0,055	0,152
347,458	378,378	338,889	361,842	335,616	402,116	371,711	398,148	376,426	382,075
237,288	202,703	227,778	217,105	226,027	253,968	250	203,704	258,555	259,434
21,186	27,027	50	59,211	20,548	15,873	39,474	27,778	53,232	37,736
131,356	94,595	94,444	111,842	82,192	121,693	88,816	157,407	102,662	84,906
16,949	0	0	0	0	21,164	6,579	18,519	0	0
4,237	0	0	13,158	0	5,291	13,158	0	0	0
0	27,027	22,222	32,895	61,644	5,291	32,895	37,037	30,418	37,736
25,424	27,027	16,667	19,737	20,548	15,873	26,316	9,259	26,616	23,585
292,373	331,082	233,334	203,948	321,917	264,55	223,684	250	266,16	283,019
114,406	101,351	105,555	118,422	130,138	100,529	138,158	129,63	140,684	132,076
72,034	114,865	122,222	92,105	109,589	74,074	92,105	74,074	87,453	113,208
59,322	67,568	77,778	111,842	61,643	58,201	62,5	46,296	68,441	37,736

59,322	54,054	150	138,158	41,096	95,238	128,29	120,37	68,441	108,491
4,237	0	11,111	0	0	15,873	62,5	0	0	0
21,186	6,757	38,889	78,947	41,096	21,164	39,474	37,037	11,407	108,491
8,475	13,514	50	0	0	5,291	0	55,556	34,221	0
0	13,514	0	0	0	0	0	0	0	0
8,475	6,757	16,667	13,158	0	37,037	6,579	0	3,802	0
2,36	2,212	2,553	2,372	2,266	2,647	2,557	2,285	2,457	2,326
3,034	2,846	3,073	3,019	2,779	3,124	3,071	3,023	3,005	2,898
1,697	0,84	1,626	1,67	1,213	1,898	1,443	1,723	1,024	1,718
334,889	407,909	328,421	344,25	353,864	347,259	326,185	357,273	327,636	363,182
584,728	576,719	585,192	577,019	576,708	592,75	583,057	577,265	580,625	573,244
354,689	356,862	337,455	322,042	362,2	356,169	361	341,621	349,386	361,9
394,691	390,125	380,987	353,604	390,969	403,764	385,81	389,588	384,606	386,802
437	433,068	440,509	391,317	452,182	438,742	426,739	428,364	421,041	432,235
4,042	3,953	4,427	3,749	4,149	4,481	3,826	4,133	3,934	3,642
6,398	6,569	6,933	6,669	6,589	6,264	6,729	7,046	6,712	6,988
1,584	1,249	1,443	1,539	1,482	1,207	1,131	1,223	1,613	1,672
1,981	2,097	1,666	1,452	2,007	1,696	1,609	1,696	1,945	2,062
64,512	44,576	66,597	50,018	50,961	66,909	68,321	58,821	56,806	52,527
7,518	11,638	6,324	10,479	8,418	7,281	6,634	7,738	10,003	9,224
20,641	14,457	30,141	16,055	20,576	23,98	24,394	15,877	21,31	23,792

