# A CROSS-SECTIONAL EVALUATION OF SYNTACTIC COMPLEXITY AND LEXICAL DIVERSITY AS PREDICTORS OF FOREIGN LANGUAGE WRITING QUALITY: A STUDY WITH PRE-SERVICE TEACHERS OF ENGLISH

**Dissertation Study**

**Zafer SUSOY**

**Eskişehir 2022**

# A CROSS-SECTIONAL EVALUATION OF SYNTACTIC COMPLEXITY AND LEXICAL DIVERSITY AS PREDICTORS OF FOREIGN LANGUAGE WRITING QUALITY: A STUDY WITH PRE-SERVICE TEACHERS OF ENGLISH

**Zafer SUSOY**

## DISSERTATION STUDY

**Department of Foreign Language Education**

**PhD Program in English Language Teaching**

**Supervisor: Prof. Dr. Gül DURMUŞOĞLU KÖSE**

**Eskişehir**

**Anadolu University**

**Institute of Educational Sciences**

**July 2022**

**FINAL APPROVAL FOR THESIS**

## ABSTRACT

A CROSS-SECTIONAL EVALUATION OF SYNTACTIC COMPLEXITY AND LEXICAL DIVERSITY AS PREDICTORS OF FOREIGN LANGUAGE WRITING QUALITY: A STUDY WITH PRE-SERVICE TEACHERS OF ENGLISH

Zafer SUSOY

Department of Foreign Language Education
Program in English Language Teaching

Anadolu University, Graduate School of Educational Sciences, July 2022

Supervisor: Prof. Dr. Gül DURMUŞOĞLU KÖSE

The main premise of the current dissertation study is to investigate to what extent syntactic complexity (SC) and lexical diversity (LD) can predict foreign language (FL) writing quality. Thus, the study scrutinizes the relationship between SC-LD scores calculated by an automated tool called Coh-Matrix and overall writing scores assigned by human judges for the English essays of 204 pre-service teachers of English of two different curricular levels (first and fourth year students). For the qualitative data, 8 instructors who had been rating student papers for 16 years on average at the time of study were interviewed. To that end, the study adopts a sequential-explanatory mixed method research design. For the quantitative part of this paradigm, English essays were first written by two different groups of ELT majors to be processed in Coh-Matrix. These essays were also scored by two independent raters following a standardized criteria and inter-rater reliability and intra-rater reliability check was also assured. The output data provided by Coh-Matrix and the overall writing scores, then, was further analyzed through descriptive statistical tests as well as t-tests, correlational and regression analyses within and across groups. As the statistical analyses reveal, the fourth year students outperformed the first year students in word counts, writing scores and in total of 5 Coh-Matrix indices of SC and LD. Though only text length as a confounding variable significantly correlated with writing quality scores on a moderate level, the correlation analysis exposed a consistent relationship among 5 SC and LD indices. Subsequent regression analyses explained the variance in overall writing scores. The qualitative results provided insights into the overall scoring procedure of the instructors and how much capable they were in grasping and considering SC and LD in their scoring. Based on the findings of the study, certain pedagogical implications and suggestions for further research were shared.



**Keywords:** Syntactic Complexity, Lexical Diversity, Foreign Language Writing Assessment.

# ÖZET

## SÖZDİZİMSEL KARMAŞIKLIK VE SÖZCÜK ÇEŞİTLİLİĞİNİN YABANCI DİLDE YAZMA KALİTESİ GÖSTERGELERİ OLARAK ÇAPRAZ-KESİŞİMSEL İNCELENMESİ: İNGİLİZCE ÖĞRETMEN ADAYLARI İLE YAPILAN BİR ÇALIŞMA

Zafer SUSOY

Yabancı Diller Eğitimi Anabilim Dalı
İngiliz Dili Eğitimi Programı

Anadolu Üniversitesi, Eğitim Bilimleri Enstitüsü, Temmuz 2022

Danışman: Prof. Dr. Gül DURMUŞOĞLU KÖSE

Bu çalışmanın temel amacı sözdizimsel karmaşıklık ve sözcük çeşitliliğinin yabancı dildeki yazma kalitesini ne ölçüde yordadığını araştırmaktır. Bu yüzden, mevcut çalışmada İngilizce öğretmenliği bölümü birinci ve dördüncü sınıf öğrencileri olan toplam 204 İngilizce öğretmeni adayının yazılarına ait Coh-Matrix isimli bir işleme aracı tarafından sağlanan sözdizimsel karmaşıklık ve sözcük çeşitliliği puanlarının, aynı yazılara insan notlandırıcılar tarafından verilen genel değerlendirme puanları ile olan ilişkisi incelenmiştir. Çalışmanın nitel verisi ise çalışmanın yürütüldüğü zaman ortalama olarak 16 yıldır öğrenci yazılarını notlandıran 8 öğretim elemanı ile mülakat gerçekleştirilmiştir. Güdülen amaçlar gereği, bu çalışmada açıklayıcı sıralı karma araştırma yöntemi benimsenmiştir. Bu yöntemin nicel boyutu için, iki farklı sınıfta okuyan katılımcı öğrenciler tarafından Coh-Matrix'e yüklenmek üzere İngilizce kompozisyonlar yazılmıştır. Bu kompozisyonlar aynı zamanda iki bağımsız notlandırıcı tarafından standart ölçütlere dayalı olarak notlandırılmış ve notlandırıcılar arası güvenirlik ve değerlendirme geçerliliği de hesaplanmıştır. Öğrenci yazılarına ait Coh-Matrix verileri ve genel değerlendirme puanları betimleyici istatistik testlerin yanı sıra, t testleri, korelasyon ve regresyon analizleri ile katılımcı gruplar içinde ve arasında detaylı analize tabii tutulmuştur. İstatistiki analizlerin ortaya koyduğu üzere, dördüncü sınıf öğrencileri, kelime sayıları, genel değerlendirme puanları ve sözdizimsel karmaşıklık ve sözcük çeşitliliğine dair toplam 5 Coh-Matrix göstergesinde de birinci sınıfları geride bırakmıştır. Korelasyon analizinde yalnızca metin uzunluğu genel değerlendirme puanları ile orta düzeyde ve de anlamlı olarak olumlu ilişki sergilemişse de sözdizimsel karmaşıklık ve sözcük çeşitliliği göstergeleri aralarında tutarlı ilişkiler bulunmuştur. Takiben, regresyon analizi ile genel değerlendirme puanlarının varyans açıklamasına gidilmiştir. Nitel bulgular ise notlandırıcıların genel değerlendirme süreçlerine ilişkin ve notlandırmalarında sözdizimsel karmaşıklık ve sözcük çeşitliliğini ne kadar fark edip, değerlendirdiklerine ilişkin kavrayışlar sağlamıştır. Çalışmanın bulgularına dayanarak, bazı eğitsel öneriler ve ileri araştırma fikirleri öne sürülmüştür.

**Anahtar Sözcükler:** Sözdizimsel Karmaşıklık, Sözcük Çeşitliliği, Yabancı Dilde Yazma Değerlendirmesi.

## ACKNOWLEDGEMENTS

Writing this PhD dissertation took a long journey which was full of several challenges as well as great pleasures and marvelous opportunities to grow and learn. The dissertation has been made possible only throughout the valuable support of a number of people. If it had not been their encouragement, this study would not have been achievable.

First and foremost, my heart-felt thanks are for my supervisor Prof. Dr. Gül DURMUŞOĞLU KÖSE, whom I deeply appreciate and respect. Hereby, I would like to express my endless gratitude for her enlightening, ever-lasting and unforgettable support at each and every phase of my student life. I owe sincere and most special thanks for her leading comments and inspiring attitudes which amended lots of obstacles on the way. Without her invaluable guidance and continuous assistance, this study would not be possible.

Besides, I would like to express my deep gratitude to the members of my dissertation committee Prof. Dr. Özgür YILDIRIM and Prof. Dr. Işıl KABAKÇI YURDAKUL who both tremendously contributed to my dissertation by their vast knowledge and experience particularly about data collection and analyses procedures. Their genuine support and generous contributions widened my research outlook from various perspectives. I also owe my thanks to other members of my defense committee Prof. Dr. İlknur SAVAŞKAN and Assoc. Prof. Dr. Hüseyin KAFES for their helpful guidance and constructive feedback.

I would like to thank Prof. Dr. Zülal BALPINAR and Bethany Kaitlyn Lucille Broberg who diligently scored our written data and provided beneficial insights. I also would like to acknowledge that I was offered a great deal of help, at the stages of data collection and interviewing, by the staff of English Language Teaching Department of Anadolu University. Besides, my sincere thanks also go for all the students who volunteered to participate in my research. Their volunteering to write essays formed the core of this study.

Last but not least, I would like to specially thank my beloved parents and dear friends who have been always there to support and encourage me. I am very grateful for their great patience and love.

Zafer SUSOY

Eskişehir 2022

20/07/2022

## ETİK İLKE VE KURALLARA UYGUNLUK BEYANNAMESİ

Bu tezin bana ait, özgün bir çalışma olduğunu; çalışmamın hazırlık, veri toplama, analiz ve bilgilerin sunumu olmak üzere tüm aşamalarında bilimsel etik ilke ve kurallara uygun davrandığımı; bu çalışma kapsamında elde edilen tüm veri ve bilgiler için kaynak gösterdiğimi ve bu kaynaklara kaynakçada yer verdiğimi; bu çalışmanın Anadolu Üniversitesi tarafından kullanılan "bilimsel intihal tespit programı"yla tarandığını ve hiçbir şekilde "intihal içermediğini" beyan ederim. Herhangi bir zamanda, çalışmamla ilgili yaptığım bu beyana aykırı bir durumun saptanması durumunda, ortaya çıkacak tüm ahlaki ve hukuki sonuçları kabul ettiğimi bildiririm.

Zafer SUSOY

20/07/2022

**STATEMENT OF COMPLIANCE WITH ETHICAL PRINCIPLES AND RULES**

I hereby truthfully declare that this thesis is an original work prepared by me; that I have behaved in accordance with the scientific ethical principles and rules throughout the stages of preparation, data collection, analysis and presentation of my work; that I have cited the sources of all the data and information that could be obtained within the scope of this study, and included these sources in the references section; and that this study has been scanned for plagiarism with "scientific plagiarism detection program" used by Anadolu University, and that "it does not have any plagiarism" whatsoever. I also declare that, if a case contrary to my declaration is detected in my work at any time, I hereby express my consent to all the ethical and legal consequences that are involved.

Zafer SUSOY

## CONTENTS

**Page**

**TITLE PAGE** ..... **i**
**FINAL APPROVAL FOR THESIS** ..... **ii**
**ABSTRACT** ..... **iii**
**ÖZET** ..... **iv**
**ACKNOWLEDGEMENTS** ..... **v**
**ETİK İLKE VE KURALLARA UYGUNLUK BEYANNAMESİ** ..... **vi**
**STATEMENT OF COMPLIANCE WITH ETHICAL PRINCIPLES AND RULES** ..... **vii**
**CONTENTS** ..... **viii**
**THE LIST OF TABLES** ..... **xi**
**THE LIST OF FIGURES** ..... **xiii**
**THE LIST OF ABBREVATIONS** ..... **xiv**

**CHAPTER 1**

**1. INTRODUCTION** ..... **1**
- **1.1. Background to the Study** ..... **1**
- **1.2. Research Purpose and Questions** ..... **10**
- **1.3. Significance of the Sxtudy** ..... **10**

**CHAPTER 2**

**2. REVIEW OF LITERATURE** ..... **13**
- **2.1. Introduction to the Chapter** ..... **13**
- **2.2. Writing Quality in Second Language** ..... **15**
- **2.3. Defining Syntactic Complexity and Browsing the Contents** ..... **17**
- **2.4. Defining Lexical Diversity and Browsing the Contents** ..... **20**
- **2.5. The Relationship between Syntactic Complexity and Foreign Language Writing Quality** ..... **23**
- **2.6. The Relationship between Lexical Diversity and Foreign Language Writing Quality** ..... **33**
- **2.7. Measurement of Syntactic Complexity and Lexical Diversity: Problems and Approaches** ..... **41**
  - **2.7.1. Measuring syntactic complexity in L2 with a critique of traditional complexity measures** ..... **43**
  - **2.7.2. Measuring lexical diversity in L2 with a critique of traditional complexity measures** ..... **46**

**Page**

**CHAPTER 3**


**3. METHODOLOGY ..... 50**

**3.1. Participants and research context ..... 50**

**3.2. Data Collection ..... 52**

**3.3. Data Analysis ..... 54**

**3.3.1. Essay quality ratings ..... 57**

**3.3.2. Inter-Rater and intra-rater reliability ..... 57**

**3.3.3. Coh-metrix measures ..... 59**


**CHAPTER 4**


**4. RESULTS ..... 62**

**4.1. Introduction to the Chapter ..... 62**

**4.2. Syntactic Complexity, Lexical Diversity and Writing Quality Scores of Participating Students ..... 62**

**4.2.1. Syntactic complexity, lexical diversity and writing quality scores of 1st year students' compositions ..... 62**

**4.2.2. Syntactic complexity, lexical diversity and writing quality scores of 4tht year students' compositions ..... 65**

**4.3. Curricular Level Differences Among the Investigated Variables ..... 68**

**4.3.1. Differences in text length ..... 68**

**4.3.2. Differences in writing scores ..... 69**

**4.3.3. Differences in syntactic complexity ..... 69**

**4.3.4. Differences in lexical diversity ..... 71**

**4.4. Inter-correlations between variables and variances explained ..... 72**

**4.4.1. Correlations of syntactic complexity, lexical diversity and text length with writing quality ..... 72**

**4.4.2. Variance in the writing quality scores explained ..... 75**

**4.5. Unfolding Syntactic Complexity: Embodying It as a Construct ..... 76**

**4.5.1. The role of syntactic complexity in scoring the students' essays ..... 80**

**4.5.2. An analogy between complexity and diversity ..... 82**

**4.6. Unfolding Lexical Diversity: Embodying It as a Construct ..... 83**

**4.6.1. The role of lexical diversity in scoring the students' essays ..... 86**

**4.7. General Outlook of the Scoring Procedure ..... 88**

**4.7.1. How do our instructors start scoring and continue? ..... 88**

**Page**

**4.7.2. Content and organizational patterns overshadow syntactic complexity and lexical diversity ..... 89**

**CHAPTER 5**

**5. DISCUSSION, CONCLUSION AND IMPLICATIONS ..... 92**

**5.1. Summary of the Study ..... 92**

**5.2. Uncovering Syntactic Complexity and Lexical Diversity ..... 94**

**5.3. Issues of Syntactic Complexity with Regards to Scoring, Scorers and Indices ..... 95**

**5.4. Issues of Lexical Diversity with Regards to Scoring, Scorers and Indices .. ..... 97**

**5.5. Issues Regarding Text Length in FL Writing Scoring ..... 99**

**5.6. Conclusions and Pedagogical Implications ..... 101**

**5.7. Suggestions for Further Research ..... 103**

**REFERENCES ..... 105**

**APPENDICES ..... 111**

**CURRICULUM VITAE ..... 152**

## THE LIST OF TABLES

**Page**

**Table 2.1.** Questions and Answers about Coh-metrix..... 48

**Table 3.1.** Distribution of participants ..... 51

**Table 3.2. I**nstructors and their year of experience in scoring..... 51

**Table 3.3.** Overview of research methodology ..... 55

**Table 3.4.** Results of Pearson correlations coefficients between two raters across two scoring procedures (for 1st year students' scores)..... 58

**Table 3.5.** Results of Pearson correlations coefficients between two raters across two scoring procedures (for 4th year students' scores) ..... 58

**Table 3.6.** Questions and Answers about Coh-metrix..... 59

**Table 4.1.** Descriptives of 1st year students' essays ..... 63

**Table 4.2.** Frequency and percentage rates of word ranges in 1st year students' essays64

**Table 4.3.** Descriptives of 1st year students' writing quality scores..... 64

**Table 4.4.** Descriptive values of lexical diversity of 1st year students' essays ..... 65

**Table 4.5.** Descriptive values of syntactic complexity of 1st year students' essays..... 65

**Table 4.6.** Descriptives of 4th year students' essays..... 66

**Table 4.7.** Frequency and percentage rates of word ranges in 4th year students' writing..... 66

**Table 4.8.** Descriptives of 4st year students' writing quality scores..... 67

**Table 4.9.** Descriptive statistics of lexical diversity and its two indices in 4th year students' essays..... 67

**Table 4.10.** Descriptive statistics of syntactic complexity and its three indices in 4th year students' essays..... 68

**Table 4.11.** A numerical comparison of 1st and 4th year students' essays..... 68

**Table 4.12.** Results of independent samples t-test for writing quality scores by curricular level ..... 69

**Table 4.13.** Results of independent samples t-test for 'number of words coming before main verb' by curricular level..... 70

**Page**

**Table 4.14.** Results of independent samples t-test for 'syntactic similarity' by curricular level ..... 70

**Table 4.15.** Results of independent samples t-test for mean number of modifiers per NP by curricular level ..... 71

**Table 4.16.** Results of independent samples t-test for 'MTLD' by curricular level ..... 72

**Table 4.17.** Results of independent samples t-test for 'VocD' by curricular level ..... 72

**Table 4.18.** Results of Pearson correlations coefficients among seven variables ..... 73

**Table 4.19.** Hierarchical Multiple Regression Analysis with a Three-Layered Model (Dependent Variable; Writing Quality Overall Scores) ..... 75

**Table 4.20.** Examples that point syntactic complexity ..... 79

**Table 4.21.** Examples that point lexical diversity ..... 85

## THE LIST OF FIGURES

**Page**

**Figure 1.1.** A layout of language sophistication ..... 3

**Figure 3.1.** Screenshot for the categorical themes ..... 56

**Figure 3.2.** A Coh-metrix screen shot displaying syntactic complexity indices..... 61

**Figure 3.3.** A Coh-metrix screen shot displaying lexical diversity indices ..... 61

**Figure 4.1.** The Thematic Display of 'Unfolding Syntactic Complexity' ..... 77

**Figure 4.2.** The Thematic Display of 'Unfolding Lexical Diversity' ..... 84

**Figure 4.3**. The Thematic Display of 'The General Outlook of Scoring Process' ..... 89

**Figure 4.4**. The Thematic Display of 'Qualitative Results Summary' ..... 91

**Figure 5.1.** The Thematic Display of Research Input..... 93

**Figure 5.2.** The Thematic Display of Research Output ..... 93

**Figure 5.3.** Major Differences in the Nature of Comparison..... 99

## THE LIST OF ABBREVATIONS

**EFL** : English as a Foreign Language

**ELT** : English Language Teaching

**FL** : Foreign Language

**INST** : Instructor

**LD** : Lexical Diversity

**MELAB** : Michigan English Language Assessment Battery

**MTLD** : Measure of Textual Lexical Diversity

**NWD** : Number of Different Words

**R** : Rater

**SC** : Syntactic Complexity

**TTR** : Type Token Ratio

**TWE** : Test of Written English

# CHAPTER 1

## 1. INTRODUCTION

### 1.1. Background to the Study

Learning to write requires a process which forms a very basic necessity in a student's whole educational life. Students who lack the ability of adequate written expression may face underachievement at school and may not even graduate. Majority of similar threats result from high stake tests which demands higher order writing skills in first language (L1) (Jenkins, Johnson and Hileman, 2004). Academic achievement in primary and higher education in L1 largely depends on developed writing skills as well as further professional enterprises (Geiser and Studley, 2001). These developed L1 writing skills have been associated with sophisticated linguistic characteristics and an elaboration of language (McNamara, Crossley and McCarthy, 2010). Highly qualified foreign language (FL) writing works have also been shown to include linguistic characteristics related to more elaborated language (McNamara, Crossley and McCarthy, 2009). The sophistication of language used in written FL production which contributes to writing quality ratings were mostly associated with syntactic complexity and lexical diversity in a great bulk of previous research (Crowhurst, 1980; Engber, 1995; Ellis and Yuan, 2004, Crossley and McNamara, 2010, 2011; Lu, 2011; Mazgutova and Kormos, 2015). Although a more complex syntax and a more diverse range of vocabulary in written production seems to hinder text comprehension from readers' perspective, these two constructs, on the other hand, also correlates with the overall FL written quality ratings assigned by human raters (Mcnamara et. al, 2009).

Syntactic complexity and lexical diversity is considered to echo the writers' skilfulness, competence and even socioeconomic status (Ransdell and Wengelin, 2003). Likewise, more competent writers are expected to write in a more complex and diverse manner both syntactically and lexically. Consequently, it is likely that essays in L2 with higher scores are portrayed as having more complex sentences and with deployment of more diverse words. Syntactic complexity, traditionally speaking, is regarded as the sphere and elaborateness of syntactic structures in language production, and the extent of refinement of such structures prevailing in the language produced. Ortega (2015) posits that syntactic complexity is an important construct that points out the extent to which a language learner can use the language more tactfully and in a more sophisticated manner.

Syntactic complexity construct can allow the language users to successfully accomplish the communicative purposes (Ortega, 2015).

The previous research conducted both in L1 and L2 has somehow related the syntactic complexity to writing quality (Crowhurst, 1983; Wolfe-Quintero, Inagaki & Kim, 1998; Ortega, 2003; Lu, 2010; 2011; Mazgutova and Kormos, 2015; Vyatkina, Hirschmann and Golcher, 2015). These studies, however, have not yet been able to explain a complete picture depicting the relationship between syntactic complexity and writing quality partly due to the ambiguous nature of this relationship (Beers & Nagy, 2009). Additionally, very few studies in the past (Beers and Naggy,2009; Donovan and Smolkin, 2006;), took the possible effect of genre into account in terms of this complicated relationship.

Writing has been defined as a multidimensional process (National Council of Teachers of English, 2004), and one of these dimensions is the ability to construct mature sentences. NCTE (2004) points out that writers should be conscious and skilful about the rhetorical resources they use to leave a favourable impression on the readers. One of the twelve NCTE guidelines clearly mentions the sentence construction and variety in language structures. Moreover, certain commonly resorted rubrics to assess writing also refer to skilful sentence construction as a predictor of overall text quality (Northwest Regional Educational Laboratory, 2004).

The research concerning the syntactic complexity and its relationship with writing quality has been on the agenda for several decades (Crowhurst, 1983; Wolfe-Quintero, Inagaki & Kim, 1998; Ortega, 2003; Lu, 2010; 2011). Nonetheless, the results and conclusions put forward by the bulk of this research are far from being consistent, possessing a number of problems such as identical proficiency profiles of learners and insufficient sampling (Ortega, 2003). Since a large variety of syntactic complexity metrics were used in these studies, it is not likely to generalize that these results and correlations are identical across different metrics (Norris & Ortega, 2009).

Equal worries are also pertinent for lexical diversity as well although it has been viewed among the most crucial independent aspects of lexical proficiency (Crossley, Salsbury, McNamara, and Jarvis, 2011) and a powerful indicator of L2 writing quality (Engber, 1995; Laufer and Nation, 1995; Olinghouse and Wilson, 2012). Additionally, there are various explicit references to the importance of lexical diversity in the rubrics used to assess L2 writing in several internationally acclaimed English language tests such

as Test of Written English offered by ETS, IELTS and Michigan English Language Assessment Battery (See Appendix 1).

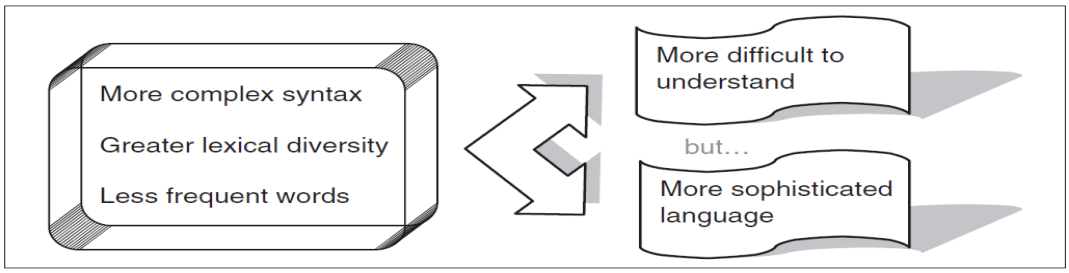



**Figure 1.1.** *A layout of language sophistication (Adopted from McNamara et.al, 2009, p.64)*

The case illustrated by Figure 1.1, though verified in L1 studies (Donovan and Smolkin, 2006; Beers and Naggy, 2009; Crossley and Mcnamara, 2009, 2011), has also been extensively studied in English as a Second Language (ESL) studies (Crowhurst, 1980; Engber, 1995; Ellis and Yuan, 2004). However, to our best knowledge there is has been a research gap to fill in the foreign language learning contexts. We acknowledge that there is a pressing need to study syntactic complexity and lexical diversity in EFL pre-service teachers' writings. The importance of complexity metric selection in the relationship of syntactic complexity and lexical diversity with writing quality will be presented in following chapters in more detail.

As already stated, there is inconsistency among the results of previous research partly due to a lack of uniformity in the complexity measures and insufficient sampling and partly due to lack of a clear definition of the complexity construct. Majority of the studies examining 'complexity' define the construct either in ambiguous or broad terms. Several definitions of FL complexity found in the literature also include the related concepts and they are as follows:

> (1) "[complexity is the] use of more challenging and difficult language … Complexity is the extent to which learners produce elaborated language" (R. Ellis & Barkhuizen 2005, p. 139)
>
> (2) "Grammatical and lexical complexity mean that a wide variety of both basic and sophisticated structures and words are available to the learner" (Wolfe-Quintero, Inagaki, & Kim 1998, p. 69)
>
> (3) "Complexity refers to … the complexity of the underlying interlanguage system developed" (Skehan 2003, p. 8).

Bulte and Housen (2014, p.46) stated that FL complexity has been mostly associated with concepts such as "better, more proficient, more advanced, more mature,

and more developed". Bulte and Housen (2012), in an attempt to define L2 complexity and refine its multidimensionality, wrote that linguistic complexity can be observed in the language behaviour in several syntactical and lexical dimensions (e.g., variation in the embededness and combination of clauses, using different verb form sor a wider range of vocabulary). Bulte and Housen (2012, p.25) indicated that L2 complexity has been handled in two basic views: global complexity and local complexity. The former refers to the learners' overall L2 system and its changing nature, while the latter refers to the specific items and structures. In our study, we follow this distinction and adopt the 'global' view to define complexity and diversity:

> "Global or system complexity refers to the degree of elaboration, the size, breadth, width, or richness of the learner's L2 system or 'repertoire', that is, to the number, range, variety or diversity of different structures and items that he knows or uses: whether he masters a small or a wide range of different words or different grammatical structures, whether he controls all or only a fraction of the sound system of the L2, and so forth." (p. 25).

Therefore, the degree of variety of both basic and elaborate structures and words in FL writing has been the key to our understanding of syntactic complexity and lexical diversity. Ortega (2003) recognizes that syntactic complexity, viewed as the extent of elaborateness of syntax in language production, has been recognized as a significant variable in second or foreign language writing research.

In the literature a large number of syntactic complexity measure has been offered. There have been considerable research attempts for decades to find and validate a reliable measure (Wolfe-Quintero et.al, 1998; Ortega,2003). Most of this research has focused on specifying which measure(s) could be objectively used either to track learners' writing development or to assess proficiency. However, there is discrepancy in the results of these studied due to variability and inconsistency among the complexity measures, data size, and the language tasks and genres operationalized in the data collection procedures (Wolfe-Quintero et. al, 1998; Ortega, 2003; Lu, 2010; Lu,2011).

Not only the inconsistency of measures used but also the small number of them and the limited data size hinders the pooling the results of previous studies. For example, Ortega's comprehensive review (2003) covered twenty-five cross-sectional studies which examined the syntactic complexity development in foreign or second language writing and only four studies included in this review used four to five different measures. The remaining twenty-one studies resorted to only three measures. Likewise, the average number of the written data collected in these studies was less than 100. Similar problems

prevailed in more recent work as well. For example, in one study three hundred learner e-mails were syntactically analyzed using only clauses per-T Unit measure (Stockwell and Harrington, 2003). In another study, Ellis and Yuan (2004) similarly used only clauses per T-Unit measure to analyze fifty-two narratives and Beers and Nagy (2009) used mean length of clause in addition to T-Unit ratio to analyze forty-one essays in two different genres. More recently, however, syntactic complexity and lexical diversity research has started to benefit from a computational tool named Coh-Metrix (Graesser, McNamara and Kulikowich, 2011) to offer subtler predictors. The syntactic and lexical indices provided by this automated tool have been validated by several recent studies (McNamara et al, 2010; Crossley and McNamara, 2011,2012; Crossley et. al., 2011).

Many distinct aspects of lexical knowledge have also been studied in writing research since there are numerous ways to characterize the lexical knowledge; however, as probably one of the most acknowledged lexical constructs, lexical diversity means the breath of words appearing in a text (Olinghouse and Wilson, 2013). Lexical diversity has been viewed as an indispensable index of learners' writing quality (Laufer and Nation, 1995) and their generic linguistic competence. The 'global' or 'system' complexity view, which was put forward by Bulte and Housen (2012, p. 25) and which emphazised the breadth of the learner repertoire, is also valid in the definition of lexical diversity; that is, lexical diversity in our studt refers to "variety, richness and diversity of different items" that a learner possesses. In other words, in the root of our understanding lexical diversity lies the variety and/or richness of vocabulary items in the FL writings of our learners.

In the literature, there have been robust efforts to devise a measure to calculate lexical diversity as well, which is not affected by text length. For example, some earlier studies only calculated the instances of unique words (Grobe, 1981) by tokens (i.e., all instances of words). In an effort to sophisticate this most commonly used tool and to eliminate the text length effect, corrected type-token ratio (CTTR; Carroll, 1964 as cited in Olinghouse and Wilson, 2013, p.48) was devised, however, CTTR has also recently been reported to have strong correlations with text length variable (McCarthy and Jarvis, 2007; Olinghouse and Leaird, 2009). To refer to this problem, with the advent of recent computational algorithms, refined and reliable measuring approaches to lexical diversity have also been developed. These lexical measures documented by Coh-Metrix included the Measure of Textual Lexical Diversity (MTLD) and VocabD, which exclude the text length as a confounding variable (McCarthy and Jarvis, 2010).

Syntactic complexity has been extensively receiving researchers' attention in the area of second language writing as well as other variables such as accuracy, fluency and lexical proficiency (Ai and Lu, 2013, p. 251). The related research has been in pursuit of reliable syntactic complexity measures to gauge the developmental path through which learners evolve in L2 writing (Wolfe-Quintero et.al, 1998; Ortega, 2003; Lu, 2011). An extensive bulk of this research has been related to the relationship between syntactic complexity in L2 writing and learners' proficiency levels. There have been several studies which examined the longitudinal changes occurring in L2 writing over time (Larsen-Freman, 1978; Ferris, 1994; Ortega, 2000; Stockwell and Harrington, 2003; Mazgutova and Kormos, 2015; Vyatkina, Hirschmann and Golcher, 2015). For example, Stockwell and Harrington (2003) reported a significant increase in writings of Japanese learners after a 5-week of e-mail exchanging with Japanese native speakers. 212 essays were written in an early study by Larsen-Freeman (1978) and syntactically analyzed at five different ESL proficiency levels and found out that syntactic complexity significantly differed at each proficiency level. Likewise, Ferris (1994) reported that syntactic complexity as a variable significantly differed between low and advanced levels of ESL. Lu (2011) also found that learners at different college levels differentiated in terms of syntactic complexity displayed in their writing.

The claim that syntactic complexity in L2 writing develops over time with more instruction and exposure has been questioned by two recent studies. Mazgutova and Kormos (2015) studied the development of syntactic complexity within a 4-week of intensive academic writing course. The researchers studied with two groups of learners; Group 1 consisted of more advanced and older ESL learners and Group 2 of less proficient and younger learners. Mazgutova and Kormos (2015) reported that less proficient and younger group of learners showed a noticeable growth over 4-week of instruction, while the more advanced and older students showed smaller growth; a result which was attributed by the researchers to different proficiency levels.

In a similar attempt, Vyatkina et al (2015) gauged the development of syntactic complexity in beginning L2 German learners' writings over a 2-year college training. The researchers showed a stable and increasing development towards a more diverse syntactic complexity usage in their writing at lexical, phrasal, clausal and sentential levels. To conclude, syntactic development can be observed over time either within a 4-week course with specific focus on writing or within a 2-year college education which did not

particularly direct the learners to more complex syntax use. It is, therefore, possible to conclude that syntactically more refined, complex and varied written production can result from the ability to benefit from a wider sphere of syntactic sources available in language, which differentiates between high and low proficiency levels.

The ability to skilfully benefit from each and every linguistic means offered by the grammar allows the writers to fulfill the intended communicative goals more successfully, thus resulting in quality in their written productions (Ortega, 2015, p.83). There are many studies which examine syntactic complexity to anticipate essay quality both in L1 and L2 (McNamara, Crossley and McCarthy, 2010; Crossley, Weston, Sullivan and McNamara, 2011). Though not adequate alone, one of the conditions essential for the production of high quality written texts in L2 is listed as the ability to construct complex sentences and each essential condition- including syntactic complexity- is of paramount significance in terms of writing quality (Beers and Nagy, 2009, p.187). Syntactic complexity is an important index of wiring quality in that complex ideas and more propositions can be pressed in complex sentences through the usage of such structures as "nominalizations, attributive adjectives, and prepositional phrases" (Beers and Nagy, 2009, p. 187).

Corrier (1996, p.321) wrote that through syntactically complicated sentences could ideas be connected to one another appropriately and, that relations among ideas, thus, can be soundly established, which contributes to writing quality. For example, to establish a compare and contrast relationship, one needs more subordinate structures like '*Although X, Y'*. Since in order for the writers to attach more attention to rhetorical and other textual issues in their writings, they need to have the ability to comfortably produce complex structures, which would otherwise undermine the global text quality. (Deane, 2013, p.13). Likewise, studies show that students' writings rated as highly qualified contain more finite verbs and a larger number of words appearing before the main verb implying that high quality texts have more complex syntactic features (McNamara et al. 2010; Crossley et al, 2011).

McNamara et. al (2010) exploited a computational tool named Coh-Metrix to uncover the linguistic characteristics of essays which were rated as high and low quality by human raters. The essays in this study were written by freshman college students. The results indicated that high quality essays containing a more sophisticated language resulted from more infrequent words and more complex syntactic structures. McNamara

et. al (2010) concluded that the three most powerful predictors of essay quality were respectively syntactic complexity, lexical variation and word count. In a similar strand, Crossley and McNamara (2011) also confirmed the relationship between syntactic complexity and essay quality in L2 through significant positive correlations. They found that writing highly qualified essays in L2 was more related to a more complex syntax and lexical diversity even more than textual cohesion, which was claimed to ease the readability of the text.

Lexical proficiency, although still lacking a clear definition as a cognitive construct, is mostly related to multiple dimensions of vocabulary knowledge such as breadth of knowledge (i.e. the size of one's knowledge), depth of knowledge (i.e. the extent of one's knowledge), and the access dimension (i.e. the ease and speed of word processing Meara, 2005). Lexical proficiency, with its multidimensional and underexploited nature, is crucial construct for L2 learners as it has already caught much scholar attention. For example, Leki and Carson (1994) found that what ESL learners in a EAP course uniformly pronounced as the most desirable skill to possess was L2 lexical knowledge. According to Ellis (1995), L2 communication breakdowns are mainly caused by lexical errors. L2 lexical proficiency is also considered essential for academic success in L2 (Daller, van Houıt and Treffers-Daller, 2003).

Lexical proficiency, as a broad term, needs clarification and specification of its individual constructs. Among attempts to characterize lexical proficiency as a construct, Crossley et. al (2011) investigated the variance in human ratings of lexical proficiency using automated lexical indices in both L1 and L2 learner writing samples. They found that 'lexical diversity' along with 'word frequency' far better predicted the human ratings than the other related dimensions under investigation. Crossley et. al (2011), therefore, uncovered that 'lexical diversity' is among the most important individual aspects of lexical proficiency.

Lexical diversity has been viewed as an indispensable index of learners' writing quality (Laufer and Nation, 1995) and their generic linguistic competence. To this end, there are several internationally acclaimed language tests and computerized language evaluation systems which consider the lexical diversity in a similar vein. IELTS, for example, uses the term 'lexical resource' to refer to 'the range of vocabulary the candidate has used' in the assessment of candidates written and spoken replies (IELTS, Handbook, 2007). Likewise, another internationally acclaimed language test TOEFL iBT explicitly

supports the claim that lexical diversity can predict the writing quality since 'appropriate and precise use of grammar and vocabulary' has been used as a criterion to evaluate the written compositions (TOEFL iBT Scores, 2005). In a similar manner, 'a wide range of appropriately used vocabulary' is written among the criteria to achieve high scores in written compositions of Micheagen English Language Assessment Battery (MELAB, Technical Manuel, 2003).

Several studies in the literature assert that lexical diversity as a construct can gauge the learners' overall language development and differs across language proficiency levels. The difference in terms of lexical diversity in writing from a proficiency point of view can be seen in studies comparing native and non-native writings. For example, Linnarud (1986) found that native speakers of English produced essays richer in lexical items in terms of diversity than Swedish learners of English. From a similar strand, Harley and King (1989) compared the compositions written by native French students and international French immersion students. They used frequency and number of verb types, lexical errors and variety as lexical diversity measures and indicated that on all measures native speakers displayed a greater diversity than non-native speakers. The difference between learners' ability to display lexical diversity in written discourse is not only apparent in native/non-native comparisons but also across different proficiency levels. For example, Laufer and Nation (1995) compared the written products of learners divided into three proficiency levels with reference to lexical frequency and diversity. They revealed that the most proficient learners could produce the most lexically diverse texts with a bigger number of infrequent words.

There are also some studies which lent empirical support to the claim that lexical diversity is closely associated and positively correlated with holistic scoring of writing quality in L2. For example, Engber (1995) uncovered that lexical diversity measured by metrics of lexical variation and density significantly and positively correlated with his overall writing quality scoring of ESL students. In a recent study, Olinghouse and Wison (2012) studied the role of lexical knowledge with its various dimensions, namely; diversity, maturity, academic words and register. The study examined the relationship of lexical knowledge to human quality ratings assigned to EFL learners' written works across three genres. Though not observed in all genres, lexical diversity was found to be the unique index of quality in story text.

## 1.2. Research Purpose and Questions

Ortega (2012) asserts that majority of L2 complexity research in the past were conducted with at least three primary purpose: "a) to gauge proficiency, b) to describe performance and c) to benchmark development" (p.128). Likewise, the premise of the present study is three-fold. First, it aims to investigate the relationship with syntactic complexity, lexical diversity and FL writing quality scores assigned by human judges by uncovering the correlations and the extent to which syntactic complexity and lexical diversity account for the variance in FL writing quality. Second, the present study intends to conduct a cross-sectional evaluation of the hypothesized relationship between syntactic complexity, lexical diversity and writing quality in learners' FL writings from a developmental and/or proficiency point of view (Crossley and McNamara, 2104; Ortega, 2012; Ortega, 2015). In other words, we aim to find out if there is any significant change in FL writings of our first and fourth year students measured by syntactic complexity and lexical diversity indices. The current study's third premise is to explore the perceptions of instructors who have been scoring undergraduates' academic writing in an ELT department. Thus, we aim to see the extent to which these instructors are aware of SC and LD in their scoring procedures. Therefore, based on the research purposes, our study aims to answer the following research questions;

1) What are the syntactic complexity, lexical diversity, text length and writing quality scores of participating students?
2) Is there a difference between syntactic complexity, lexical diversity, text length and writing quality scores of learners at different curricular levels?
3) What is the relationship between syntactic complexity, lexical diversity, text length and L2 writing quality scores assigned by human raters?
4) To what extent are syntactic complexity and lexical diversity engaged in the perception of writing instructors who evaluate undergraduates' academic writings?

## 1.3. Significance of the Sxtudy

It has not been easy to compound reliable results in the area of syntactic complexity, lexical diversity and L2 writing research due to the labour-intense nature of manual analysis and lack of computational tools which automate a reliable analysis. Consequently, most of previous research yielded inconsistent results as a result of varying

and small amount of data and measure (Wolfe-Quintero et. al, 1998; Ortega, 2003). Therefore, Lu (2010) calls for the adaptation of a reliable computational approach to analyze larger sets of written data by applying multiple and most commonly proposed complexity measures in the literature to paint a reliable and complete picture of the role played by syntactic complexity and lexical diversity in L2 writing. To our existing knowledge, majority of the related research by combining multiple measures whereby a computational tool has been carried out in first language writing studies (see for comprehensive reviews Crowhurts, 1983; Wolfe-Quintero et.al, 1998; Ortega, 2003). Apart from few pioneering exceptions (Crossley and McNamara, 2011, 2012; Ai and Lu, 2013; Bi and Jiang, 2020), large sums of data written in L2 have not yet been systematically analyzed in a similar vein.

Conceptualization of proficiency in developmental studies greatly varies, obstructing direct comparisons, from program level (Larsen-Freeman, 1978; Maamuujav, Olson and Chang, 2021) to month-long changes in an intensive writing course (Mazgutova and Kormos, 2015; Casal nad Lu, 2021) and to longitudinal tracing of the writing development through syntactic complexity and linguistic sophistication indices (Kyle, Crossley and Verspoor, 2021). It is, therefore, highly important to decide how to characterize the proficiency in such studies. In a book-length, comprehensive research synthesis, Wolfe-Quintero et. al (1998) wrote that "program level may be the most valid developmentally" (p.9) to differentiate different proficiency levels. Following this assertion and previous research, (Linnarud, 1986; Harley and King, 1995; Mazgutova and Kormos, 2015; Vyatkina, 2015; Treffers-Daller, Parslow and Williams, 2016; Maamuujav et al, 2021) we, in our study, regard the program level, i.e., first and fourth year students in a four-year degree of ELT program as the proficiency index.