Appendix 9. Scores A11-A17

A11	A12	A13	A14	A15	A16	A17
4	5	5	4	3	4	6
15	15	12	17	11	21	19
175	189	188	195	175	221	176
3,75	3	2,4	4,25	3,667	5,25	3,167
1,893	1,225	1,673	2,754	3,055	3,304	1,835
11,667	12,6	15,667	11,471	15,909	10,524	9,263
4,435	5,742	9,882	5,444	13,118	5,115	3,902
1,777	1,603	1,75	1,518	1,474	1,579	1,881
1,105	0,998	1,068	0,916	0,801	0,972	1,021
4,823	4,683	5,043	4,349	4,52	4,566	5,483
2,914	2,763	2,945	2,66	2,489	2,754	2,734
1,79	0,726	0,553	0,777	0,552	1,553	-0,323
96,25	76,42	70,88	77,94	70,88	93,94	37,45
-1,109	0,585	0,239	0,841	-0,333	0,405	1,879
13,57	71,9	59,1	79,95	37,07	65,54	96,93
-3,287	-1,083	-0,395	-1,501	-0,894	-0,989	0,341
0,05	14,1	34,83	6,68	18,67	16,35	63,31
1,455	0,601	0,632	0,417	-1,123	1,411	0,815
92,65	72,57	73,57	65,91	13,14	92,07	79,1
10,2	2,237	0,897	3,416	3	2,014	2,444
100	98,71	81,33	99,97	99,86	97,78	99,27
0,346	0,101	-1,552	-0,204	0,116	0,311	-1,361
63,31	53,98	6,6	42,07	54,38	62,17	8,69
-1,985	-3,387	-3,092	0,387	-2,293	-1,466	-1,465
2,39	0,04	0,1	64,8	1,1	7,21	7,21
2,539	-0,409	-1,127	0,408	0,75	-0,663	0,597
99,43	34,09	13,14	65,54	77,34	25,46	72,24
0,643	0,357	0,545	0,375	0,4	0,3	0,444
0,786	0,714	0,636	0,563	0,5	0,75	0,611
0,643	0,5	0,727	0,5	0,5	0,3	0,556
0,568	0,316	0,492	0,191	0,2	0,142	0,356
0,684	0,516	0,538	0,33	0,345	0,71	0,481
0,579	0,379	0,631	0,296	0,309	0,142	0,415
0,168	0,211	0,104	0,163	0,067	0,233	0,197
0,111	0,148	0,113	0,168	0,081	0,23	0,205
0,177	0,136	0,135	0,106	0,053	0,173	0,185
0,167	0,152	0,18	0,147	0,083	0,159	0,205
0,297	0,198	0,218	0,21	0,155	0,231	0,268
0,198	0,189	0,144	0,217	0,179	0,227	0,225
0,281	0,134	0,241	0,194	0,106	0,179	0,219
0,185	0,119	0,12	0,19	0,158	0,16	0,241

0,577	0,411	0,274	0,304	0,323	0,524	0,618
0,062	0,204	0,318	0,038	0,072	0,186	0,166
0,402	0,291	0,336	0,347	0,238	0,325	0,414
0,185	0,149	0,199	0,159	0,144	0,175	0,181
0,577	0,679	0,673	0,615	0,854	0,7	0,556
0,497	0,55	0,534	0,472	0,629	0,471	0,497
58,733	56,107	64,754	45,766	93,573	40,517	54,332
53,188	73,412	63,014	52,142	107,7	45,843	46,948
165,714	111,111	106,383	97,436	125,714	113,122	119,318
108,571	42,328	37,234	66,667	62,857	67,873	56,818
131,429	58,201	47,872	76,923	57,143	45,249	68,182
17,143	37,037	0	10,256	17,143	9,5	11,364
5,714	26,455	5,319	10,256	22,857	4,525	17,045
51,429	21,164	15,957	30,769	34,286	18,1	34,091
68,571	52,91	79,787	25,641	51,429	49,774	45,455
0	0	0	0	0	0	0
0	0	0	0	0	0	0
0	21,164	15,957	10,256	22,857	31,674	34,091
51,429	26,455	37,234	35,897	45,714	67,873	45,455
22,857	5,291	21,277	30,769	11,429	40,724	62,5
9	0,2	1	1,667	0,8	1	0,286
3,8	4	1	1,857	3,333	1,5	0,75
0,077	0,104	0,056	0,059	0,051	0,166	0,037
0,409	0,471	0,297	0,395	0,391	0,48	0,563
0,893	0,786	0,727	0,844	0,9	0,775	0,917
2,8	3	2,083	1,588	0,909	2,571	2
0,531	0,596	0,538	0,514	0,596	0,534	0,918
0,708	0,673	0,663	0,631	0,711	0,645	0,581
0,866	0,838	0,885	0,908	0,901	0,839	0,88
0,856	0,8	0,872	0,881	0,897	0,815	0,824
0,115	0,157	0,095	0,142	0,037	0,143	0,197
0,096	0,107	0,076	0,151	0,033	0,114	0,163
382,857	343,915	414,894	471,795	371,429	393,665	392,045
194,286	259,259	207,447	184,615	211,429	271,493	170,455
28,571	26,455	42,553	10,256	28,571	40,724	39,773
131,429	63,492	117,021	169,231	137,143	76,923	119,318
5,714	21,164	15,957	0	5,714	0	5,682
0	10,582	0	0	11,429	0	0
17,143	10,582	5,319	25,641	28,571	0	28,409
17,143	26,455	10,638	10,256	45,714	27,149	0
251,428	253,968	292,553	282,051	222,858	230,77	335,227
102,858	169,312	111,701	107,693	114,286	122,173	73,864
102,857	74,074	69,149	66,667	80	76,924	136,364
108,571	58,201	47,872	30,769	57,142	54,299	45,455