The short review so far indicates that the relationship between syntactic complexity, lexical diversity and L2 writing has a complicated nature lacking even patterns. The results and conclusions put forward by previous research are far from being consistent, possessing a number of problems such as a limited sum of data, identical proficiency profiles of learners and insufficient sampling (Ortega, 2003). However, the current study hypothesizes that the incorporation of different measures at a large data set with different proficiency levels using a reliable text processing tool may suggest a more intelligible view of the relationship of syntactic complexity, lexical diversity with FL writing. Furthermore, our participating students, different from the reviewed studies, are EFL pre-

service teachers who are supposed to teach English language and FL writing at various levels. They, before certified as EFL teachers, experience an extensive four-year degree program. Therefore, another premise of our study is to find out whether there is any difference in EFL pre-service teachers' FL writings in terms of SC, LD and overall FL writing scores. We also set out to explore the human perception which is set to work while scoring undergraduate students' academic papers. Our study's another aim is to find out how much room in the perception of human scorers is taken by SC and LD when it comes to scoring. Put differently, we are curious about the extent to which the human scorers are aware of and consider SC and LD in their scoring.

# CHAPTER 2

## 2. REVIEW OF LITERATURE

### 2.1. Introduction to the Chapter

Learning to write requires a process which forms a very basic necessity in a student's whole educational life. Students who lack the ability of adequate written expression may face underachievement at school and may not even graduate. Majority of similar threats result from high stake tests which demands higher order writing skills in first language (L1) (Jenkins, Johnson and Hileman, 2004). Academic achievement in primary and higher education in L1 largely depends on developed writing skills as well as further professional enterprises (Geiser and Studley, 2001). These developed L1 writing skills have been associated with sophisticated linguistic characteristics and an elaboration of language (McNamara, Crossley and McCarthy, 2010). Highly qualified foreign language (FL) writing works have also been shown to include linguistic characteristics related to more elaborated language (McNamara, Crossley and McCarthy, 2009). The sophistication of language used in written FL production which contributes to writing quality ratings were mostly associated with syntactic complexity and lexical diversity in a great bulk of previous research (Crowhurst, 1980; Engber, 1995; Ellis and Yuan, 2004, Crossley and McNamara, 2010, 2011; Lu, 2011; Mazgutova and Kormos, 2015). Although a more complex syntax and a more diverse range of vocabulary in written production seems to hinder text comprehension from readers' perspective, these two constructs, on the other hand, also correlates with the overall FL written quality ratings assigned by human raters (Mcnamara et. al, 2009).

Syntactic complexity and lexical diversity is considered to echo the writers' skilfulness, competence and even socioeconomic status (Ransdell and Wengelin, 2003). Likewise, more competent writers are expected to write in a more complex and diverse manner both syntactically and lexically. Consequently, it is likely that essays in L2 with higher scores are portrayed as having more complex sentences and with deployment of more diverse words. Syntactic complexity, traditionally speaking, is regarded as the sphere and elaborateness of syntactic structures in language production, and the extent of refinement of such structures prevailing in the language produced. Ortega (2015) posits that syntactic complexity is an important construct that points out the extent to which a language learner can use the language more tactfully and in a more sophisticated manner.

Syntactic complexity construct can allow the language users to successfully accomplish the communicative purposes (Ortega, 2015).

The previous research conducted both in L1 and L2 has somehow related the syntactic complexity to writing quality (Crowhurst, 1983; Wolfe-Quintero, Inagaki & Kim, 1998; Ortega, 2003; Lu, 2010; 2011; Mazgutova and Kormos, 2015; Vyatkina, Hirschmann and Golcher, 2015). These studies, however, have not yet been able to explain a complete picture depicting the relationship between syntactic complexity and writing quality partly due to the ambiguous nature of this relationship (Beers & Nagy, 2009). Additionally, very few studies in the past (Beers and Naggy,2009; Donovan and Smolkin, 2006;), took the possible effect of genre into account in terms of this complicated relationship.

Writing has been defined as a multidimensional process (National Council of Teachers of English, 2004), and one of these dimensions is the ability to construct mature sentences. NCTE (2004) points out that writers should be conscious and skilful about the rhetorical resources they use to leave a favourable impression on the readers. One of the twelve NCTE guidelines clearly mentions the sentence construction and variety in language structures. Moreover, certain commonly resorted rubrics to assess writing also refer to skilful sentence construction as a predictor of overall text quality (Northwest Regional Educational Laboratory, 2004).

The research concerning the syntactic complexity and its relationship with writing quality has been on the agenda for several decades (Crowhurst, 1983; Wolfe-Quintero, Inagaki & Kim, 1998; Ortega, 2003; Lu, 2010; 2011). Nonetheless, the results and conclusions put forward by the bulk of this research are far from being consistent, possessing a number of problems such as identical proficiency profiles of learners and insufficient sampling (Ortega, 2003). Since a large variety of syntactic complexity metrics were used in these studies, it is not likely to generalize that these results and correlations are identical across different metrics (Norris & Ortega, 2009).

Equal worries are also pertinent for lexical diversity as well although it has been viewed among the most crucial independent aspects of lexical proficiency (Crossley, Salsbury, McNamara, and Jarvis, 2011) and a powerful indicator of L2 writing quality (Engber, 1995; Laufer and Nation, 1995; Olinghouse and Wilson, 2012). Additionally, there are various explicit references to the importance of lexical diversity in the rubrics used to assess L2 writing in several internationally acclaimed English language tests such

as TOEFL iBT, IELTS and Michigan English Language Assessment Battery. This chapter will present an extensive review of literature about syntactic complexity, lexical diversity and their relationship with foreign language writing quality as well as the issues covering the measurement of these main constructs.

## 2.2. Writing Quality in Second Language

Second language writing research historically divided the academic written texts into two; those texts written in circles of academia in the form of published articles and dissertations, and those written by university students in the form of homework and essays tests (Hinkel, 2002). Thus, it is likely to propose that student essays as a common form of academic text, are worthy of textual and linguistic analysis to discover their definite content and form since these kind of written academic texts prevail in number. Therefore, bearing the ability to generate well-built and written essays of high quality is among the necessities of a successful school life and a professional career (Geiser and Stundley, 2001). From the standardized language tests of proficiency, to various placement tests and personal certification, student essays are supposed to be of high quality from several textual and linguistic respects. To exemplify these requirements for text quality, we can cite syntactic accuracy of both sentences and phrases, the degree of variety in grammatical structures, the proper selection of vocabulary as well as adequate organizational support and appropriate rhetorical structure (ETS, 2000).

Likewise, the assessment criteria employed by Michigan English Language Assessment Battery (MELAB) impose similar textual and linguistic expectations to those of ETSs'. The MELAB specifically signifies that in order for an essay to be highly rated, the essay topic should be "richly and fully developed" and it should pose "a flexible use of a wide range of syntactic structures" (MELAB Technical Manual, 1994, p.7) as well as a large diversity and a proper use of words. Apart from linguistic characteristics that ETS and MELAB writing assessment criteria openly mention, they also make clear references to the place of cohesion and coherence of a text with "control of connections" (p.7).

Global impressions of human scorers who assess student writing also have parallels with the quality criteria expressed in the body of standardized tests. That is to say, that textual organization and topic development as well as grammatical accuracy along with variety and lexical width and propriety affect the perceptions of human raters, thus their

scores (Santos, 1988; Davidson, 1991). It is notable to note here that even standardized tests applications employ trained human raters to make holistic judgements about student writing quality. Crossley et al. (2010) wrote that there are three basic ways to quantify writing quality; primary trait, analytic and holistic. While primary trait scoring refers to the "rhetorical situations (e.g. the purpose, audience and assignment)" (p.284), analytical scoring distinctly emphasizes the single units that are related to quality (Crossley et. al., 2010). Holistic scoring, on the other hand, is realized through an impressionist view of human raters involves and according to Haut (1990), holistic scoring became the ordinary means of writing quality assessment as it equates analytic scoring well and it is cost-effective. To specify, holistic scoring is defined by Cohen (1994, p.314) as the evaluation "that is based on a single, integrated score of writing behavior". There are several reasons why holistic scoring of writing quality has been largely exploited in related writing research. Among its advantages is that its being effective as Weigle (2002) puts it. That is, holistic scoring takes place within a short time without much effort to read thoroughly several times a text to touch on individual aspects separately like in the analytical approach. Additionally, holistic scoring focuses on the strengths of students; that is the writing quality is determined based on what students can achieve rather than their weak points (White, 1984). White (1984) also pointed out that "holistic scoring is able to achieve acceptably high reliability" (p.403). Six different approaches were proposed by White (1984) to ensure the reliability of holistic scoring;

- *Controlled essay reading*: a group of raters come together and rate the papers by the same amount of time and labor allocated like working in a workshop
- *Scoring criteria guide*: scoring through a pre-conditioned sets of descriptive statements for different aspects
- *Sample papers*: these papers can be called the practiced examples of scoring guides. Raters, before starting a reliable scoring, practice their scoring criteria and reach an agreement about what and what not to expect from papers.
- *Checks on the reading in progress:* Raters are grouped around a table and a chief reader goes around each rater to check the scoring process of each rater and sometimes stops the progress and make a mini-workshop.
- *Multiple independent scoring:* Two raters simultaneously score a paper independent of each other, and then compare their scores and opinions and resolve any possible discrepancies.

- *Evaluation and record keeping:* Consistency rates of scorers with others and the scoring criteria are recorded to determine the most reliable and consistent raters.

Although there are clear advantages of using a holistic approach to assessing L1 and L2 writing quality as documented above, Weigle (2002), in her book devoted to writing assessment, warns that a single score would not be enough to display manifold aspects of writing proficiency and quality, additionally, raters engaging in holistic scoring do not assign same scores based on the same criteria. Weigle (2002) wrote, for example, "a certain script might be given a 4 on a holistic scale by one rater because of its rhetorical features (content, organization, development), while another rater might give the same script a 4 because of its linguistic features (control of grammar and vocabulary)" (p. 114).

For the very reason, it is of vital importance to discover which features (rhetorical or linguistic) contribute more to the writing quality rated by human judges. We also deem it necessary for our research purposes to discover whether and to what extent these features are related to L2 writing quality. At this point of argumentation, it does not seem unwise at all to characterize writing quality as successful writing to which high scores were given by raters in line with the assessment criteria followed which certainly involved specific linguistic and/or rhetorical features. Hence, we understand that writing quality is neither a single, simple nor a concrete construct but rather complex and visible through its expected linguistic and textual indicators. A crucial question arises at this specific point: what are these linguistic and features that make a text qualified?

### 2.3. Defining Syntactic Complexity and Browsing the Contents

Complexity and complex schemes, in various fields, have been a popular phenomenon which is under scientific investigation since 1990s and among those fields is Second Language Acquisition (SLA) as put forward by Bulte and Housen (2014). There is, yet, inconsistency among the results of previous research;

- partly due to a lack of uniformity in the complexity measures
- insufficient sampling
- and more importantly due to lack of a clear definition of the complexity construct.

Problems and approaches to solutions on the measurement of complexity and other methodological concerns in the studies will be presented in next sections. In this section,

we focus more on the definition and contents of complexity. As stated, there is still no unanimity reached on a fixed definition of complexity, however, there are foundation piers upon which several definitions have been established so far. Though, in this section, we first try to clarify the term 'complexity' and then, finalize 'syntactic complexity', it is important to note that some definitions found in the literature also include the lexical dimension of complexity like that of Lennon's. Lexical dimension of complexity will be discussed in the following section. Among different definitions of complexity are "using a wide range of structures and vocabulary" (Lennon, 1990, p. 390), "progressively more elaborate language and a greater variety of syntactic patterning" (Forster and Skehan, 1996, p.303). As can be understood from the core of definitions reported, a sense of elaborateness and a width of linguistic devices arise as two key terms to understand 'complexity'. These two key terms greatly contribute to our understanding of complexity as well in this study. Likewise, Bulte and Housen (2012, p.22) wrote that complexity is associated with "(1) the number and the nature of the discrete components that the entity consists of, and (2) the number and the nature of the relationship between the constituent components". What is implied by 'the number' in their definitions can be equal to width and range of structures in other definitions, while 'the nature' can be equal to the sophistication and/or elaborateness of structures.

We should remember a notable classification made to specify the complexity construct. In a vigorous attempt to operationalize 'complexity' as s construct, Bulte and Housen (2012, p 23-24) outlined a basic distinction between "relative" and "absolute" complexity. At the very onset of this distinction, we want to highlight that relative complexity is also called as cognitive complexity or simply difficulty and it is presented rather as a subjective issue since it is prone to the effects of language features and learner-dependent factors. In Bulte and Housen (2012) difficulty is considered subjective since it, in fact, is the extent or amount of cognition someone has to employ to work out a task. This can clearly change, for example, according to several learner-related factors such as motivation, anxiety, memory or learners' language proficiency. For the very reason, what is perceived as cognitively more complex by someone can be well found less or more cognitively complex by someone else depending on these factors. Apart from these learner-related factors, relative complexity is also related with language system or particular linguistic features, that is, the cognitive burden that linguistic items possess. The order of acquisition observed in some structures might provide examples to make

this point clear. For instance, active structures are found easier to be processed and earlier to be produced by language learners.

As for 'absolute complexity', it tackles language complexity in an objective and quantitative manner. It describes the language complexity, as written earlier, in the form of "(1) the number and the nature of the discrete components that the entity consists of, and (2) the number and the nature of the relationship between the constituent components" (Bulte and Housen, 2012, p.22). 'Absolute complexity' suits more to the traditional definitions of complexity reported earlier in this section in that absolute complexity also highlights the importance of width and elaborateness of structures which are accepted as two fundamental piers of complexity in our study. Therefore, upon adopting an absolute complexity approach we can say that difficulty is left out of the scope of this study. It is vital to note, still, that difficulty is only one of the embedded aspects that might contribute to complexity or not. Bulte and Housen (2102) reminds us that there is not a necessarily positive correlation between cognitive complexity and absolute complexity; namely between difficulty and complexity for short. Therefore, difficulty and complexity are two different constructs that were separated on a theoretical ground. It is, additionally, important to remember that difficulty does not always result in complexity, but "rather reflects it" (Rescher, 1998, p.17). In other words, it might be sometimes easier to produce structurally more complex items than others. Pallotti (2015) exemplifies this issue through an analogy and went on writing that "A Sudoku with 18 digits is structurally less complex but cognitively more complex, or difficult than one with 25 digits" (p.119).

An alternative definition was proposed by Wolfe-Quintero et al. (1998), putting 'grammatical' and 'complexity' together in a way that "grammatical complexity means that a wide range of both basic and sophisticated structures are available" (p.69). Wolfe-Quintero et. al's definition (1998) is an attention grabbing one in that it includes a large array of basic forms too, besides sophisticated structures. We need to make an important note right here to remind that grammatical complexity and syntactic complexity substitute each other and are used reciprocally in the literature. To provide more clarity to syntactic complexity, Pallotti (2015, p.120) brought a distinction between syntactic and stylistic complexity. Syntactic complexity, according to this view, is associated with rules that are compulsory to be followed to construct syntactically correct sentences. Stylistic

complexity, on the other hand, has more to do with the extent of individual or even culturally imposed rhetorical choices.

Bulte and Housen (2014, p.46) stated that FL complexity has been mostly associated with concepts such as "better, more proficient, more advanced, more mature, and more developed". Bulte and Housen (2012), in an attempt to define L2 complexity and refine its multidimensionality, wrote that linguistic complexity can be observed in the language behaviour in several syntactical and lexical dimensions (e.g., variation in the embeddedness and combination of clauses, using different verb forms or a wider range of vocabulary). Bulte and Housen (2012, p.25) indicated that L2 complexity has been handled in two basic views: global complexity and local complexity. The former refers to the learners' overall L2 system and its changing nature, while the latter refers to the specific items and structures. In our study, we follow this distinction and adopt the 'global' view to define complexity and diversity:

> "Global or system complexity refers to the degree of elaboration, the size, breadth, width, or richness of the learner's L2 system or 'repertoire', that is, to the number, range, variety or diversity of different structures and items that he knows or uses: whether he masters a small or a wide range of different words or different grammatical structures, whether he controls all or only a fraction of the sound system of the L2, and so forth." (p. 25).

Therefore, the degree of variety of both basic and elaborate structures in FL writing has been the key to our understanding of syntactic complexity. Ortega (2003) recognizes that syntactic complexity, viewed as the extent of elaborateness of syntax in language production, has been recognized as a significant variable in second or foreign language writing research. So far, we tried to distill cognitive complexity (difficulty) and stylistic issues out of syntactic complexity to highlight the intermediate boundaries. We also stressed different typological approaches (relative vs. absolute, global vs. local) to define complexity and stated what our understanding and adoption.

### 2.4. Defining Lexical Diversity and Browsing the Contents

Lexical proficiency, although still lacking a clear definition as a cognitive construct, is mostly related to multiple dimensions of vocabulary knowledge such as breadth of knowledge (i.e. the size of one's knowledge), depth of knowledge (i.e. the extent of one's knowledge), and the access dimension (i.e. the ease and speed of word processing Meara, 2005). Lexical proficiency, with its multidimensional and underexploited nature, is crucial construct for L2 learners as it has already caught much scholar attention. For

example, Leki and Carson (1994) found that what ESL learners in a EAP course uniformly pronounced as the most desirable skill to possess was L2 lexical knowledge. According to Ellis (1995), L2 communication breakdowns are mainly caused by lexical errors. L2 lexical proficiency is also considered essential for academic success in L2 (Daller, van Hout and Treffers-Daller, 2003).

Lexical proficiency, as a broad term, needs clarification and specification of its individual constructs. Among attempts to characterize lexical proficiency as a construct, Crossley et. al (2011) investigated the variance in human ratings of lexical proficiency using automated lexical indices in both L1 and L2 learner writing samples. They found that 'lexical diversity' along with 'word frequency' far better predicted the human ratings than the other related dimensions under investigation. Crossley et. al (2011), therefore, uncovered that 'lexical diversity' is among the most important individual aspects of lexical proficiency.

Lexical diversity has been viewed as an indispensable index of learners' writing quality (Laufer and Nation, 1995) and their generic linguistic competence. To this end, there are several internationally acclaimed language tests and computerized language evaluation systems which consider the lexical diversity in a similar vein. IELTS, for example, uses the term 'lexical resource' to refer to 'the range of vocabulary the candidate has used' in the assessment of candidates' written and spoken replies (IELTS, Handbook, 2007). Likewise, another internationally acclaimed language test TOEFL iBT explicitly supports the claim that lexical diversity can predict the writing quality since 'appropriate and precise use of grammar and vocabulary' has been used as a criterion to evaluate the written compositions (TOEFL iBT Scores, 2005). In a similar manner, 'a wide range of appropriately used vocabulary' is written among the criteria to achieve high scores in written compositions of Michigan English Language Assessment Battery (MELAB, Technical Manuel, 2003).

As can be understood, lexical array of one's linguistic production, or the range of words used in language production can be named as lexical diversity. The term 'diversity' is, after all, associated with "the range of vocabulary and avoidance of repetition" (Malvern, Richards, Chiepere and Duran, 2004, p.3). Lexical diversity relates to the size and amount of different word usage in any text. That is, less repetition of the same word along with a varied profile of lexical usage leads to lexical diversity. In the literature, different terms have been interchangeably used to refer to lexical diversity; among them

were ‘lexical variation’ (Engber, 1995), ‘lexical density’ (O’Loughlin,1995), “a combination of lexical variation and lexical sophistication” (Laufer, 2003, p.24), and ‘lexical richness’ as coined by Daller, von Haut and Treffers-Daller, 2003). Different ways to operationalize and, then, to quantify lexical diversity in previous studies resulted in a confusion about the construct, which made it difficult to synthesize the related research for bigger and clearer pictures.

In order to exceed the barrier made up of incomprehensibility of the term ‘lexical diversity’, Malvern et. al., (2004) outlined and divided the measures into basics. According to these researchers, lexical diversity can be possibly measured in two broad ways; 1) traditional approaches to measurement, 2) mathematical approaches to measurement. Traditional approaches were token, types and type-token ratio (TTR) that is better and more advanced than individual token and type analyses. TTR, however, is not without flaws in that it mainly relies on the text length and sample size (Malvern et. al., 2004). Mathematical approaches to lexical diversity measurement, on the other hand, have been found free from text length effect (Malvern et. al., 2004, McCarthy and Jarvis, 2007). These validated and novel approaches are Vocab-D or D measure and Measure of Textual Lexical Diversity (MTLD) which is automated by Coh-Metrix- an automated computational tool for advanced textual analysis. In the following section that is dedicated to the measurement of syntactic complexity and lexical diversity, these measures and the justification of using an automated analysis tool will be presented in detail. For now, we suffice it to say that in our study we adopted a lexical diversity view and definition that is free from text length and measured differently from traditional methods.

Many distinct aspects of lexical knowledge have been studied in writing research since there are numerous ways to characterize the lexical knowledge; however, as probably one of the most acknowledged lexical constructs, lexical diversity means the breath of words appearing in a text (Olinghouse and Wilson, 2013). Lexical diversity has been viewed as a strong indicator of learners’ writing quality (Laufer and Nation, 1995) and their generic linguistic competence. The ‘global’ or ‘system’ complexity view, which was put forward by Bulte and Housen (2012, p. 25) and which emphasised the breadth of the learner repertoire, is also valid in the definition of lexical diversity; that is, lexical diversity in our study refers to “variety, richness and diversity of different items” that a learner possesses. In other words, in the root of our understanding of lexical diversity lies

the variety and/or richness of vocabulary items in the FL writings of our learners in line with the diversity measures we used.

## 2.5. The Relationship between Syntactic Complexity and Foreign Language Writing Quality

Syntactic complexity has almost become one of the regular gauges for judgement of language performance and proficiency in the area of Second Language Acquisition and Applied Linguistics (Norris and Ortega, 2009). As also earlier pointed out by Larsen-Freeman (1978), syntactic complexity – as an index of linguistic development- is likely to increase by the time learners attain more command of their language production. Since some complex ideas and the connections between them could only be exchanged via particular complex structures, it is possible to view syntactic complexity as one of trivets of writing proficiency (Beers and Nagy, 2009).

Studies have been scrutinizing syntactic complexity and its relationships with various SLA and Applied Linguistics-related concepts. Prior to syntactic complexity and L2 writing quality relationship which is the main focus of this section, we deem it notable to shortly document several studies investigated 'these other associations' of syntactic complexity in the field of SLA. To begin with, for example, different planning times allocated to students to construct their compositions led to differences in the syntactic complexity of their L2 writing as reported by Ellis and Yuan (2004). Likewise, Lu (2011) also showed a strong correlation between the amount of time given and the syntactic complexity measured in students' writings. Task conditions such as audience and topic have also been subject to investigation in the relationship of syntactic complexity with L2 writing. For example, Li (2000) examined the linguistic features of 132 e-mails written by ESL students. The study showed that interaction with audience was a task condition that led to more syntactic complexity than without-audience condition. Besides, e-mails tended to have higher levels of syntactic complexity when students had a freedom of topic and content choice.

The relationship between syntactic complexity and genre has also been examined and confirmed. Way, Joiner and Seaman (2000), in a study to see whether and to what extent three genres associate with syntactic complexity, confronted with more syntactic complexity in expository writings of learners of French as a foreign language in contrast with descriptive and narratives. A similar comparison of genre effect was conducted by

Lu (2011) with college level ESL learners and argumentation was found to lead to more syntactic complexity than narrative writing. Each genre, as socially occurring language patterns, fulfill a distinct social function and by doing so, they draw upon different language constructions (Halliday & Hassan, 1985). The genres vary mainly in two aspects; first, the linguistic characteristics vary at the micro-level; second, the global organization of the text structures vary at macro-level (Donovan & Smolkin, 2006). Generating qualified texts across genres, thus, is likely to associate with genre specific syntactic structures that facilitate the realization of communicative functions of the genre. We suffice it to say this much about genres since our study does not aim to compare different genres in terms of syntactic complexity.

Syntactic complexity has been extensively receiving researchers' attention in the area of second language writing as well as other variables such as accuracy, fluency and lexical proficiency (Ai and Lu, 2013, p. 251). The related research has been in pursuit of reliable syntactic complexity measures to gauge the developmental path through which learners evolve in L2 writing (Wolfe-Quintero et.al, 1998; Ortega, 2003; Lu, 2011). An extensive bulk of this research has been related to the relationship between syntactic complexity in L2 writing and learners' proficiency levels. There have been several studies which examined the longitudinal or short-term changes occurring in L2 writing over time (Larsen-Freman, 1978; Ferris, 1994; Ortega, 2000; Stockwell and Harrington, 2003; Stockwell, 2005; Mazgutova and Kormos, 2015; Vyatkina, Hirschmann and Golcher, 2015). For example, Stockwell and Harrington (2003) reported a significant increase in writings of Japanese learners after a 5-week of e-mail exchanging with Japanese native speakers. Following Stockwell and Harrington (2003), Stockwell (2005) designed a very similar study this once to find out whether email exchanges between NNS students for a five-week period would result in any change in syntactic complexity of a group of students studying different universities in Japan. In this study, syntactic complexity was traced by mean length of T-unit. As results indicated, there was not a significant change in the syntactic complexity values of emails unlike Stockwell and Harrington (2003) where the email exchange took place between NS and NNS students. Stockwell (2005) attributed the different results of the two studies to the fact that emails exchanged between NNS students were much shorter in length compared to those exchanged between NS and NNSs. We should also highlight that mean length of T-unit as a complexity index is very much dependent on the length of written production, which might be the real cause of

inconsistency between results. 212 essays were written in an early study by Larsen-Freeman (1978) and syntactically analyzed at five different ESL proficiency levels and found out that syntactic complexity significantly differed at each proficiency level. Likewise, Ferris (1994) reported that syntactic complexity as a variable significantly differed between low and advanced levels of ESL. Lu (2011) also found that learners at different college levels differentiated in terms of syntactic complexity in their writing.

The claim that syntactic complexity in L2 writing develops over time with more instruction and exposure has been questioned by two recent studies. Mazgutova and Kormos (2015) studied the development of syntactic complexity within a 4-week of intensive academic writing course. The researchers studied with two groups of learners; Group 1 consisted of more advanced and older ESL learners and Group 2 of less proficient and younger learners. Mazgutova and Kormos (2015) reported that less proficient and younger group of learners showed a noticeable growth over 4-week of instruction, while the more advanced and older students showed smaller growth; a result which was attributed by the researchers to different proficiency levels.

In a similar attempt, Vyatkina et al (2015) gauged the development of syntactic complexity in beginning L2 German learners' writings over a 2-year college training. The researchers showed a stable and increasing development towards a more diverse syntactic complexity usage in their writing at lexical, phrasal, clausal and sentential levels. To conclude, syntactic development can be observed over time either within a 4-week course with specific focus on writing or within a 2-year college education which did not particularly direct the learners to more complex syntax use. It is, therefore, possible to conclude that syntactically more refined, complex and varied written production can result from the ability to benefit from a wider sphere of syntactic sources available in language, which differentiates between high and low proficiency levels.

160 ESL compositions belonging ot low and advanced groups of learners were examine din a study by Ferris (1994). Al the compositions were holistically scored by three independent scorers on a 1-10 scale. The study aimed to see whether and to what extent the syntactic complexity indices could predict proficiency groups and to find out to see how well these indices predicted the holistic scores assigned to student writings. Ferris (1994) conducted correlation coefficients and as well as stepwise multiple regression analysis to achieve the research purposes. The results announced that more advanced learners used more syntactic devices such as a common use of "specific lexical

categories (e.g., emphatics and hedges) and a greater production of difficult syntactic constructions (e.g., stative forms, relative clauses)“ (p. 417). As for the variance in the holistic scores given, number of words best predicted the scores with a variance percentage of 37.6. Becker (2010) also studied with ESL students I a university’s Intensive English Program (IEP) and divided the students into three proficiency groups based on their IEP placement test scores. Becker (2010) analyzed a total of ten discourse characteristics to uncover whether and which of them could distinguish between different proficiency levels. The results revealed that syntactic complexity indices such as word per T-unit, clauses per T-unit could well differentiate across three proficiency levels. Though other textual features (e.g., cohesion, coherence) were also investigated in Becker’s study (2010), they did not seem to be correlating with proficiency levels at all, which is warranting why syntactic complexity as a text variable should be investigated further in L2 writing. However, one important caution we must take from this study is about the text length. Lower level students wrote longer sentences and texts than advanced groups as advanced learners could well condense their ideas into smaller units and could write in short, but effectively as well. Johansson and Geisler (2011) studied the syntactic complexity development across different curricular levels of EFL students. They examined the junior and senior Swedish high school students’ writings in English. Number of T-units and number of error-free T-units and the proporsion of subordinate clauses were the investigated indicators of syntactic complexity. The results demonstrated that senior students displayed longer and more complex T-units as they use a higher number of relative clauses.

Lu (2010) developed an automated tool (see: http://nlp.stanford.edu/software/lex-parser.shtml) which involved fourteen different syntactic complexity measures that were widely used and recommended in the related literature. The measures used in the study fell into five types; length of production unit with three measures, sentence complexity with one measure, subordination with four measures, coordination with three measures, and particular structures with three measures. Lu (2010) exploited Written English Corpus of Chinese Learners (WECCL) as the source of college level written data. The data for Lu’s study was comprised of 3,554 essays with a mean word count of 315 and with a standard deviation of 87. The written data was analyzed in the automated tool of fourteen measures which was specifically developed for this sort of research purposes. The analysis results showed that six out of fourteen measures significantly discriminated

three proficiency levels in terms of syntactic complexity and demonstrated a linear increasing across proficiency groups. It is safe to understand from the conclusions of this study that measures which detected a linear increasing in the complexity of students' writings were the measures targeting phrasal level complexity such as coordinate phrasal and complex nominal. In sum, the two most important assumptions of this study were confirmed; first, the automated complexity analyzer managed to produce reliable results; second, syntactic complexity of Chinese EFL learners displayed a developing character across different proficiency levels.

Lu (2010) devised an automated tool combining fourteen syntactic complexity measures and Lu (2011) later refined ten measures as best indicators of syntactic complexity out of originally proposed fourteen measures. Hence, Mancilla, Polat and Akcay (2015) conducted a study using the same automated analyzer tool employing ten syntactic complexity measures. Their aim was to compare NS and NNS students' written responses on an asynchronous learning environment in terms of syntactic complexity. Mancilla et. al. (2015) compared the groups on the basis of gender and language proficiency in terms of syntactic complexity. 102 NNS ans 142 NS of English participated in the study, 169 of whom were female, while 74 of whom were male. The data for the study was piled from 486 discussion board postings between years of 2009 and 2013. According to Second Language Syntactic Complexity Analyzer and between-groups ANOVA comparisons, NS students turned out to be using more subordination as an index of syntactic complexity, on the other hand, NNS students attended more to coordination and phrasal sophistication. Additionally, their data showed no difference between high and low proficiency levels of NNS students, and only minor difference between NS and low level NNS students only regarding to the amount of subordination. The study, all in all, highlighted the firm claim that NS and NNS student writings were not to compete in terms of syntactic complexity and could not easily be on exactly same levels of complexity. However, what Mancilla et. al. (2015) concludes is that NNS students could well display a native like syntactic complexity in their writing by the time they attend to college level study.

Ai and Lu (2013), in a similar strand of methodology but with a different research purpose, conducted another study to examine syntactic complexity of non-native speakers (NNS) and native speakers' writing (NS) to see whether these two groups' writings differed in selected ten syntactic complexity measures, and if so, to what extent it would.

The ten syntactic complexity measures used in this study were about length of production unit, amount of subordination, amount of coordination and the degree of phrasal sophistication. The researchers divided the written data into three groups of 200; NNS-low proficiency, NNS-high proficiency and NS learners' writings. The result of the study displayed statistically significant differences in the four investigated areas of syntactic complexity. Namely, NS produced longer clauses, longer sentences and T-units, bigger amounts of subordination and more complex nominals than two groups of NNS's. More proficient NNS group, however, could better approximate the NS group in terms of length of production and phrasal sophistication by comparison to less proficient NNS students. The study concludes that the results determined the areas of syntactic complexity where NNS students were weak and called for proper pedagogical interventions to fill in this gap. In a seminal work of research synthesis, Ortega (2003) concluded that in syntactic complexity and writing relationship research which was conducted in ESL settings, participants generated more complex writings compared to those in the studies conducted in FL instructional settings. Ortega (2003) attributed this difference of complexity level to the basic differences between ESL and EFL instructional settings. Namely, in contrast with ESL settings, FL instructional settings do not allow learners to achieve a fast process of development. Beside, learners at FL instructional settings may not have as long a history of learning a second language as ESL learners and may only draw near to the ultimate language acquisition.

The ability to skilfully benefit from each and every linguistic means offered by the grammar allows the writers to fulfill the intended communicative goals more successfully, thus resulting in quality in their written productions (Ortega, 2015, p.83). There are many studies which examine syntactic complexity to anticipate essay quality both in L1 and L2 (McNamara, Crossley and McCarthy, 2010; Crossley, Weston, Sullivan and McNamara, 2011). Though not adequate alone, one of the conditions essential for the production of high quality written texts in L2 is listed as the ability to construct complex sentences and each essential condition- including syntactic complexity- is of paramount significance in terms of writing quality (Beers and Nagy, 2009, p.187). Syntactic complexity is an important index of wiring quality in that complex ideas and more propositions can be pressed in complex sentences through the usage of such structures as "nominalizations, attributive adjectives, and prepositional phrases" (Beers and Nagy, 2009, p. 187).

Bulte and Housen (2014) carried out a recent study. The study aimed to investigate whether and to what extent objective syntactic complexity measures correlate with subjective human ratings of L2 writing quality. Their study also aimed to discover which syntactic and lexical complexity measure(s) could best predict writing quality ratings of human raters. Forty-five randomly selected essays from a learner corpus were used as the data for the study. The study used ten measures of syntactic complexity and three measures of lexical complexity. The essays were evaluated for quality whereby a rating scale for 'language use' and 'vocabulary' as well as for content, organization and mechanics. As in line with the first aim of the study, the findings suggested that, though there occurred changes in the syntactic complexity level of the student writings, lexical profile of students' writings did not become "more lexically diverse, rich or sophisticated in the course of observed period." (p.53). However, significant changes in the syntactic complexity of student writings did occur; significant growth of sentential and phrasal production units and of clause coordination. Secondly, the study found strong correlations between L2 writing quality and "lexical richness, clausal subordination and mean lengths of clauses sentences and T-Units" (Bulte and Housen, 2014, p.54). Thus, the researchers provided empirical support to the claims that highly qualified writings which were scored high by human scorers contained a wide range of different words and longer units of sentences and phrases. The study also concluded that use of simple sentence constructions was perceived as indicators of poor writing quality.

Corrier (1996, p.321) wrote that through syntactically complicated sentences could ideas be connected to one another appropriately and, that relations among ideas, thus, can be soundly established, which contributes to writing quality. For example, to establish a compare and contrast relationship, one needs more subordinate structures like 'Although X, Y'. Since in order for the writers to attach more attention to rhetorical and other textual issues in their writings, they need to have the ability to comfortably produce complex structures, which would otherwise undermine the global text quality. (Deane, 2013, p.13). Likewise, studies show that students' writings rated as highly qualified contain more finite verbs and a larger number of words appearing before the main verb implying that high quality texts have more complex syntactic features (McNamara et al. 2010; Crossley and Weston et al, 2011).

McNamara et. al (2010) exploited a computational tool named Coh-Metrix to uncover the linguistic characteristics of essays which were rated as high and low quality

by human raters. The essays in this study were written by freshman college students. The results indicated that high quality essays containing a more sophisticated language resulted from more infrequent words and more complex syntactic structures. McNamara et. al (2010) concluded that the three most powerful predictors of essay quality were respectively syntactic complexity, lexical variation and word count. As already stated in McNamara et. al (2010), writers start producing more elaborate language which is both syntactically complex and lexically diverse as they progress across proficiency levels. Crossley et. al (2011) also hypothesized the same claim that syntactic complexity and lexical diversity increased as the proficiency level increased and conducted a study with the support of a computational Coh-metrix. The researchers made use of essays written by NS students of English from 9$^{th}$, 11$^{th}$ high school grades and 1$^{st}$ year university grade. The researchers also aimed to trail whether holistic writing quality scores correlated with the investigated linguistic variables; they, namely are; syntactic complexity (e.g., the number of modifiers per NP) and cohesion (e.g., word overlap and connectives). The results suggested that syntactic complexity increased as the grade level increased, however, cohesion decreased as the grade level increased. The findings so far, in Crossley et. al (2011), supported the claims that linguistic sophistication, which can well be expressed in the form of syntactic complexity- as an indispensable ingredient-, is acquired at later stages of language development over time and across grade levels. This study also showed that linguistically more advanced students used less cohesive devices. The use of less cohesive devices was attributed to the employment of more syntactically complex constructions that might serve a similar function as connectives. First year university students in this study wrote the essays which got the highest scores (M=3.75 SD=0.92) while 9$^{th}$ graders' essays were scored the lowest (M=1.65, SD=0.76). As the 9$^{th}$ graders' essays also had the fewest number of modifiers per noun phrase (as an index of syntactic complexity) and 1$^{st}$ year university students writings displayed the largest number of the same index, we can say that the study found a positive correlation between writing quality scores and syntactic complexity values. In a similar strand, Crossley and McNamara (2011) also confirmed the relationship between syntactic complexity and essay quality in L2 through significant positive correlations. They found that writing highly qualified essays in L2 was more related to a more complex syntax and lexical diversity even more than textual cohesion, which was claimed to ease the readability of the text.

In another study, similarly, Guo, Crossley and McNamara (2013) aimed to evaluate the extent to which linguistic features in both integrated and independent writing tasks were prognostic of human raters' scores. The linguistic features that were taken into quest in this study are syntactic complexity, lexical sophistication and text cohesion, expressed by a number of measures provided by Coh-metrix used to analyze the writings. 240 essays, comprising of both integrative (e.g., using reading and/or listening materials as a stimuli) and independent writing tasks (e.g., timed, impromptu writing), were provided to the researchers by the TOEFL iBT admiration. The essays were rated by two experienced ETS raters through a holistic rubric. Regression analysis yielded that text length was the strongest predictive of essay quality ratings with 26.4% of the variance of human scores. Participle verbs, accepted as an index of syntactic complexity, came second in predictive power of human scores. As another index of syntactic complexity, use of 3rd person singular verbs was the forth strongest predictive measure which correlated with human scores. On the other hand, the proportion of verbs in the base form, as a syntactic feature, negatively correlated with human scores. Essays with more verbs in base form got lower scores. The researchers wrote that this was because majority of the students did not conjugate the verbs correctly, which diminished the accuracy in the essays. The results we reported so far in this study were about integrative writings. As for independent writings, once again, text length was found to be the strongest predictive element of human scores followed by average syllables per word and past participle verbs as indices of syntactic complexity. As can be seen, regardless of the writing task, syntactic complexity is at the top of predictive linguistic features that predict writing quality scores.

Though syntactic complexity has mostly been found related in writing quality scores assigned by human judges in both L1 and L2 writings, there are some scarce studies where syntactic complexity as a separate construct did not correlate with writing quality scores. It is important to note that even in these studies linguistic features (apart from syntactic complexity) almost always positively correlated with writing scores and accounted for the majority of the variance in the writing scores. An example to these studies is Crossley and McNamara (2012). They carried out this study to predict L2 writing scores via linguistic characteristics appearing in student writings. The researchers particularly addressed "language features related to cohesion (i.e. the use of connectives and wordlaps) and linguistic sophistication (i.e. lexical difficulty and syntactic complexity)" (p.116). 1.200 essays were collected from Hong Kong high school students.

Text length was controlled as a confounding variable with a minimum word count of 500. Essays were analyzed by means of related measures provided by Coh-metrix. Syntactic complexity was measured by three Coh-metrix indices; a) mean number of words coming before the main verb, b) the number of high level constituents per word and c) syntactic similarity at phrasal level and the amount of parts of speech. According to results, in contrast to numerous past studies, syntactic complexity did not correlate significantly with essay scores. Lexical diversity (measured as D variable in Coh-metrix), however, operationalized as the other aspect of linguistic sophistication, explained 18% of the variance alone in writing scores of students. The other investigated textual variables in the study related to cohesion were found to be negatively correlating with essay scores, which means that highly qualified essays contained less cohesion, and more linguistic sophistication. Thus, Crossley and McNamara (2012) concluded that essays written by more proficient students tended to be linguistically more sophisticated, but less cohesive at the same time. The linguistic sophistication found in more proficient students' writings mainly came from lexical aspect of linguistic sophistication, rather than being syntactic. The researchers, still reminded that more proficient students might presume that their audience would not need much connectives for successful comprehension, whereas low-proficiency students who wot lower writing scores might need more to rely on cohesion devices to get their ideas across. Park (2017) recently completed a doctoral dissertation which was to investigate syntactic complexity as an indicator of second language writing development and quality. Specifically, the researcher aimed to add a diversity dimension to syntactic complexity by means of type/token frequency of different verb-argument constructions (VACs). Thus, this study asked whether syntactic complexity of Korean EFL students' linguistic proficiency and writing scores could be predicted by 14 syntactic complexity measures as offered by Syntactic Complexity Analyzer (Lu, 2010) and a newly introduced syntactic diversity measure (VACs). Data for the study came from 390 Korean EFL learners. Collected data was, then, analyzed in the automated analyzing tool Syntactic Complexity Analyzer. VACs were analyzed by a concordance and part-of-speech tagging tools. Participants proficiency levels were determined by a cloze test developed, piloted and validated specifically by the researcher. Each students essay was independently rated by a group of seven raters by using an analytic rating score. The findings revealed that, firstly; all the syntactic complexity values went up as proficiency levels went up, yielding a positive correlation with language proficiency. Secondly; the

study found all 16 (i.e. 14 elaboration indices and two diversity indices) significantly and positively correlated with writing scores of human raters. VAC types demonstrated the strongest correlation, among other syntactic complexity indices, with writing scores, which meant that the presence of various verb-argument structures were related to better writing quality. VACs were followed by text length-related variables of syntactic complexity (i.e. mean length of clausal, sentential and T-Units). As one may remember, Bulte and Housen (2014) also stated that clausal and sentential length of production units could well predict essay quality ratings.