91,429	121,693	111,702	102,564	91,429	185,52	113,636
11,429	42,328	31,915	15,385	28,571	0	0
74,286	21,164	47,872	56,41	5,714	158,371	85,227
0	0	0	0	34,286	0	11,364
0	0	0	0	0	0	0
0	15,873	26,596	15,385	5,714	13,575	5,682
2,648	2,46	2,275	2,493	2,406	2,485	2,192
3,109	2,984	2,984	3,109	3,069	3,059	2,693
1,964	1,413	1,393	1,68	1,202	1,696	1,657
401,167	337,167	346,846	336,375	337	381,308	368,826
583,416	583,656	570,672	578,892	579,667	587,917	567,859
297,634	330,667	359,529	349,911	338,923	336,4	395,306
349,961	373,578	392,966	388,077	366,228	381,694	410,437
392,525	408,478	438,625	410,653	413,261	450,034	465,294
4,568	4,033	3,343	3,887	4,047	4,849	3,614
7,074	7,47	7,091	6,607	5,997	6,815	6,846
1,703	1,268	1,725	1,603	1,333	1,69	1,556
1,896	1,989	2,161	1,891	1,492	1,755	2,223
44,659	58,432	42,883	66,769	65,987	62,57	38,3
9,929	8,239	11,17	6,796	8,008	7,147	10,218
29,59	30,248	16,74	27,553	14,167	31,087	26,017

Appendix 10. Scores B1 benchmarking samples

B1,1	B1,2	B1,3	B1,4	B1,5	B1,6	B1,7	B1,8
6	3	3	4	4	3	4	4
19	19	18	13	17	36	12	28
406	335	344	304	268	643	279	474
3,167	6,333	6	3,25	4,25	12	3	7
1,472	5,774	2,646	0,957	3,403	7	0,816	3,559
21,368	17,632	19,111	23,385	15,765	17,861	23,25	16,929
13,66	9,4	7,459	17,788	8,623	11,528	10,244	9,302
1,416	1,448	1,308	1,359	1,429	1,384	1,387	1,563
0,847	0,867	0,604	0,645	0,708	0,734	0,796	0,974
4,16	4,319	4,073	3,993	4,291	3,908	3,986	4,709
2,467	2,413	2,424	2,282	2,455	2,325	2,455	2,726
1,541	1,348	0,914	-0,381	-0,165	1,208	1,476	0,092
93,82	90,99	81,86	35,2	43,64	88,49	92,92	53,59
-0,685	-0,437	-0,461	-0,905	-0,823	-0,555	-0,444	-0,116
24,83	33,36	32,28	18,41	20,61	29,12	33	45,62
-0,778	-0,561	1,586	1,345	1,264	-0,342	-0,526	-0,169
22,06	28,77	94,29	90,99	89,62	36,69	30,15	43,64
0,544	0,274	0,552	-0,552	-0,464	0,84	0,743	-0,666
70,54	60,64	70,88	29,12	32,28	79,95	77,04	25,46
0,302	0,945	-0,207	1,342	-0,399	1,504	0,942	0,134
61,79	82,64	42,07	90,99	34,46	93,32	82,64	55,17
0,26	0,548	-0,143	0,221	1,943	0,419	1,318	0,489
59,87	70,54	44,43	58,71	97,38	65,91	90,49	68,44
-4,03	-3,373	-3,422	-5,586	-2,028	-4,528	-3,472	-3,221
0	0,04	0,03	0	2,17	0	0,03	0,06
-0,219	0	-1,128	0,888	0,196	0,074	-2,037	-0,303
41,29	50	13,14	81,06	57,53	52,79	2,12	38,21
0,222	0,278	0,294	0,417	0,25	0,429	0,091	0,222
0,611	0,722	0,706	0,5	0,313	0,829	0,727	0,481
0,333	0,5	0,412	0,5	0,25	0,486	0,273	0,37
0,17	0,163	0,2	0,253	0,235	0,177	0,215	0,284
0,711	0,459	0,648	0,387	0,313	0,656	0,723	0,462
0,215	0,281	0,248	0,28	0,243	0,239	0,338	0,373
0,144	0,173	0,12	0,085	0,104	0,181	0,13	0,085
0,125	0,135	0,104	0,084	0,116	0,121	0,09	0,095
0,168	0,096	0,11	0,061	0,067	0,137	0,143	0,08
0,127	0,122	0,096	0,074	0,089	0,118	0,102	0,091
0,18	0,243	0,17	0,19	0,16	0,25	0,124	0,139
0,132	0,134	0,131	0,155	0,147	0,157	0,086	0,116
0,177	0,129	0,167	0,18	0,1	0,148	0,141	0,146
0,165	0,145	0,134	0,153	0,129	0,152	0,111	0,145