## 2.6. The Relationship between Lexical Diversity and Foreign Language Writing Quality

Several studies in the literature assert that lexical diversity as a construct can gauge the learners' overall language development and differs across language proficiency levels (Laufer and Nation, 1995; Olinghouse and Leaird, 2009; Bulte and Housen, 2014; Mazgutova and Kosmoz, 2015). What these studies suggested is that the nature of lexical diversity is open to development in short and long terms of language instruction and that lexical diversity is apt at distinguishing different linguistic proficiency levels. The difference in terms of lexical diversity in writing from a proficiency point of view can be seen in studies comparing native and non-native writings. For example, Linnarud (1986) found that native speakers of English produced essays richer in lexical items in terms of diversity than Swedish learners of English. From a similar strand, Harley and King (1989) compared the compositions written by native French students and international French immersion students. They used frequency and number of verb types, lexical errors and variety as lexical diversity measures and indicated that on all measures native speakers displayed a greater diversity than non-native speakers. The difference between learners' ability to display lexical diversity in written discourse is not only apparent in native/non-native comparisons but also across different proficiency levels. For example, Laufer and Nation (1995) compared the written products of learners divided into three proficiency levels with reference to lexical frequency and diversity. They revealed that the most proficient learners could produce the most lexically diverse texts with a bigger number of infrequent words. Olinghouse and Leaird (2009) tested lexical diversity in a study where they aimed to find out whether four lexical measures (lexical diversity, less frequent vocabulary, mean syllable length and number of polysyllable words) are related to writing

quality or not, and whether they display developmental patterns across different curricular levels. Participants of the study were 92 2nd grade and 101 4th grade students of a local public elementary school in the US. It s important to note here that English was spoken as the primary language by all the participants. To collect data, the students were asked to follow a visual prompt to write a story about it and they were given 5 minutes to get prepared and write the story within 15 minutes. Stories were rated by two experienced RAs and the inter-rater reliability was found by a Pearson product-moment correlation to be ".77, .81 and .84 for the organization, plot development and creativity scales respectively" (p.552), which formed the three subset of the analytic scale to assess writing quality. Olinghouse and Leaird (2009) measured lexical diversity by a corrected type-token ratio (CTTR), a recent variant of TTR to minimize the text length effect. The independent samples t test showed a significant difference in lexical diversity and frequency measures between 2nd graders' (t=8.67, df=176, p=.000) and 4th graders' essays (t=3.37, df= 176, p=.001), which means that 2nd graders displayed less lexical diversity than 4th graders. As for the variance of writing scores explained by the independent variables in question, mean syllable length in 2nd graders'' writing (11%) and text length in 4th graders' writings (6.1%) was found to be the strongest indicators of writing quality. The most important implication to be drawn from Olinghouse and Leaird's study is that above and beyond the effect of text length, vocabulary is a very vital construct in written language quality perception with a variance of 12% it explains.

Bulte and Housen (2014), in a similar vein, conducted a study to track the characteristics of development in both syntactic and lexical complexity of learners' writing and to see whether and to what extent these characteristics are related to overall writing quality scores. 90 essays written by 45 randomly selected ESL students in a short intensive English language program were used as the data of the study. 45 of these essays were written at the beginning of the program while the other half were written at the end. There were four months between the first and the last essay writing. All the essays were rated subjectively by two experienced judges. Bulte and Housen (2014) employed three different measures to gauge lexical complexity; D index for lexical diversity, Guiraud index for lexical richness (G) and Advanced Guiraud (AG) for lexical sophistication (see Bulte and Housen, 2014 for a detailed description of the selected indices). In contrast with many studies, the findings indicate that D index of lexical diversity and G index of lexical richness did not increase from first writing to the last. Only AG index exerted an increase

which is not statistically significant. The results, additionally, indicated a difference in overall writing quality scores of the first time (M=48.56, SD=10.56) and last time writing (M=57.16, SD=8.24) on a significant level p<.000. As for predicting overall over all writing quality, D-value and G-value exerted a weak and non-significant correlation. However, when the researchers entered GA index (as lexical sophistication) into a model of four variables, it was seen that this model explained 45% of the variance in the perceived overall writing quality ($F(4,89)=17.672; p<0.001; r=0.67; R^2=0.45$).

Likewise, to track short term changes in lexical complexity and diversity of students' productive language, Mazgutova and Kormos (2015) examined the lexical development (if any) within a one-month-long intensive Academic English program at a British University. 39 undergraduate ESL students were asked to write two argumentative essays at the beginning and end of the language program. Lexical variation (or lexical diversity) was measured by Coh-metrix through MTLD index (Measure of Textual Lexical Diversity). MTLD index is the most recent alternative of type-token ratio that Coh-Metrix readily automate. The study was also interested in measuring syntactic complexity and accuracy development, the related results of which were already reported in the previous section. Mazgutova and Kormos (2015) divided the students into 2 groups of proficiency based on their generic scores of previously taken IELTS. The study suggested an overall increase in the mean of MTLD scores of both groups from first essays to second essays. While the MTLD mean was 73.55 (SD=19.79) for the high proficient group at the beginning, it rose to 86.21 (SD=17.25) at the end of the program. Similarly, for the low-proficient group, the mean MTLD rose from 72.36 to 87.66 (SD=1.15 and 12.2 respectively) during the EAP program.

In order to probe whether and how lexis and lexical errors are one of the primary indicators of writing quality, we must also take a closer look at the issue from the raters' perspectives like Santos (1988) earlier did. Santos (1988) chose two compositions which were equal in length and topic and which were written by a Chinese and a Korean EFL college level learner. The two compositions were then rated by a large cohort of (N=178) university professors of varying branches and years of teaching experience. The compositions were rated on two broad sense; first, in compliance with content, second; in compliance with language. The study also aimed to compare the professors' ratings based on their ages and departments, however, as our study's major concern, we only report the language features which were taken into the upper most consideration by the

professors. According to the results, content of the compositions was rated lower and more harshly than language features on a significant scale. As for the language features, however, lexical errors were found the most irritating errors of all, being very severely rated by the professors as academically unacceptable. The study proposed a 'seriousness rank of errors' and lexical errors formed the first most serious four errors in the rank only followed by article errors in the fifth order.

There are also some studies which lent empirical support to the claim that lexical diversity is closely associated and positively correlated with holistic scoring of writing quality in L2. For example, Ferris (1994) investigated the textual features that were appearent in college level ESL students' writings to find out whether and to what extent these features were related to overall quality scores. Ferris (1994) analyzed a total of 160 essays; 40 from each L1 groups of Arabic, Chinese, Japanese and Spanish. The students wrote the essays under exam conditions, within 35 minutes and about a prompt on culture shock. Based on their essay scores, the students were divided into writing proficiency groups as high (n=100) and low (n=60). The essays were scored by three raters. According to correlation and multiple regression analyses in which a total of 28 determined textual variables were entered as independent variables, five most significant indicators of writing quality scores were: "number of words, synonym/antonym, word length, passives and 3rd person/impersonal pronouns" (p.418). The results also suggested that more proficient group had a wider range of vocabulary and produced longer words and texts as well as more synonym/antonym.

Engber (1995) tested four lexical indices to find out whether and to what extent they were related to holistic scores given to 66 placement compositions written by a group of EL students of different L1 ground. The lexical indices which were tested in this study were; lexical variation or diversity, error-free variation, percentage of lexical errors and lexical density which was operationalised as "the ratio of total number of lexical items with total number of words in the essay" (p.147). The essays were rated by ten experienced raters using a 6-point scale TOEFL writing rubric. The inter-rater reliability was calculated to be high, r=.93. Lexical density, accepted as one of the four indicator of lexical proficiency in this study, exerted a non-significant and low correlation with writing quality scores (r=.23), which means that "percentage of lexical words has little, if any, relationship to quality" (p.148). As for percentage of lexical error, the results indicated a negative and a moderate correlation which was statistically significant (r=-

.43, p<.01). This finding suggested, as expected, that writing scores increased as the number of lexical errors decreased. Lexical variation- operationalized as number of different lexical items and/or lexical diversity- correlated moderately and positively with writing quality scores (r=.45, p<.01), however, this correlation value rose to .57 on a significant level when lexical errors were eliminated. Thus, Engber (1995) concluded that lexical variation with accurate lexical items; that is the amount of accurate and different lexical items predicted the writing scores the best.

Laufer and Nation (1995) devised a then-new approach to measure lexical diversity and named it as Lexical Frequency Profile (LFP). LFP, basically, examines the lexical items in a given text in contrast to the word lists piled up on frequency of usage. Goodfellow, Lamy and Jones (2002) used the Lexical Frequency Profile (Laufer and Nation, 1995) to examine the feasibility of using its data as an index of EFL learners' overall writing quality scores. The learners in this study were French learners as a FL and 36 essays were gathered and analyzed for the research purpose. First, the researchers made adaptations to LFP for assessing French vocabulary items; the first 2.000 most frequent word list and an Academic Word List were produced in a similar strand with the original LFP. The study showed that there is a strong correlation between holistic ratings of students essays and their lexical proficiency characterized by LFP. Based on the LFP profiles, Goodfellow et. al. (2002) concluded that LFP would be a beneficial construct at assessing FL learners' vocabulary levels, which has been found quite related to overall writing ability. In the same year, Jarvis (2002), likewise, measured lexical diversity in the short written compositions of 140 Finnish, 70 Swedish and 66 Native English students of similar age and educational background. To measure lexical diversity, Jarvis (2002) relied on D-Value as the lexical diversity index upon having determined "how well the D formula models the actual TTR curves" (p.63). With methodological advantages of D, Jarvis (2002) also lent support to the positive relationship between lexical diversity and written composition scores of EFL learners. Jarvis (2002) presented, though moderate, a significant and positive correlation only between Swedish students' lexical diversity and writing scores. The same study, however, showed statistically non-significant and low correlations between lexical diversity and writing scores of American and Finnish students. On the other hand, the study confirmed that native speakers always achieved more lexical diversity in their writing compared to total of non-native group.

The literature also holds some studies which examines and compares the written and spoken student performances in terms of lexical diversity. Yu (2009), for example, investigated the relationship between lexical diversity and EFL learners' global writing and speaking scores. The study drew the spoken and written data from the archives of Michigan English Language Assessment Battery (MELAB). 200 written compositions and 25 transcribed candidate interviews were used as the data of the study. Only the compositions and interviews which got the same score from two raters were selected for further analysis to ensure inter-rater reliability. Apart from our own study's particular interest which is lexical diversity/quality scores relation, Yu (2009) also examined this relationship in terms of different student L1 backgrounds (i.e., Chinese, Filipino, Russian, Persian), gender, test taking purpose (i.e., college admission vs. professional certification), and composition topic (personal vs. impersonal). According to linear regression analysis, lexical diversity measured by the D-value was found closely related and positively correlated with overall writing quality scores ($r=0.294$, $p<0.001$, $N=200$). Yu (2009) reported a variance of 11% in writing scores that was explained by lexical diversity alone, which was perceived quite high by the researcher since there might well be other lexical and syntactical issues at play in score variance. Yet, Yu (2009) found that lexical diversity was more successful at anticipating speaking test scores than it was at writing scores. Additionally, overall language proficiency of students was found positively correlated with lexical diversity in writing ($t=4.497$, $p<.001$, $N=199$) and speaking performances ($t=2.748$, $p<.01$, $N=25$). This finding shows that 9.3% of the variance in written compositions and 24.7% of the variance in spoken interviews were explained by lexical diversity. As for so-called 'topic effect', Yu (2009) concluded that topic familiarity displayed a positive correlation with overall writing scores as well as the extent of lexical diversity. Lexical diversity, similar to syntactic complexity as might be remembered, is associated with various aspects of SLA and applied linguistic. One of them is genre. In a recent study, Olinghouse and Wison (2012) studied the role of lexical knowledge with its various dimensions, namely; diversity, maturity, academic words and register. The study examined the relationship of lexical knowledge to human quality ratings assigned to EFL learners' written works across three genres. Though not observed in all genres, lexical diversity was found to be the unique index of quality in story text.

Likewise, Mellor (2011) studied the relationship between lexical diversity, text length and writing quality ratings. Mellor aimed to find out whether text length and lexical

diversity -as combined in a single model- could predict FL writing scores better than these two construct could individually do. 34 Japanese learners of English as a FL participated in the study and they wrote 34 essays which were to be analyzed in terms of text length (characterized by word count) and lexical diversity (measured by six measures). The essay quality was rated by a native speaker rater as "good, above average, below average and poor" (p.2). Lexical diversity indices used in the study are D-measure, TTR, Guiraud Index, Yule's K, Hapax and Advanced Guiraud. Correlation and multiple regression analyses yielded that "lexical diversity together with text length can more accurately predict essay quality than either feature alone in this set of essays" (Mellor, 2011, p.9). Essay length, however, was found superior over lexical diversity indices in predicting essay quality. We regard it important to remind that text length is largely reliant on the particular diversity measure used. We are already informed that D-measure is either little or never affected by text length, yet Mellor (2011) did not discuss this point. The results went on showing that 60% of the variance in quality ratings was explained by text length alone and lexical diversity brought about only 4% of increase in the variance. Mellor (2011) warns us that this proportion of variance can drastically change depending on the learners and tasks and states that the results are only limited to the set of essays used in the study.

Crossley, Salsbury, McNamara and Jarvis (2010) designed a robust study to explore which lexical proficiency features could better predict human quality ratings assigned to essays of L2 learners and native speakers. Lexical proficiency was operationalized by the researchers under three broad categories "breadth of lexical knowledge, depth of lexical knowledge and the accessibility to core lexical items" (p.1). These three broad categories of lexical proficiency – lexical diversity included in the breadth dimension- were measured by a total of 10 measures provided by the computational analysis tool Coh-metrix. A total of 240 essays were collected as the data of this study; 60 essays from beginner, intermediate and advanced level students and 60 essays from native students of similar backgrounds. The researchers assured a variety of linguistic proficiency. All the essays were holistically scored by three experienced raters. The correlation and multiple regression analysis showed that there is a strong positive relationship between three lexical knowledge types and writing quality (r=.66). These lexical dimensions included lexical diversity, word hypernym and word frequency as respectively measured by D-value, average of word hypernym and content word frequency values that were all

provided by Coh-metrix. Crossley et. al. (2010) found that this three-faceted model of lexical proficiency (included lexical diversity) could account for 44% of the variance in human quality ratings of lexical and writing proficiency.

As to show why lexical diversity is important in terms of writing quality assessments, Crossley, Salsbury and McNamara (2011) examined whether and to what extent lexical competence as characterized by Coh-metrix indices could predict students' writing divided into different proficiency levels. For the study, 100 essays were analyzed in consideration of three main categories of lexical proficiency as in Crossley. et. al (2010); these categories are breadth knowledge, depth knowledge and the access to core lexical items. Lexical diversity, as our particular interest, was included in breath dimension of lexical proficiency. Students engaged in 15-minute free writing of their own choice to eliminate the topic effect. The student writings were then grouped into three proficiency levels (beginner, intermediate, advanced) based on three proficiency tests that the students previously participated; TOEFL PBT, TOEFL iBT and ACT ESL Compass reading and grammar tests. The investigated three faceted model of lexical proficiency could successfully discriminate and classify the writings of different levels at a percentage of 69.7, $x^2$ (4) = 24.175, p<.001. As for our specific interest, lexical diversity as measured by measure of Textual Lexical Diversity (MTLD) by Coh metrix was found to be the third strongest indicator of proficiency classification of students' writings with a medium effect size of .250. lexical diversity as the results suggested increased almost linearly across proficiency levels from beginning to native.

Stating that good quality in L2 writing has got multiple facets and relations – rather than linear- with various linguistic features, Jarvis, Grant and Bikowski (2003) conducted a cluster analysis to find out whether there are multiple profiles of highly rated essays which co- occur within clusters. They analyzed two different data sets which were previously used by Ferris (1994; 160 ESL compositions) and by Grant and Ginther (2000; 178 EFL compositions). They examined the two data sets with respect to a total of 21 linguistic features, one of which was lexical diversity measured through type/token ratio. After they tagged all the 21 linguistic features in the data sets, the first data set yielded five different clusters and "three linguistic features for which all five clusters show positive mean Z scores; text length, diversity of vocabulary and emphatics" (p.387). In the second data set (178 EFL compositions) which fell into three clusters "four linguistic features where all three clusters show above-average levels; text length, diversity of

vocabulary, downtowners and adverbials" (p.398). With these results, Jarvis et. al. (2003) suggested that quality in second language writing is more closely related to a set of linguistics features that co-occur and overlap. They also concluded that lexical diversity (measured by type-token ratio) always falls into the scope and characterization of high writing quality.

Years after Jarvis et.al (2003), a similar study was conducted by Friginal, Li and Weigle (2014) to test Jarvis et al's findings and also to explore whether multiple profiles of highly rated essays occur across native and non-native essays. In a similar vein, Friginal et al (2014) also carried out a cluster analysis using SPSS 17.0 to find the distributional patterns of 23 linguistic features in quest. The written non-native data came from TOEFL iBT administration and consisted of 353 graduate and undergraduate essays and 150 native essays were taken from the students of a university in the U.S. The essays were rated by two experienced raters using a TOEFL rubric on a five-point scale. Only those essays which got five from two raters were used for the analysis (N=24 for NNs, N=51 for NS). For our particular interest, it is important to note that lexical diversity was measured by type-token ratio in this study, too. Other linguistic features were tagged in the essays by an automated tagging software. The data fell into 6 clusters. Cluster 1, as the most common profile, consisted of 30 of 75 highly rated essays. In this cluster, "8 out of 24 NNS essays and 22 out of 51 NS essays" (p.9) exerted a profile of longer texts and more diverse vocabulary. Friginal et. al. (2014) concluded that one of the manifold features of highly rated essays was lexical diversity in both native and non-native times essays. From cluster analyses, it is likely to see that lexical diversity is an important member of co-occurring clusters of linguistic features that are visible in highly rated student essays.

## 2.7. Measurement of Syntactic Complexity and Lexical Diversity: Problems and Approaches

In the literature a large number of syntactic complexity measure has been offered. There have been considerable research attempts for decades to find and validate a reliable measure (Wolfe-Quintero et.al, 1998; Ortega,2003). Most of this research has focused on specifying which measure(s) could be objectively used either to track learners' writing development or to assess proficiency. However, there is discrepancy in the results of these studies due to variability and inconsistency among the complexity measures, data size,

and the language tasks and genres operationalized in the data collection procedures (Wolfe-Quintero et. al, 1998; Ortega, 2003; Lu, 2010; Lu,2011).

Not only the inconsistency of measures used but also the small number of them and the limited data size hinders the pooling the results of previous studies. For example, Ortega's comprehensive review (2003) covered twenty-five cross-sectional studies which examined the syntactic complexity development in foreign or second language writing and only four studies included in this review used four to five different measures. The remaining twenty-one studies resorted to only three measures. Likewise, the average number of the written data collected in these studies was less than 100, and the mean number of words in each written sample is 234 with a standard deviation of 110. Similar problems prevailed in more recent work as well. For example, in one study three hundred learner e-mails were syntactically analyzed using only clauses per-T Unit measure (Stockwell and Harrington, 2003). In another study, Ellis and Yuan (2004) similarly used only clauses per T-Unit measure to analyze fifty-two narratives and Beers and Nagy (2009) used mean length of clause in addition to T-Unit ratio to analyze forty-one essays in two different genres. Text length as a measure of syntactic complexity, however, poses serious problems of reliability. Although text length was often associated with overall writing quality scores assigned by human judges (Guo et al. 2013) ,studies showed that text length does not necessarily increase along with syntactic complexity indices (Stockwell, 2005; Becker, 2010).

More recently, however, syntactic complexity and lexical diversity research has started to benefit from a computational tool named Coh-Metrix (Graesser, McNamara and Kulikowich, 2011) to offer subtler predictors. The syntactic and lexical indices provided by this automated tool have been validated by several recent studies (McNamara et al, 2010; Crossley and McNamara, 2011,2012; Crossley et. al., 2011).

As can be seen, there is an inconsistency in the number and type of complexity and diversity measures applied to different sets of written data in the literature which makes it very difficult to compare and combine previous results. There has been, as well, recent computational approaches to the measurement of syntactic complexity and lexical diversity beside some refinements of measurement formulations. In the following of this section, a detailed view of these novel approaches and refinements as well as a critical evaluation of former traditional measures will be presented.

### 2.7.1. Measuring syntactic complexity in L2 with a critique of traditional complexity measures

In second language writing syntactic complexity has been prominent and associated with the degree of sophistication and variation of syntactic structures (Foster and Shekan, 1996; Wolfe-Quintero et al., 1998; Ortega, 2003). Syntactic complexity in second language writing research has been attributed a paramount importance as its relationship with overall writing quality ratings and writing development has been manifested in a large bulk of previous research (Beers and Naggy, 2009; Norris and Ortega, 2009; Lu, 2011; Mazgutova and Kormos, 2015). A large set of different measures has been asserted to quantify syntactic complexity in the area of second language writing research.

Before reporting the L2 literature about the approaches to the measurement of written syntactic complexity, we must also take a brief look at the issue in L1 studies. Although L1 and L2 writing and their developmental characteristics widely differ, syntactic complexity measurement has its origins in L1 writing (Lu, 2010). Length of production unit has traditionally been accepted as shared index of both L1 and L2 complexity. Length of production units as an index will be evaluated in the following paragraphs more in detail. When this similarity (i.e., length of production units) is set aside, it is likely to see that sets of complexity measures in L1 and L2 are fundamentally different. Majority of measures in L1 largely depend on frequency of use of particular structures (Covington et al., 2006). Additionally, some syntactic complexity measures in L1 aim to characterize the cognition load that different syntactic structures exert on learners' mind as well as calculations of ratio of words over constituent length (Hawkings, 1994). The operational difference seen in L1 syntactic complexity measurement is not unexpected when we think of the differences in two areas; L1 and L2 writing. Approaches to syntactic complexity measurement in L1 which focus on processing, frequency or comprehension would be more eligible to complexity research in reading (Biber, Gray and Poonpon, 2011).

When it comes to syntactic complexity measurement in L2 writing, it is possible to see that majority of these measures are proposed somehow to quantify either clausal, sentential or T-unit length, amount of subordination and coordination (Ortega, 2003). There has been a popular stance of syntactic complexity which emphasizes that averaged length of production units and the rate of subordination (i.e., embedding elaborated clauses to a main clause) echo syntactic complexity; that is the longer and the more

subordinated the units are, the more complex they are (Biber et.al., 2011). This point of view reflects a very straightforward logic, however, many studies widely made use of T-unit as the production unit and relied on its average length and subordination ratio (Ellis and Yuan, 2004; Larsen-Freeman, 2006). It is noteworthy to define and exemplify a T-unit as it is one of the most commonly applied yardsticks of syntactic complexity despite its shortcomings. A T-unit, as exemplified below, consists of a grammatically correct sentence (a main clause and its dependent clauses, if any);

> I don't know [ [ why I was expecting [to see something else]] (Biber et. al, 2011, p.7)

The above example shows a single T-Unit which includes 11 words and two dependent clauses which are shown in brackets. We already mention a heavy reliance on T-unit based syntactic complexity measures in related research; either on mean length of T-unit (MLTU) or on clauses per T-unit C/TU). Ortega (2003), in her comprehensive meta-analysis, examined 27 studies on written syntactic complexity of college level EFL/ESL students. She found out that 25 studies used MLTU, while 11 of them also employed C/TU as the complexity index. Despite this wide use, the shortcomings of T-unit analysis in assessing syntactic complexity have been shown in several studies (Ortega, 2003; Ravid, 2005; Norris and Ortega, 2009; Biber et. al., 2011). What these studies conclude is that students do not demonstrate a linear progression from simple to complex by adding more elaboration to a simple base clause, that is, syntactic complexity is above and beyond of a straightforward logic which follow a linear increase in the length of T-unit and its subordination ratio.

Likewise, Ortega (2003) reported that syntactic complexity in 27 studies, included in her research synthesis sample, did not produce distinguishing findings with a T-unit based measure (mean length of T-unit) about different proficiency levels. Ortega (2003), in this research synthesis, conducted a total of 68 comparisons of different proficiency levels, some of which were adjacent and the other were nonadjacent. In 43 of these 68 proficiency comparisons, mean length of T-unit as a proficiency index only differ on a scale smaller than -/+1.8 words between different proficiency groups. As also commented in Ortega (2003), the biggest drawbacks of these studies is that they do not discuss the controversial issues in syntactic complexity measurement and adapt in advance a T-unit based approach (either MLTU or C/TU). As seen, however, this acceptance without questioning resulted in findings showing that syntactic complexity in students' essays do

not increase at all as they progress across levels, which is hard to interpret. Biber et al., (2011) provided an example to clarify the reasons why T-unit based measures as a syntactic complexity and/or proficiency index are difficult to be trusted. They provided two sentences cited verbatim below. The first sentence was taken from a natural conversation, while the second one was from an academic course book:

1. Well, since he got so upset, I just didn't think we would want to wait for Tina to come back

T-Unit length:20

Number of dependent clauses: 4

2. This may be part of the reason for statistical link between schizophrenia and membership in the lower socioeconomic classes

T-unit length: 20

Number of dependent clauses: 0 (Biber et al., 2011, p.14).

The researchers wrote that although the two sentences have the same T-unit length, they differ in the number dependent clauses they have. This is to say, according to mean length of T-unit complexity measure, there is no difference at all between the complexities of the two sentences. On the other hand, when measured by clauses per T-unit (C/TU), the first sentence is far more complex than the second sentence. Biber et. al., (2011) stated that both sentences are inherently complex in their own ways and that how problematic T-unit based measures could be to determine syntactic complexity. Biber et.al., (2011) designed their own study to examine the true syntactic characteristics of written academic texts and to examine whether and to what extent T-unit based measures could capture the complexities of academic text types. For the research purposes, the researchers used 429 research articles published in 11 academic journals. The study also drew back on a conversation corpus consisting of 723 text files of face-to-face conversation recordings. By the help of a syntactic tagger software, the researchers determined the occurrences of 28 investigated syntactic features in their samples. All in all, the research showed that academic writing had different syntactic complexity features than speech as expected. More importantly, however, syntactic complexity in academic writing is manifested through complex noun phrases rather than complex clauses. Biber et. al., (2011) also showed that clausal subordination measured by clauses per T-unit (C/TU) is more visible in speech complexity rather than in writing.

### 2.7.2. Measuring lexical diversity in L2 with a critique of traditional complexity measures

As earlier stated, measurement to lexical diversity is far from being simple and there have been numerous approaches proposed to quantify the diversity. At the onset of this section, we first need to address in some detail the basic and earlier approaches to measurement before a thorough discussion of computational formulas of lexical diversity. One of the earliest ways of measurement of lexical diversity is 'the number of different words' (NDW). As cited in Malvern et. al. (2004, p.16), Klee (1992) used this measure of lexical diversity which clearly and simply depended on range. Klee (1992) studied with children between ages of 24 and 50 months and showed that NDW could potentially discriminate age and vocabulary deployment range of normally developing children and those with a language impairment. As deBoer (2014) puts it, NDW is the number of the simplest quantification technique of lexical diversity by counting the number of different words, or so-called types. NDW, however, is not free from flaws although it could have worked reliably in some earlier studies, like that of Klee's, in specific language impairment areas.

The biggest flaw of NDW stems from the fact that it is largely reliant on text length. Thus, it is not likely to soundly and authentically compare two texts in different lengths with NDW. Differences in text size result in serious problems of reliability and judgement as Malvern et. al. (2004) stated "how many different words appear in a language sample will in all probability depend on how many words there are in total and this is the heart of many problems in the measurement of lexical diversity" (p.16).

Another one of the most widely applied lexical diversity measures is the type-token ratio (TTR). Type is counted as the number of different words in a text, while token is counted as the number of all words in the text. TTR is estimated through the division of token by the type count, yielding an index between 0 and 1- the higher the score, the more diverse the vocabulary range is. TTR was proposed to modify NDW and to increase its reliability. It is doubtless that taking a ratio rate is more reliable than simply counting the number of different words, however TRR also suffers from the same problem of text length.

During natural language production, some repetitions, especially of functional language parts such as prepositions and articles may often take place, which increases the token but not necessarily the type. Malvern et. al. (2004) described this situation and

stated that the usage of a new word increases the type only if it has not been used before. Malvern et. al (2004) points out that type and token do not increase at the same rate when different words are added to the language sample.

One of the computational formulas to quantify lexical diversity is called vocd-D (Malvern et al., 2004). It is proposed in the literature as a novel and a reliable approach to the assessment of lexical diversity and, in fact, vocd-D has been derived from the famous TTR method. Voc-D, as described by Malvern et al (2004), first randomly takes 100 samples of 35 tokens. TTR is then calculated for each sample, the mean of which is stored. The same process is repeated with samples of 36 to 50 tokens to create some empirical TTR curves from the means of these samples. The D coefficient as described by Malvern et al (2004, p.51), is applied to create the best-fitting TTR curve among the empirical curves. As all these process is randomly computed, the same processes are repeated to reach the uppermost accurate results. An ultimate D value is the last product of these procedures and it generally varies from 10 to 100, higher numbers indicating greater diversity. Vocd- D is automatized and readily offered by the recent computational text processing tool of Coh-Metrix.

Another one of recent assessment approaches of lexical diversity is called the measure of textual lexical diversity (MTLD). McCarthy and Jarvis (2010) described the rationale behind this formulation in detail. According to McCarthy and Jarvis (2010), we must begin understanding the rationale of MTLD by first looking into segmental TTR coined by Jonhson (1994). This version of TTR splits the texts into segments of typically 100 words length and calculates the TTR of each segment to reach a mean TTR value of all segments. However, it is useful to remind that the words that fall out of the set segments are disregarded from the index in segmental TTR. McCarthy and Jarvis (2010) stated that this version of TTR could work well with long texts where it is possible to split the text into larger segments. Nonetheless, when the text size is small, the segmental sizes also become small, which decreases the sensitivity of the index. In short, MTLD is an index of lexical diversity which is derived from segmental TTR that is widely dependent on text length. MTLD, on the other hand, is supposed to minimize text length effect by taking factor sizes smaller than 100 words. In MTLD, all factors must first reach a default TTR of .720. Thus, it is possible to claim that MTLD could well eliminate the text length and sensitivity problems (McCarthy and Jarvis, 2010). Several studies confirmed that MTLD, as automatically offered by Coh-Metrix, is one of the most trustworthy and

distinctive indicators in the full inventory of indicators of Coh-metrix (Crossley and McNamara, 2009; Crossley, Salsbury and McNamara, 2009).

Computational approaches to the measurement: Coh-Metrix

The current availability of computational tools of discourse processing have enabled the analysis of large textual data in terms of linguistic components. Certain syntactic complexity and lexical diversity indices are readily provided by Coh-Metrix, an automated tool of accurate and detailed textual analysis (Graesser et. al., 2004). A general overview of Coh-metrix can be seen in Table 2.1 below;

**Table 2.1.** *Questions and Answers about Coh-metrix*

| **Questions** | **Answers** |
|---|---|
| What is Coh-Metrix? | Computational linguistics and recent advents in text processing technologies have lately created a large sum of complicated discourse indicators. A team at the Institute for Intelligent Systems at The University of Memphis have developed a text processing tool named Coh-metrix that incorporates these novel and sophisticated text indices (McNamara, Graesser, McCarthy and Cai, 2014, p.164) |
| What function does it serve? | Coh-metrix provides a wide number of linguistic and discourse features of a text through plentiful indices of readability, language and cohesion. Coh-Metrix provides its textual analysis whereby automated syntactic trees and parsing, and latent semantic analysis as well as "traditional textual measures such as average word length, average sentence length, and the readability formulas of Flesch Reading Ease and Flesch-Kincaid Grade Level (Klare 1974–1975)" (McNamara, et. al., 2014). |
| Why should we rely on Coh-Metrix? | Syntactic complexity and lexical diversity research has started to widely benefit from Coh-Metrix for analysis of multilevel textual features (Graesser, McNamara and Kulikowich, 2011) to offer subtler predictors. There has been a broad approval and employment of the tool in the related research community. The syntactic and lexical indices provided by this automated tool have been validated by several recent studies that investigated linguistic textual features as well as textual cohesion, coherence and lexical diversity and lexical proficiency (McNamara et al, 2010; Crossley and McNamara, 2011,2012; Crossley et. al., 2011). |
| What are we specifically using it for? | In our study, we are peculiarly interested in 3 syntactic complexity and 2 lexical diversity indices, which are summarized below |

A summary of syntactical and lexical indices exclusively used in L2 writing quality research and in the current study as well is outlined below.

*Syntactic complexity indices*: Coh-metrix measures syntactic complexity in three fundamental ways. First, it calculates the mean number of words appearing before the main verb in a sentence with the assumption that the higher this number is, the more complex a sentence is. Second, syntactic similarity is measured by Coh-metrix as an index of syntactic complexity with the assumption that more complex sentences have less uniform and inconsistent constructions. Third, Coh-metrix provides noun phrase (NP) density and the mean number of modifiers per NP as a syntactic complexity index.

*Lexical diversity indices:* As stated before, lexical diversity refers to the number of words a learner has in his lexicon. Traditional lexical diversity measurement includes the number of word types by tokens (i.e., the division of unique word number by all instances of words) known as type-token ratio (TTR). However, TTR does not produce reliable results as the tokens are very much dependent on the text length. To eliminate this problem, we use lexical diversity indices reported by Coh-metrix which are more sophisticated, reliable and free from text length effect. They, namely, are the *Measure of Textual Lexical Diversity* (MTLD: McCarthy and Jarvis, 2010) and *VocD* (Malvern, Richards, Chipere and Duran, 2004).

# CHAPTER 3

## 3. METHODOLOGY

In this chapter, an overview of the research design and context, the participants, the data collection instruments as well as data analysis procedures are presented.

### 3.1. Participants and research context

The participants of the present study consist of three cohorts. The largest participating group is ELT undergraduate students, while 3 raters and 8 instructors also took part in paper-scoring and providing qualitative probe. First of all, a total of 204 undergraduate ELT students participated in the current study majoring at the English Language Teaching (ELT) department of a large Turkish public university. The participants were recruited by convenience sampling method which emphasizes the ease of access as the sampling principle (Creswell, 2002). The participation was voluntary and the participants were informed by consent forms, being assured that their volunteering decision would not bias their course grades (see Appendix-2). In accordance with our research purposes, the participants were selected from two groups; freshman and senior students. 102 students were freshman at the end of their first year of study having taken 'Academic Writing and Report Writing' courses I and II in two consecutive semesters. The students, before entering the four-year ELT degree program, have to pass an extensive English test. Upon being admitted to the program, the university also mandates another extensive language test which assesses the writing, speaking, listening, reading skills as well as vocabulary and grammar dimensions of students' English proficiency. The students either pass this exam, and directly embark on studying in the degree program or fail and pursue one-year long English preparatory program offered by the university's School of Foreign Languages. The four-year ELT degree program offers a wide range of courses from skill based language courses at the first year to literature, linguistics and teaching methodology courses throughout the remaining three years.

**Table 3.1.** *Distribution of participants*

| Participants | Number |
|---|---|
| 4th Year Students | 102 |
| 1st Year Students | 102 |
| Instructors | 8 |
| Raters | 3 |
| **Total** | **215** |

The students are exposed to various academic genres both in spoken and written mode and expected to produce language in the forms of manifold homework, reports and presentations. As Wolfe-Quintero et. al (1998) wrote, "program level may be the most valid developmentally" (p.9). With this in mind and based on the claim that syntactic complexity and lexical diversity in L2 writing develops over time with more instruction and exposure and differ across proficiency levels (Linnarud, 1986; Harley and King, 1995; Mazgutova and Kormos, 2015; Vyatkina, 2015; Treffers-Daller, Parslow and Williams, 2016), we acknowledge that our first and fourth year students can be different in terms of linguistic proficiency.

Secondly, we collected our qualitative data by means of semi-structured interview questions which were asked to eight instructors who had been working in an ELT four-year degree program of the same public university where we collected the quantitative data. These instructors held different years of experience in scoring undergraduate student writing at the time of data collection. The mean year of experience of interviewees is 16,6, which possibly indicates that our qualitative data was provided by highly experienced instructors.

**Table 3.2.** *Instructors and their year of experience in scoring*

| Instructors | Year of Experience in Scoring |
|---|---|
| Inst.1 | 20 |
| Inst.2 | 14 |
| Inst.3 | 15 |
| Inst.4 | 40 |
| Inst.5 | 3 |
| Inst.6 | 9 |
| Inst.7 | 15 |
| Inst.8 | 17 |
| Mean of Experience in Scoring | 16,6 |

Thirdly, the essays were rated by two separate scorers: one with over thirty year of experience in teaching and grading various kinds of academic writing, one with over ten years of experience in teaching and assessing academic writing and a native speaker of English who is following her MA degree in the ELT program. A third scorer was recruited to resort to when there was an inconsistency of 1 point and more between two.

### 3.2. Data Collection

We compiled a learner corpus from undergraduate students. The corpus was collected in a way to control the confounding variables of text and task conditions such as;

- genre (i.e., opinion essay)
- task conditions (i.e., timed and unplanned writing within classroom)

One of the variables at play in the relationship of syntactic complexity and its predictive power of writing quality appears to be genre. Each genre, as socially occurring language patterns, fulfill a distinct social function and by doing so, they draw upon different language constructions (Halliday & Hassan, 1985). The genres vary mainly in two aspects; first, the linguistic characteristics vary at the micro-level; second, the global organization of the text structures vary at macro-level (Donovan & Smolkin, 2006). Generating qualified texts across genres, thus, is likely to associate with genre specific syntactic structures that facilitate the realization of communicative functions of the genre. Therefore, we decided on 'opinion essay' as the genre to fix the so called 'genre effect' that might differentiate the complexity and diversity outputs (Ravid, 2005; Beers and Nagy, 2009).

Although there are contradictions, metacognitive stages of L2 writing has been found somehow related to planning time and task conditions. Planning time refers to conditions such as pre-task or free writing while task conditions are mainly concerned with topic. Planned writing in an L2 writing setting contributed to greater to writing accuracy and fluency while unplanned instant and free writing resulted in less accuracy and fluency (Ellis and Yuan, 2004, p.78). Kroll (1990) also found that compositions written by a class of L2 learners at home contained more accuracy and were scored higher than those written at class within a 60-minute time constraint. Kroll (1990, p. 153), however, warns that not knowing the exact amount of time spent at home and the particular task conditions might blur the results. Contrary to the claims and findings that

planned and untimed writing comprise more accurate language and better writing quality, there is a plenty number of study which claims the opposite. Likewise, free writing conditions without a preparation through a pre-task and within a time restraint was found to produce:

- a larger number of ideas
- greater global writing quality
- more lexical variety and writing fluency (Ong, 2013a; Ong and Zhang, 2010, 2013)

In our study, therefore, we set our task conditions as unplanned and timed writing as to minimize worries similar to those of Kroll's (1990) and so that conditions beyond our control would not interfere in our findings. Another study examined the effect of planning time and topic on language and idea aspects as well as writing organization in 106 EFL compositions (Ong, 2013b). She revealed that topic of writing as a task condition showed the most significant effect on the dependent variables in question. Thus, we ran a captious process to decide on our writing topic, which is outlined below.

Before the data collection procedure was started, we applied for a research ethics approval to the institutional review board, and the necessary formal permission were acknowledged (see Appendix-3) The students were provided with a topic on which they were asked to write an opinion essay. The selection of the opinion essay topic was based on a decision-making process in which we resorted to experts' ideas through a specially designed questionnaire (See Appendix-4). This procedure aimed to immobilize the so-called topic effect. The questionnaire was comprised of 10 topics, all of which were compiled from an IELTS study recommendation page found on http://ieltsliz.com/100-ielts-essay-questions/education/ web address. The selected topics were about education, university and campus life, learning and teaching in general. The candidate topics were presented to 20 experts who all have been teaching in a public university's ELT department for about 10 years. 5 of these experts were teaching 'Academic Writing and Report Writing' at the time of questionnaire application, and the rest of them had previously taught the same course for at least one semester. The experts were supposed to prioritize 3 most likely topics that they taught our participants could write over with maximum ease and amount. The questionnaire also included a part where it demanded the experts' optional topic recommendations except from given 10 topics. (if any). Only one expert provided a topic recommendation which was about a futuristic view of

teaching profession. The topic prioritized the most and thus selected for the current study was:

> "Students at universities often have a choice of places to live. They may choose to live in university dormitories, in private student residences, or they may choose to live in apartments in the town. Where would you prefer to live? Why? Give reasons for your preference"

The above-given opinion essay writing topic was prioritized by 15 experts by being put in a triple order of priority. Then, the topic was placed on a writing sheet that was designed for the data collection procedure (see Appendix-5). The writing sheet included the name and the surname of the participants. The students were assured that their names would remain confidential and were only to be used for classification purposes. There was also an instruction on the sheet with the topic and the duration of the writing task, which was one hour- slightly more than a regular class hour.