0,475	0,39	0,419	0,35	0,292	0,632	0,229	0,378
0,074	0,107	0,067	0,164	0,126	0,025	0,068	0,197
0,332	0,318	0,295	0,255	0,251	0,348	0,258	0,306
0,135	0,118	0,119	0,13	0,118	0,11	0,127	0,105
0,725	0,698	0,747	0,744	0,783	0,551	0,717	0,73
0,453	0,493	0,459	0,503	0,567	0,345	0,52	0,506
60,074	58,917	60,231	70,77	59,355	60,876	66,321	84,984
61,654	95,398	57,846	70,765	68,698	72,005	64,654	108,886
108,374	113,433	98,837	125	70,896	122,862	118,28	88,608
29,557	23,881	29,07	36,184	22,388	29,549	32,258	16,878
49,261	50,746	31,977	65,789	29,851	66,874	43,011	42,194
24,631	20,896	8,721	26,316	11,194	38,88	17,921	21,097
7,389	23,881	5,814	3,289	0	26,439	14,337	14,768
36,946	20,896	20,349	13,158	11,194	24,883	10,753	10,549
71,429	65,672	63,953	85,526	44,776	69,984	64,516	59,072
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
27,094	14,925	31,977	13,158	11,194	20,218	25,09	16,878
41,872	23,881	46,512	26,316	22,388	40,435	39,427	21,097
12,315	8,955	20,349	9,868	7,463	10,886	10,753	8,439
0,5	0,5	0,417	0,8	0,75	0,929	0,5	0,222
1,667	1,25	0,875	2,75	1,333	1,125	2	1,6
0,119	0,076	0,086	0,126	0,11	0,09	0,1	0,084
0,468	0,486	0,426	0,22	0,631	0,476	0,66	0,489
0,806	0,861	0,735	0,917	0,844	0,857	0,636	0,833
3,789	4	2,389	7,385	2,647	5,389	2,5	4,036
0,694	0,63	0,874	1,099	1,038	0,804	0,75	0,688
0,699	0,683	0,662	0,631	0,636	0,693	0,649	0,678
0,927	0,891	0,875	0,891	0,846	0,88	0,849	0,898
0,897	0,867	0,86	0,876	0,814	0,865	0,836	0,871
0,088	0,103	0,092	0,031	0,103	0,077	0,075	0,114
0,102	0,088	0,084	0,047	0,084	0,063	0,065	0,111
320,197	343,284	351,744	388,158	388,06	359,254	301,075	364,979
209,36	232,836	220,93	128,289	186,567	194,401	340,502	221,519
59,113	26,866	31,977	19,737	22,388	38,88	35,842	21,097
100,985	98,507	107,558	118,421	100,746	90,202	78,853	113,924
0	0	2,907	0	3,731	0	3,584	6,329
7,389	5,97	2,907	3,289	3,731	4,666	14,337	4,219
14,778	23,881	11,628	16,447	18,657	7,776	7,168	29,536
17,241	14,925	29,07	3,289	7,463	34,215	86,022	10,549
182,266	211,941	218,024	276,316	294,776	213,064	186,379	246,836
103,448	134,329	101,745	69,077	126,866	96,422	161,289	120,252
96,059	74,627	63,953	118,421	59,701	108,864	86,021	130,802
98,522	65,672	61,047	49,342	33,582	74,65	64,516	44,304

145,32	152,239	101,744	55,921	63,433	127,527	111,111	88,608
137,931	68,657	72,674	42,763	33,582	97,978	89,606	67,511
2,463	32,836	0	3,289	0	12,442	7,168	0
0	20,896	2,907	0	3,731	0	0	8,439
0	0	2,907	0	7,463	0	0	0
0	2,985	2,907	0	3,731	0	0	6,329
2,493	2,694	2,323	2,225	2,464	2,565	2,653	2,466
3,09	3,133	3,08	2,946	3,202	3,074	3,142	2,983
1,633	1,572	1,042	1,167	0,991	1,674	1,852	1,413
365,459	361,818	260,2	311,214	303,267	355,123	341,87	378,985
586,237	587,59	588,53	581,103	578,293	587,358	586,193	586,429
346,241	338,211	415,231	415,336	429,716	346,171	332,787	373,357
395,424	383,785	455,664	447,966	461,485	389,303	378,076	408,561
433,218	437,637	451,376	441,622	447	438,561	427,25	437,32
3,83	4,481	3,507	3,267	3,661	3,859	5,478	4,163
7,003	6,921	7,275	6,774	6,36	6,731	6,851	6,509
1,23	1,34	1,402	1,301	1,574	1,199	1,57	1,289
1,333	1,565	1,69	1,781	1,907	1,425	1,556	1,673
65,353	66,438	76,781	68,128	69,94	71,62	65,896	57,422
9,452	8,373	7,298	9,566	7,421	7,707	9,844	9,456
23,233	30,128	18,448	10,719	21,437	26,088	25,268	21,145