### 3.3. Data Analysis

Upon being collected, all essays were typed on Microsoft Word 2016 to be processed on Coh-Metrix. Coh-metrix provided the intended indices about syntactic complexity and lexical diversity. The syntactic complexity and lexical diversity indices provided by Coh-metrix as well as writing quality scores were transferred into a statistical analysis software SPSS for further analysis. For research questions 1 and 2, a multiple regression analysis was conducted to discover the relationship of syntactic complexity, i.e., measured by a) mean number of words before main verb, b) mean number of modifiers per noun clause, c) syntactic similarity), lexical diversity, i.e., a) MTLD, b) VocD) with L2 writing quality. Multiple regressions also led to the discovery of the extent to which syntactic complexity and lexical diversity indices, both jointly and separately, explain the variance in L2 writing quality. For research question 3, we computed 5 independent samples t-tests to find out if there is any significant difference in the writings of freshmen and seniors in terms of syntactic complexity and lexical diversity. For the 4$^{th}$ research question, we employed a content analysis within qualitative analysis paradigm. The research questions, the number of participants, the Coh-metrix indices and statistical analysis were displayed in Table 3.3.

The present study is based on a mixed research paradigm. Therefore, it utilizes a qualitative inquiry approach as well as quantitative and statistical data analysis methods. As to explore the extent to which syntactic complexity and lexical diversity are engaged

in the perceptions of ELT instructors, we benefited from semi-structured interview questions. Eight instructors who have been scoring student academic papers in an ELT department were interviewed and their responses were recorded. The participating instructors were asked for their written consent before their responses were voice-recorded (see Appendix-6) The semi-structured interview questions were derived from the related literature by the researcher. The questions were, then, presented to expert opinion and only after two sessions of feedback, the questions were refined and took their final forms. The refinement process of the interview questions can be tracked in Appendix

**Table 3.3.** *Overview of research methodology*

| **Research Questions** | **Number of Participants (n)** | **Coh-metrix indicies** | **Statistical Analysis** |
|---|---|---|---|
| What are the syntactic complexity, lexical diversity and writing quality scores of participating students? | 204 | *mean number of words before main verb<br>*mean number of modifiers per noun clause<br>*syntactic similarity<br>*Measure of Textual Lexical Diversity (MTLD: McCarthy and Jarvis, 2010)<br>**VocD* (Malvern et.al., 2004).<br>*Overall Writing Quality Scores | Descriptive Statistics (Mean, Standart Deviations) |
| What is the relationship between syntactic complexity, lexical diversirty and L2 writing quality scores assigned by human raters? | 204 | | Bivariate Regression Analysis<br>Multiple Regression Analysis |
| Is there a difference between syntactic complexity, lexical diversity and writing quality scores of learners at different curricular levels? | 102<br>102 | | Independent Samples t-Tests |
| To what extent are syntactic complexity and lexical diversity engaged in the perception of ELT instructors who evaluate undergraduates' academic writings? | 8 | | Content analysis |

As suggested by Seidman (2012), interviews can function to uncover ‘stories ‘and “stories are ways of knowing” (p.7). Bertaux (1981; as cited in Seidman, 2021, p.8) also points out that people could convey much information on a given matter if given an opportunity to express. The interviews were transcribed by the researcher and transferred into Microsoft Word documents. The transcriptions were checked for spelling or punctuation and mistakes were corrected. The transcribed data were analyzed with thematic content analysis as outlined by Weber (1990). Analysis procedures suggested by both Weber (1990) and Creswell (2012) were employed. Firstly, the researcher broadly read the transcribed data on several occasions by taking margin notes by hand. These margin notes, afterwards, evolved into broad themes which were few in number. The first themes, after having been discussed for feedback with the advisor, were transferred into NVivo 11 pro, which is a qualitative analysis tool for further and detailed analysis. The first drawn themes were labelled as codes in NVivo and thoroughly read more than once to define persistent and interesting codes. NVivo automatically appointed the selected data chunks under the title of ‘references’, and these data chunks afterwards were used as verbatim quotations and proof for the codes. The codes and the appointed references were checked for further refinement. In these checks with the supervision of a qualitative data analysis expert, some misunderstandings were resolved, several names of the codes changed, several new codes emerged and some themes changed their place in the tree diagrams offered by NVivo.

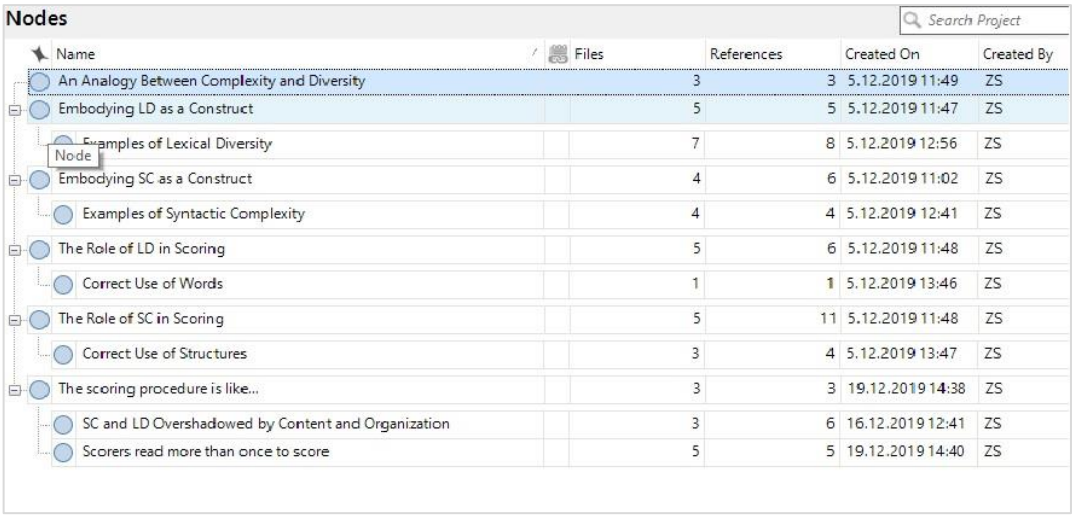



**Figure 3.1.** *Screenshot for the categorical themes*

The end version of theme-code refinement procedure can be seen in Appendix 8 and a screenshot from NVivo displays the first and the latest version of code-theme-reference organization in Figure 3.1 above.

### 3.3.1. Essay quality ratings

The essays were copied and filed. Labels were assigned to each essay such as, for example, 4-1, which indicates that the essay was written by a 4th year student and it continued until 4-102. The same labelling was conducted for first year students as well. The essays were rated by two separate raters: one with over thirty year of experience in teaching and grading various kinds of academic writing, one with over ten years of experience in teaching and assessing academic writing and a native speaker of English who is following her MA degree in the ELT program. A third rater was recruited to resort to when there was an inconsistency over 1 point between two raters. To evaluate the quality of essays, a standardized rubric used in assessing TOEFL iBT essays was resorted (see Appendix-9). This rubric globally evaluates the quality of essays having a scores ranging from 0 to 5, 5 indicating the best maximum score.

### 3.3.2. Inter-Rater and intra-rater reliability

The extent to which a test produces consistent results is known as its reliability. For a test to be considered reliable, its results should be very much alike across different administrations and different raters, which is also called 'inter-rater reliability'. Our raters scored the student papers twice to ensure intra-rater reliability as well. The second scoring was carried out 6 months after the first one. With a six-month time-lapse, the two raters scored the same papers again. One way to measure the inter and intra-rater reliability is to run the Pearson product moment correlation between two different raters and two scoring procedures. This correlation will yield the overall rate of agreement of all raters who read the compositions. Therefore, we computed the Pearson-product moment correlation to assess the inter-rater and intra-rater reliability between our first and second raters. As can be remembered, our first rater is retired professor of applied linguistics who got over a thirty years of experience in scoring various kinds of scholarly papers (Rater Z). As for our second rater, she is a native American speaker working as an English lecturer in a Turkish state university and has been delivering foreign language writing

courses for several years (Rater B). The raters' pseudonym were the initial letters of their actual names.

**Table 3.4.** *Results of Pearson correlations coefficients between two raters across two scoring procedures (for 1st year students' scores)*

| | **Rater Z 1st Scoring** | **Rater Z 2nd Scoring** | **Rater B 1st Scoring** | **Rater B 2nd scoring** |
|---|---|---|---|---|
| Rater Z 1st Scoring | 1 | | | |
| Rater Z 2nd Scoring | .449** | 1 | | |
| Rater B 1st Scoring | .342** | .198* | 1 | |
| Rater B 2nd Scoring | .003 | .047 | .132 | 1 |

** Correlation is significant at the 0.01 level (2-tailed).

In the Table above, the Pearson-product moment correlation of the two raters could be seen across two rating procedures. Table 3.4 above displays both the intra-rater reliability and the inter-rater reliability across scorings for the first year students.

First, the intra-rater reliability is given here. The scores assigned to first-year students by Rater Z in two scoring procedures were positively correlated with each other ($r=(102), .449, p<0.01$). This positive correlation was found to be statistically significant. This means, although weak, there is a consistency in Rater Z's scores across two rating procedures. On the other hand, the other rater, that is, Rater B's scores were not correlated with each other across two scoring. When it comes to the inter-rater reliability values for the first students' essays, Rater B and Rater Z were only consistent at the r. value of .342 ($p<0.01$) at the first scoring procedure. However, at the second scoring procedure, namely after six months, the two raters lost all the consistency between each other ($r(102)= .047, p>0.5$).

**Table 3.5.** *Results of Pearson correlations coefficients between two raters across two scoring procedures (for 4th year students' scores)*

| | **Rater Z 1st Scoring** | **Rater Z 2nd Scoring** | **Rater B 1st Scoring** | **Rater B 2nd scoring** |
|---|---|---|---|---|
| Rater Z 1st Scoring | 1 | | | |
| Rater Z 2nd Scoring | .509** | 1 | | |
| Rater B 1st Scoring | .546** | .346** | 1 | |
| Rater B 2nd Scoring | .331** | .316** | .469** | 1 |

When it comes to the intra and inter rater reliability values at the fourth-year students' essays, we witness some higher correlation values. Rater Z, for example, displayed an r value of .509 on a significant level, which means Rater Z's scores tended

to increase or decrease the same direction. However, we should highlight that this consistency was statistically significant only on a moderate level. Likewise, Rater B's scores also showed a positive correlation in two of the ratings with an r value of .469 on a statistically significant level. To sum up, two of the raters scores were moderately consistent across two scoring procedures.

At the first scoring, Rater Z and Rater B compromised again on a moderate strength. The correlation between Rater Z and Rater B was found to be statistically significant with an r value of .546. However, the correlation, though still positive and significant, decreased to .316 between two raters at the second rating.

In order for a relationship to be strongly correlated, the r value should be at least between .60 and .79, while the higher scores are considered to be 'very strong' (Evans, 1996). As the Test of Written English Guide (TWE) showed that a strong and a positive correlation between .76 and .82 can indicate an acceptably high inter-rater reliability. As the TWE guide points out, 213,221 essays in 10 TWE administrations between years of October 2001 and November 2003 yielded inter-rater reliability scores ranging from .76 to.82.

### 3.3.3. Coh-metrix measures

The current availability of computational tools of discourse processing have enabled the analysis of large textual data in terms of linguistic components. Certain syntactic complexity and lexical diversity indices are readily provided by Coh-Metrix, an automated tool of accurate and detailed textual analysis (Graesser et. al., 2004). A general overview of Coh-metrix can be seen in Table 3.6 below;

**Table 3.6.** *Questions and Answers about Coh-metrix*

| Questions | Answers |
|---|---|
| What is Coh-Metrix? | Computational linguistics and recent advents in text processing technologies have lately created a large sum of complicated discourse indicators. A team at the Institute for Intelligent Systems at The University of Memphis have developed a text processing tool named Coh-metrix that incorporates these novel and sophisticated text indices (McNamara, Graesser, McCarthy and Cai, 2014, p.164) |

Table 3.6. (continued) *Questions and Answers about Coh-metrix*

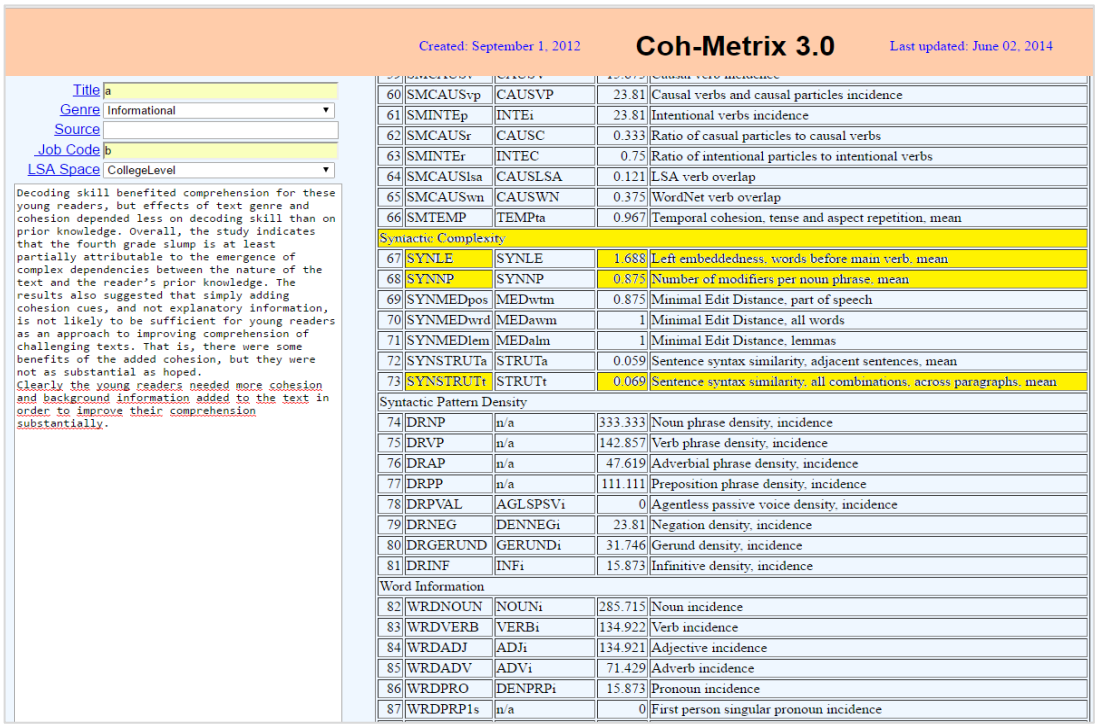
| | |
|---|---|
| What function does it serve? | **Coh-metrix provides a wide number of linguistic and discourse features of a text through plentiful indices of readability, language and cohesion. Coh-Metrix provides its textual analysis whereby automated syntactic trees and parsing, and latent semantic analysis as well as "traditional textual measures such as average word length, average sentence length, and the readability formulas of Flesch Reading Ease and Flesch-Kincaid Grade Level (Klare 1974–1975)" (McNamara, et. al., 2014).** |
| Why should we rely on Coh-Metrix? | Syntactic complexity and lexical diversity research has started to widely benefit from Coh-Metrix for analysis of multilevel textual features (Graesser, McNamara and Kulikowich, 2011) to offer subtler predictors. There has been a broad approval and employment of the tool in the related research community. The syntactic and lexical indices provided by this automated tool have been validated by several recent studies that investigated linguistic textual features as well as textual cohesion, coherence and lexical diversity and lexical proficiency (McNamara et al, 2010; Crossley and McNamara, 2011,2012; Crossley et. al., 2011). |
| What are we specifically using it for? | In our study, we are peculiarly interested in 3 syntactic complexity and 2 lexical diversity indices, which are summarized below |

A summary of syntactical and lexical indices exclusively used in L2 writing quality research and in the current study as well is outlined below.

Syntactic complexity indices: Coh-metrix measures syntactic complexity in three fundamental ways. First, it calculates the mean number of words appearing before the main verb in a sentence with the assumption that the higher this number is, the more complex a sentence is. Second, syntactic similarity is measured by Coh-metrix as an index of syntactic complexity with the assumption that more complex sentences have less uniform and inconsistent constructions. Third, Coh-metrix provides noun phrase (NP) density and the mean number of modifiers per NP as a syntactic complexity index.

Lexical diversity indices: As stated before, lexical diversity refers to the number of words a learner has in his lexicon. Traditional lexical diversity measurement includes the number of word types by tokens (i.e., the division of unique word number by all instances of words) known as type-token ratio (TTR). However, TTR does not produce reliable results as the tokens are very much dependent on the text length. To eliminate this problem, we use lexical diversity indices reported by Coh-metrix which are more sophisticated, reliable and free from text length effect. They, namely, are the *Measure of*

*Textual Lexical Diversity* (MTLD: McCarthy and Jarvis, 2010) and *VocD* (Malvern, Richards, Chipere and Duran, 2004). The following two screen shots of Coh-metrix display syntactic complexity and lexical diversity indices of a sample short text.



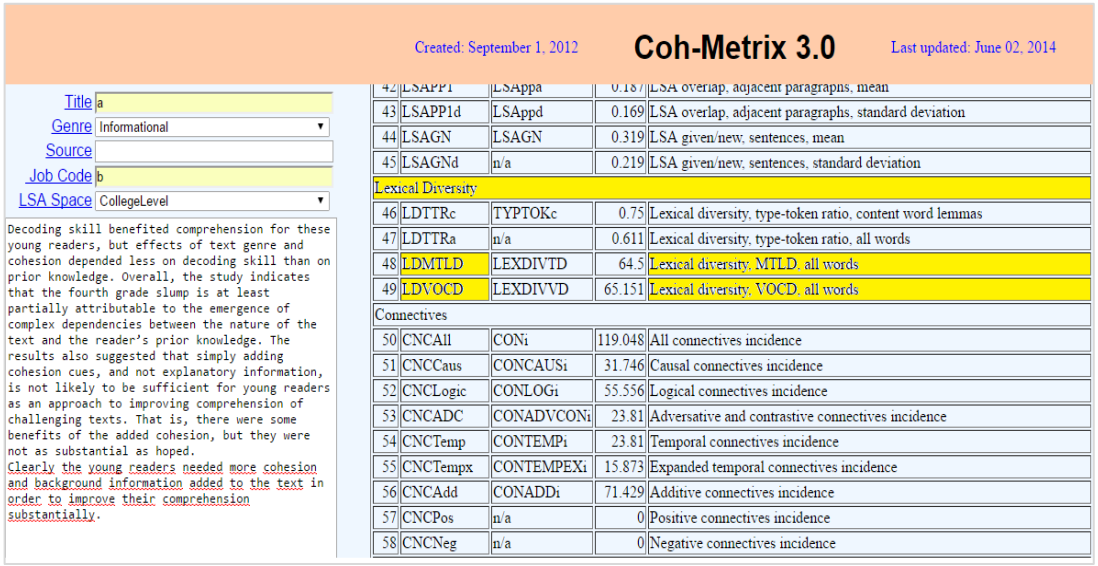
Created: September 1, 2012 **Coh-Metrix 3.0** Last updated: June 02, 2014

Title a
Genre Informational
Source
Job Code b
LSA Space CollegeLevel

Decoding skill benefited comprehension for these young readers, but effects of text genre and cohesion depended less on decoding skill than on prior knowledge. Overall, the study indicates that the fourth grade slump is at least partially attributable to the emergence of complex dependencies between the nature of the text and the reader's prior knowledge. The results also suggested that simply adding cohesion cues, and not explanatory information, is not likely to be sufficient for young readers as an approach to improving comprehension of challenging texts. That is, there were some benefits of the added cohesion, but they were not as substantial as hoped.
Clearly the young readers needed more cohesion and background information added to the text in order to improve their comprehension substantially.