Appendix 11. Scores B2 benchmarking samples

B2,1	B2,2	B2,3	B2,4	B2,5	B2,6	B2,7	B2,8	B2,9
6	4	4	4	4	4	4	3	4
33	20	9	14	17	12	18	13	15
591	425	244	282	384	286	266	256	362
5,5	5	2,25	3,5	4,25	3	4,5	4,333	3,75
2,429	2,16	0,5	1	1,5	1,414	3,786	2,887	1,5
17,909	21,25	27,111	20,143	22,588	23,833	14,778	19,692	24,133
9,997	7,638	10,768	6,111	10,869	15,497	7,527	12,841	5,902
1,613	1,445	1,385	1,766	1,365	1,545	1,278	1,543	1,414
0,906	0,76	0,672	0,985	0,68	0,919	0,561	0,811	0,762
4,876	4,365	4,398	5,007	4,313	4,385	3,88	4,707	4,34
2,504	2,362	2,245	2,84	2,239	2,658	2,021	2,595	2,23
-0,032	1,375	0,417	-0,612	0,73	0,464	0,497	-0,307	0,759
48,8	91,47	65,91	27,09	76,42	67,72	68,79	38,21	77,34
-0,694	-0,574	-1,788	-0,149	-0,737	-0,916	0,139	0,003	-0,94
24,51	28,43	3,75	44,43	23,27	18,14	55,17	50	17,36
-0,984	-0,935	0,651	0,13	0,358	0,405	0,472	-0,557	0,559
16,35	17,62	74,22	55,17	63,68	65,54	68,08	29,12	70,88
-0,724	0,493	1,206	-0,86	-0,416	-1,014	-0,363	-1,327	0,995
23,58	68,79	88,49	19,49	34,09	15,62	35,94	9,34	83,89
-0,042	0,571	1,34	0,658	-0,239	0,614	1,049	3,212	2,415
48,4	71,57	90,82	74,22	40,9	72,91	85,08	99,93	99,2
0,161	0,075	1,343	-0,972	0,119	-0,778	1,816	-0,388	0,766
56,36	52,79	90,99	16,6	54,38	22,06	96,49	35,2	77,64
-3,272	-3,842	-5,202	-2,524	-1,798	-4,178	-3,714	-2,451	-3,491
0,05	0	0	0,59	3,67	0	0	0,71	0,02
1,516	0,624	1,507	-1,864	-2,271	1,074	-0,402	0,209	0,097
93,45	73,24	93,32	3,14	1,16	85,77	34,46	57,93	53,59
0,188	0,368	0,875	0,308	0,188	0	0,412	0,167	0,571
0,469	0,842	0,875	0,462	0,5	0,545	0,529	0,25	0,929
0,406	0,368	0,875	0,462	0,313	0,091	0,412	0,333	0,714
0,24	0,269	0,778	0,224	0,139	0,108	0,184	0,24	0,526
0,411	0,676	0,806	0,341	0,348	0,492	0,296	0,4	0,716
0,375	0,317	0,778	0,471	0,191	0,262	0,216	0,467	0,642
0,088	0,152	0,145	0,094	0,074	0,082	0,121	0,053	0,179
0,098	0,097	0,066	0,078	0,068	0,097	0,156	0,08	0,125
0,074	0,147	0,126	0,077	0,062	0,079	0,064	0,045	0,147
0,091	0,112	0,083	0,083	0,085	0,088	0,107	0,067	0,122
0,132	0,137	0,308	0,213	0,179	0,146	0,184	0,197	0,179
0,103	0,118	0,197	0,146	0,146	0,112	0,144	0,108	0,096
0,149	0,126	0,291	0,214	0,162	0,132	0,101	0,213	0,181
0,15	0,145	0,221	0,154	0,143	0,118	0,135	0,188	0,096

0,527	0,447	0,518	0,503	0,263	0,287	0,37	0,62	0,484
0,078	0,092	0,013	0,2	0,094	0,022	0,242	0,027	0,067
0,326	0,297	0,327	0,317	0,301	0,233	0,296	0,32	0,318
0,126	0,11	0,179	0,159	0,111	0,128	0,126	0,169	0,121
0,697	0,619	0,669	0,764	0,795	0,865	0,713	0,818	0,63
0,483	0,445	0,531	0,563	0,521	0,612	0,519	0,598	0,456
83,581	85,974	80,335	96	81,14	93,156	86,596	105,821	72,424
96,654	92,231	79,499	99,005	99,295	109,139	91,207	110,365	87,475
79,526	101,176	122,951	102,837	72,917	108,392	116,541	117,188	99,448
20,305	23,529	20,492	28,369	15,625	34,965	22,556	58,594	41,436
37,225	56,471	61,475	46,099	33,854	52,448	48,872	74,219	77,348
35,533	30,588	16,393	21,277	23,438	27,972	26,316	23,438	35,912
3,384	16,471	12,295	17,73	13,021	3,497	26,316	15,625	11,05
16,92	11,765	20,492	28,369	10,417	24,476	15,038	31,25	19,337
49,069	58,824	90,164	49,645	36,458	66,434	56,391	50,781	52,486
0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0
11,844	28,235	8,197	28,369	20,833	17,483	22,556	19,531	27,624
27,073	40	20,492	46,099	31,25	20,979	37,594	46,875	49,724
6,768	11,765	0	14,184	13,021	13,986	26,316	15,625	5,525
1,125	0,385	1	0,556	0,444	0,167	0,571	1,167	0,727
1,2	1	3	0,8	0,667	2	0,625	2,2	3,333
0,132	0,081	0,136	0,092	0,092	0,039	0,129	0,084	0,101
0,58	0,405	0,612	0,391	0,458	0,346	0,639	0,465	0,505
0,953	0,921	1	0,654	0,625	0,955	0,824	0,833	0,857
3,121	5,2	4,111	6,214	7,882	5,583	2,833	5,308	4,933
0,667	0,659	0,82	0,986	0,717	0,976	0,463	0,862	0,742
0,69	0,644	0,679	0,696	0,661	0,713	0,647	0,723	0,654
0,901	0,881	0,837	0,917	0,902	0,874	0,9	0,937	0,863
0,873	0,864	0,837	0,891	0,885	0,874	0,882	0,929	0,835
0,081	0,075	0,034	0,093	0,073	0,048	0,052	0,055	0,063
0,067	0,08	0,036	0,101	0,077	0,041	0,058	0,052	0,072
358,714	350,588	352,459	336,879	333,333	405,594	379,699	312,5	350,829
228,426	218,824	213,115	195,035	304,688	143,357	255,639	253,906	256,906
57,53	32,941	32,787	60,284	23,438	52,448	30,075	15,625	46,961
98,139	80	94,262	120,567	85,938	111,888	97,744	117,188	60,773
6,768	2,353	0	10,638	10,417	0	15,038	31,25	11,05
1,692	23,529	12,295	3,546	2,604	13,986	7,519	11,719	2,762
15,228	11,765	28,689	17,73	23,438	3,497	0	7,813	16,575
28,765	11,765	8,197	7,092	31,25	13,986	26,316	27,344	11,05
241,964	192,942	254,098	255,319	200,521	290,21	225,564	261,719	204,42
130,287	108,235	98,362	127,66	187,501	87,415	109,022	117,189	104,972
87,986	120	81,968	102,837	75,521	108,392	93,986	85,937	107,735
79,526	89,412	57,377	60,284	36,458	76,923	45,113	42,969	69,061

74,45	105,882	49,18	53,191	96,354	108,392	78,947	54,688	85,635
0	47,059	8,197	31,915	49,479	83,916	7,519	11,719	8,287
13,536	2,353	8,197	0	0	3,497	37,594	15,625	13,812
8,46	4,706	12,295	0	0	0	7,519	0	0
0	2,353	0	10,638	5,208	0	0	23,438	13,812
33,841	30,588	8,197	0	28,646	6,993	3,759	11,719	46,961
2,126	2,504	2,517	2,099	2,499	2,218	2,653	2,216	2,499
2,888	3,041	3,111	2,936	3,146	2,861	3,183	2,98	3,03
0,716	2,085	0,847	0,69	0,623	1,773	1,572	1,31	1,187
384,632	317,815	342,464	411,226	290,357	352,211	276,083	380,1	318,231
566,545	587,376	585	564,238	580,482	573,5	591,991	563,727	589,597
349,061	342,364	372,29	386,8	374,682	386,879	385,78	351,35	362,263
387,853	373,055	396,52	418,475	406,187	415,945	407,44	386,284	404,06
422,185	424,218	432,21	436,238	437,083	443,867	438,407	432,143	438,704
3,861	3,747	4,72	3,323	4,62	3,37	4,603	4,062	4,645
6,728	5,901	7,795	5,969	6,029	6,682	6,931	6,622	6,838
1,522	1,459	1,533	1,693	1,402	1,599	1,54	1,348	1,544
1,807	1,331	2,084	1,642	1,459	1,804	1,712	1,797	1,63
52,198	63,019	62,146	36,986	68,429	51,938	83,717	56,31	62,716
10,428	9,749	11,326	13,105	9,326	11,936	5,254	10,297	10,507
11,734	23,113	20,517	12,195	18,802	11,444	23,377	10,311	23,671

Curriculum Vitae

Personal Details

Name and Surname: Vilém Nájemník

Place of Birth and Date: Jablonec nad Nisou-1991

E-mail: vilem.najemnik@gmail.com / vilem.najemnik@seznam.cz

Telephone: +420 608 967 062

Educational Background

Degree	Institution	Year
Bachelor	University of Hradec Králové Faculty of Education	2015

Taking Part in the Organisation of International Summits

1-) III. International Eurasian Educational Research Congress, 2016