| | | | | |
|---|---|---|---|---|
| 60 | SMCAUSvp | CAUSVP | 23.81 | Causal verbs and causal particles incidence |
| 61 | SMINTEp | INTEi | 23.81 | Intentional verbs incidence |
| 62 | SMCAUSr | CAUSC | 0.333 | Ratio of casual particles to causal verbs |
| 63 | SMINTEr | INTEC | 0.75 | Ratio of intentional particles to intentional verbs |
| 64 | SMCAUSlsa | CAUSLSA | 0.121 | LSA verb overlap |
| 65 | SMCAUSwn | CAUSWN | 0.375 | WordNet verb overlap |
| 66 | SMTEMP | TEMPta | 0.967 | Temporal cohesion, tense and aspect repetition, mean |
| Syntactic Complexity | | | | |
| 67 | SYNLE | SYNLE | 1.688 | Left embeddedness, words before main verb, mean |
| 68 | SYNNP | SYNNP | 0.875 | Number of modifiers per noun phrase, mean |
| 69 | SYNMEDpos | MEDwtm | 0.875 | Minimal Edit Distance, part of speech |
| 70 | SYNMEDwrd | MEDawm | 1 | Minimal Edit Distance, all words |
| 71 | SYNMEDlem | MEDalm | 1 | Minimal Edit Distance, lemmas |
| 72 | SYNSTRUTa | STRUTa | 0.059 | Sentence syntax similarity, adjacent sentences, mean |
| 73 | SYNSTRUTt | STRUTt | 0.069 | Sentence syntax similarity, all combinations, across paragraphs, mean |
| Syntactic Pattern Density | | | | |
| 74 | DRNP | n/a | 333.333 | Noun phrase density, incidence |
| 75 | DRVP | n/a | 142.857 | Verb phrase density, incidence |
| 76 | DRAP | n/a | 47.619 | Adverbial phrase density, incidence |
| 77 | DRPP | n/a | 111.111 | Preposition phrase density, incidence |
| 78 | DRPVAL | AGLSPSVi | 0 | Agentless passive voice density, incidence |
| 79 | DRNEG | DENNEGi | 23.81 | Negation density, incidence |
| 80 | DRGERUND | GERUNDi | 31.746 | Gerund density, incidence |
| 81 | DRINF | INFi | 15.873 | Infinitive density, incidence |
| Word Information | | | | |
| 82 | WRDNOUN | NOUNi | 285.715 | Noun incidence |
| 83 | WRDVERB | VERBi | 134.922 | Verb incidence |
| 84 | WRDADJ | ADJi | 134.921 | Adjective incidence |
| 85 | WRDADV | ADVi | 71.429 | Adverb incidence |
| 86 | WRDPRO | DENPRPi | 15.873 | Pronoun incidence |
| 87 | WRDPRP1s | n/a | 0 | First person singular pronoun incidence |

***Figure 3.2.*** *A Coh-metrix screen shot displaying syntactic complexity indices*



Created: September 1, 2012 **Coh-Metrix 3.0** Last updated: June 02, 2014

Title a
Genre Informational
Source
Job Code b
LSA Space CollegeLevel

Decoding skill benefited comprehension for these young readers, but effects of text genre and cohesion depended less on decoding skill than on prior knowledge. Overall, the study indicates that the fourth grade slump is at least partially attributable to the emergence of complex dependencies between the nature of the text and the reader's prior knowledge. The results also suggested that simply adding cohesion cues, and not explanatory information, is not likely to be sufficient for young readers as an approach to improving comprehension of challenging texts. That is, there were some benefits of the added cohesion, but they were not as substantial as hoped.
Clearly the young readers needed more cohesion and background information added to the text in order to improve their comprehension substantially.

| | | | | |
|---|---|---|---|---|
| 42 | LSAPP1 | LSAppa | 0.187 | LSA overlap, adjacent paragraphs, mean |
| 43 | LSAPP1d | LSAppd | 0.169 | LSA overlap, adjacent paragraphs, standard deviation |
| 44 | LSAGN | LSAGN | 0.319 | LSA given/new, sentences, mean |
| 45 | LSAGNd | n/a | 0.219 | LSA given/new, sentences, standard deviation |
| Lexical Diversity | | | | |
| 46 | LDTTRc | TYPTOKc | 0.75 | Lexical diversity, type-token ratio, content word lemmas |
| 47 | LDTTRa | n/a | 0.611 | Lexical diversity, type-token ratio, all words |
| 48 | LDMTLD | LEXDIVTD | 64.5 | Lexical diversity, MTLD, all words |
| 49 | LDVOCD | LEXDIVVD | 65.151 | Lexical diversity, VOCD, all words |
| Connectives | | | | |
| 50 | CNCAll | CONi | 119.048 | All connectives incidence |
| 51 | CNCCaus | CONCAUSi | 31.746 | Causal connectives incidence |
| 52 | CNCLogic | CONLOGi | 55.556 | Logical connectives incidence |
| 53 | CNCADC | CONADVCONi | 23.81 | Adversative and contrastive connectives incidence |
| 54 | CNCTemp | CONTEMPi | 23.81 | Temporal connectives incidence |
| 55 | CNCTempx | CONTEMPEXi | 15.873 | Expanded temporal connectives incidence |
| 56 | CNCAdd | CONADDi | 71.429 | Additive connectives incidence |
| 57 | CNCPos | n/a | 0 | Positive connectives incidence |
| 58 | CNCNeg | n/a | 0 | Negative connectives incidence |

***Figure 3.3.*** *A Coh-metrix screen shot displaying lexical diversity indices*

# CHAPTER 4

## 4. RESULTS

### 4.1. Introduction to the Chapter

In this section, four research questions of our study will be respectively answered. The statistical tests will be displayed through tables and/or figures as well as through relevant written explanations. We first mention the inter-rater reliability process in short (see Inter-Rater Reliability subsection in Methodology Chapter), then, we separately illustrate the descriptive statistical results that depict the 1st year and 4th year students' compositions in terms of word count (e.i text length), categories of word ranges, writing scores and five Coh-Metrix indices. To be specific, we would like to remind the audience of the Coh-Metrix indices. Three Coh-Metrix indices that we employed to assess the syntactic complexity of our students' compositions were;

- the mean number of words appearing before the main verb in a sentence
- the mean number of modifiers per noun phrase
- syntactic similarity (e.g the extent to which the syntactic constrıctions are varied)

To assess lexical diversity, we ran two Coh-Metrix indices. They, namely, are the Measure of Textual Lexical Diversity (MTLD: McCarthy and Jarvis, 2010) and VocD (Malvern, Richards, Chipere and Duran, 2004).

### 4.2. Syntactic Complexity, Lexical Diversity and Writing Quality Scores of Participating Students

#### 4.2.1. Syntactic complexity, lexical diversity and writing quality scores of 1st year students' compositions

In the first research question, we aimed to find out the the syntactic complexity, lexical diversity and writing quality scores of our participating students. To do so, 102 first year and 102 fourth year students (N=204) wrote an approximately one page long English composition on a topic which is related to the accommodation types and preferences at university. These compositions were then rated following a TOEFL writing rubric by two human raters. A Pearson product-moment correlation was run to determine the inter-rater reliability between two raters of our study. There was a moderate, positive

correlation between the scores of the native scorer and the experienced scorer, which was statistically significant ($r$ = .546, $n$ = 102, $p$ < .001). In order for a relationship to be strongly correlated, the r value should be at least between .60 and .79, while the higher scores are considered to be 'very strong' (Evans, 1996). As the Test of Written English Guide (TWE) showed, a strong and a positive correlation between .76 and .82 can indicate an acceptably high inter-rater reliability. As the TWE guide points out, 213,221 essays in 10 TWE administrations between years of October 2001 and November 2003 yielded inter-rater reliability scores ranging from .76 to.82. Thus, to increase our inter-reliability score, we restored to a third rater. First, we labeled the inconsistent papers. The 68 compositions which got a one-point and higher discrepancy were rated by the third rater using the same TOEFL rubric on a 5-scale. The third rater was also a native speaker American lecturer. After the third rater assigned her scores, another inter-reliability check was run and this once the r value jumped to .78 with a p value of 0.01 (r(204)=0.78, p<.001). Hence, we can report that our inter-rater reliability falls into high and strong category. The overall writing quality scores of our participants were determined by the mean of two raters' scores. To assign a quality score in 68 inconsistent compositions, we agreed to the third rater's score to eliminate the one point or higher discrepancy.

To increase the interrater reliability, a third rater also scored the inconsistent 68 compositions and another inter-reliability check was run and this once the r value jumped to .78 with a *p* value of 0.01 (r(204)=0.78, p<.001).

We collected 102 written compositions from the 1st year students at our ELT department. Table 4.1 below shows basic numbers about these 1st year compositions. As can be seen, 1st year students wrote a sum of 28.648 words with a mean number of 281 words.

**Table 4.1.** *Descriptives of 1st year students' essays*

| N | Total Word Count | Min. | Max. | M | Std. Deviation |
|---|---|---|---|---|---|
| 102 | 28.648 | 113 | 473 | 281 | 71.619 |

Number of word appearing in these essays change up to 473, as the highest number of word. Table 4.2 shows the frequency and percentage rates of word ranges in 1st year students' essays.

**Table 4.2.** *Frequency and percentage rates of word ranges in 1st year students' essays*

| Word Range | Frequency (n) | Percentage (%)* |
|---|---|---|
| 0-200 | 12 | 11 |
| 200-300 | 49 | 48 |
| 300-473 | 41 | 40 |
| **Total** | **102** | **100** |

***Percentages were rounded off to the nearest number*

1st year students' essays which were anatomized so far were evaluated by three independent raters to be assessed in terms of writing quality. The evaluation of the essays was carried out twice to ensure intra-rater reliability. Each essays were given a point between 0 and 5 which indicated a scale of writing quality. The higher the score is, the more qualified the essays are supposed to be according to the TOEFL writing rubric used. Table 4.3 below displays the writing quality descriptive scores which 1st year student essays were given by three independent raters in the first and second scoring. The mean score of 1st year writing quality, at the first scoring, is 3.2 and 3.5 at the second scoring as can be seen in the table. The lowest score assigned to a 1st year student is 2.0 while the highest quality score is 4.0. Along with the mean score, the minimum and maximum scores also rose up at the second scoring. None of the 1st year students were assigned 5 as the highest possible score to be achieved at neither of the scoring times.

**Table 4.3.** *Descriptives of 1st year students' writing quality scores*

| Scoring Time | n | Min. | Max. | M | Std. Deviation |
|---|---|---|---|---|---|
| First Time | 102 | 2.0 | 4.0 | 3.2 | 3.20 |
| Second Time | 102 | 2.8 | 4.5 | 3.5 | .329 |

We also conducted a paired samples t-test to see if there is a statistically significant difference between the mean scores in writing quality grades assigned in the first and second procedure. As Table 4.3 above illustrates, 1st year students' mean writing quality scores rose from 3.2 to 3.5 in the second scoring. This mean difference was found statistically non-significant [t(101)=-.904, p=.368], which means that 1st year students got higher scores in the second scoring procedure at statistically significant level. This statistically non-significant result suggests that our raters displayed a consistency in rating the 1st year student essays across two rating procedures.

Table 4.4 below shows the descriptive statistics of lexical diversity and its two Coh-Metrix indices – MTLD and VocD. The lexical diversity mean measured by MTLD in 1st

year students' writing is 68.37 (*SD*= 14,27). VocD measure on Coh-Metrix provided another mean score of lexical diversity *(M*=75,70 *SD*= 15,46). When combined, these two lexical diversity measures provided by Coh- Metrix offers a unified value which indicates a total body of lexical diversity in 1st year students' writing, that is; (*M*=137).

**Table 4.4.** *Descriptive values of lexical diversity of 1st year students' essays*

| | n | Minimum | Maximum | Mean | Std. Deviation |
|---|---|---|---|---|---|
| MTLD | 102 | 40 | 100 | 68.37 | 14.27 |
| VocD | 102 | 48 | 115 | 75.70 | 15.46 |

Table 4.5 outlines three Coh-Metrix syntactic complexity indices and a total number describing the syntactic complexity value calculated in 1st year students' essays. As can be seen, mean number of word coming before main verb, which is also called left embeddedness, is 3,76. Mean number of modifiers used per noun phrase is ,5775 while it is important to note that it is optional to use a modifier in each noun phrase. Mean of syntactic similarity refers to "the proportion of intersection tree nodes between all sentences and across paragraphs" as Coh-metrix defines (McNamara et al., 2014, p.71). The mean of tree similarity of all sentences and across paragraphs was found to be ,1116.

**Table 4.5.** *Descriptive values of syntactic complexity of 1st year students' essays*

| | n | Minimum | Maximum | Mean | Std. Deviation |
|---|---|---|---|---|---|
| Mean number of words | 102 | 2 | 11 | 3.76 | 1.249 |
| Mean Number of Modifiers | 102 | .29 | .87 | .5775 | .12132 |
| Syntactic Similarity | 102 | .05 | .18 | .1116 | .02883 |

When combined, these three syntactic complexity measures provided by Coh-Metrix offers a unified value which indicates a total body of syntactic complexity in 1st year students' writing, that is; (*M*=4,5).

### 4.2.2. Syntactic complexity, lexical diversity and writing quality scores of 4th[t] year students' compositions

This section briefly presents the descriptive analysis findings of 4th year students' opinion essay writings. At the very onset of the section, we should warn that all the values reported here are only to anatomize several extents of 4th year students' writings beyond any intention and/or means of inferential purposes for the moment.

Table 4.6 below displays the number of words appearing in 102 essays collected from 4th year participating students. 102 4th year student essays generated a sample of opinion essay writing with 36.861 words on total. The mean number of words in each essay was calculated to be 361. The shortest essay contained 127 words while the longest essay contained 685 words. When we extract the lowest number in a data set from the highest number, we get the range score, which is 558 in our case.

**Table 4.6.** *Descriptives of 4th year students' essays*

| N | Total Word Count | Min. | Max. | M | Std. Deviation |
|---|---|---|---|---|---|
| 102 | 36.861 | 127 | 685 | 361.3 | 113.792 |

**Table 4.7.** *Frequency and percentage rates of word ranges in 4th year students' writing*

| Word Range | Frequency (n) | Percentage (%) |
|---|---|---|
| 0-200 | 4 | 3.9 |
| 200-300 | 31 | 30.1 |
| 300-400 | 33 | 32 |
| 400-500 | 20 | 19.4 |
| 500-685 | 14 | 13.6 |
| **Total** | **102** | **100** |

Table 4.7 illustrates frequency and percentage rates of word ranges in 4th year students' writings. As can be seen, only 4 students wrote 200 words length at most. 33 essays contained words between 300 and 400, which forms the widest range. The number of students who wrote essays with words above 500 is 14. The 4 shortest essays contained 127, 168 190 and 197 words respectively. The 4 longest essays, on the other hand, contained 603, 630, 638 and 685 words respectively.

4th year students' essays which were anatomized so far were evaluated by three independent raters to be assessed in terms of writing quality. The evaluation of the essays was carried out twice to ensure intra-rater reliability. Each essays were given a point between 0 and 5 which indicated a scale of writing quality. The higher the score is, the more qualified the essays are supposed to be according to the TOEFL writing rubric used. Table 4.8 below displays the writing quality descriptive scores which 4th year student essays were given by three independent raters at two scoring procedures. The mean score of 4th year writing quality is, at the first scoring, 3.7 and it rose up to 3.8 at the second

scoring as can be seen in the table. The lowest score assigned to a 4th year student is 2.2 while the highest quality score is 5.0. Although the maximum score that a 4th year student got did not change across scoring times, the minimum score became 3.0. As the minimum and maximum scores along with the slight increase in the mean scores suggest, the 4th year students got slightly higher scores in the second procedure and this slight difference at the means was found to be statistically significant [t(101)=-2.464, p<.05)].

**Table 4.8.** *Descriptives of 4st year students' writing quality scores*

| Scoring Time | n | Min. | Max. | M | Std. deviation |
|---|---|---|---|---|---|
| First Time | 102 | 2.2 | 5.0 | 3.7 | .619 |
| Second Time | 102 | 3.0 | 5.0 | 3.8 | .344 |

Table 4.9 below shows the descriptive statistics of lexical diversity and its two Coh-Metrix indices – MTLD and VocD. The lexical diversity mean measured by MTLD in 4th year students' writing is 6.585 (*SD*= 1.882). VocD measure on Coh-Metrix provided another mean score of lexical diversity *(M*=7.065, *SD*= 2.261). When combined, these two lexical diversity measures provided by Coh- Metrix offers a unified value which indicates a total body a lexical diversity in 4th year students' writing, that is; (*M*=13.352, *SD*=3.869).

**Table 4.9.** *Descriptive statistics of lexical diversity and its two indices in 4th year students' essays*

| | n | Min. | Max. | M | Std. Deviation |
|---|---|---|---|---|---|
| MTLD | 102 | 1.004 | 9.932 | 6.585 | 1.882 |
| VocD | 102 | 1.035 | 9.976 | 7.065 | 2.261 |

Table 4.10 outlines three Coh-Metrix syntactic complexity indices and a total number describing the syntactic complexity value calculated in 4th year students' essays. As can be seen, mean number of word coming before main verb, which is also called left embeddedness, is 4.177. Mean number of modifiers used per noun phrase is ,630 while it is important to note that it is optional to use a modifier in each noun phrase. Mean of syntactic similarity refers to "the proportion of intersection tree nodes between all sentences and across paragraphs" as Coh-metrix defines (McNamara et al., 2014, p.71). The mean of tree similarity of all sentences and across paragraphs was found to be ,121.

**Table 4.10.** *Descriptive statistics of syntactic complexity and its three indices in 4th year students' essays*

| | n | Min. | Max. | M | Std. Deviation |
|---|---|---|---|---|---|
| Mean Number of Words Before Main Verb | 102 | 1.608 | 7.952 | 4.177 | 1.29 |
| Mean Number of Modifiers per NP | 102 | ,351 | ,916 | ,630 | ,131 |
| Mean of Syntactic Similarity | 102 | ,054 | ,228 | ,121 | ,032 |

## 4.3. Curricular Level Differences Among the Investigated Variables

In this subsection, we aimed to answer the second research question of our study which was questioning whether there was a difference between syntactic complexity, lexical diversity and writing quality scores of learners at different curricular levels (e.g. 1st and 4th year students).

### 4.3.1. Differences in text length

We first present the difference in terms of total word count (Text length henceforth). Table 4.11 below displays the comparison of word counts and mean number of words of 1st year and 4th year students' essays. The 1st year students wrote visibly shorter essays than 4th year students. In other words, our 4th year students produced 8.213 more words on total comparing to 1st year students.

**Table 4.11.** *A numerical comparison of 1st and 4th year students' essays*

| Curricular Level | n | Total Word Count | Min. | Max. | M | Std. Deviation |
|---|---|---|---|---|---|---|
| 1st Year Students | 102 | 28.648 | 113 | 473 | 281 | 71.619 |
| 4th Year Students | 102 | 36.861 | 127 | 685 | 361 | 113.792 |

An independent samples t-test was run to find out if the mean differences of word counts between groups are statistically significant or not. The t-test finding showed that 4th year students' essays (M=361,38; SD=113,7) contain more words than 1st year students' essays (M=280,86; SD=71,6) and that this mean difference is statistically significant $t$ (202)=6,048, $p$=.000.

### 4.3.2. Differences in writing scores

An independent samples t-test was conducted to see if there is a significant difference between 1st year and 4th year students' writing quality scores. There was a significant difference between the means of 1st year students' writing quality scores (M=3.2, SD=.3.20) and 4th year students' writing quality scores (M=3,7, SD=.619). Specifically, these results suggest that our 4th year students scored higher than the 1st year students and this difference in the mean scores was found to be statistically significant [t(202)=-9.957, p=.000)]. As Table 4.12 below also displays, another independent samples t-test was carried out to see if there is a difference between 1st and 4th year students in the second scoring. The findings suggested that, also in the second scoring, the 4th year students (M=3.8, SD=.344) outdid the 1st year students (M=3.5, SD=.329). The slight difference found in the writing quality scores assigned in the second scoring time was also statistically significant [t(202)=-6.669, p=.000].The results showed that at both scoring procedures the 4th year students scored higher than the first year students and the mean differences were statistically significant.

**Table 4.12.** *Results of independent samples t-test for writing quality scores by curricular level*

| | **1st Year Students** | | | **4th Year Students** | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | ***M*** | ***SD*** | ***n*** | ***M*** | ***SD*** | ***n*** | ***t*** | ***df*** | ***p*** |
| Writing Quality in the First Scoring | 3.2 | 3.20 | 102 | 3.7 | .619 | 102 | -9.95 | 202 | .000 |
| Writing Quality in the Second Scoring | 3.5 | .329 | 102 | 3.8 | .344 | 102 | -6.66 | 202 | .000 |

### 4.3.3. Differences in syntactic complexity

Coh-metrix measures syntactic complexity in three fundamental ways. The independent samples t-test results suggest that in all three syntactic complexity indices that Coh-Metrix provided, 4th year students excelled the 1st year students. To state syntactic complexity Coh Metrix first, calculates the mean number of words appearing before the main verb in a sentence with the assumption that the higher this number is, the

more complex a sentence is. In independent samples t- test results, we found a mean difference in the number of words coming before the main verb in each sentence of the compositions of 1st (M=3.76 SD=1,24) and 4th year students (M=4,15 SD=1,26), these mean differences are statistically significant according to the independent samples t-test results as shown in the table 4.13 below.

**Table 4.13.** *Results of independent samples t-test for 'number of words coming before main verb' by curricular level*

| | **1st Year Students** | | | **4th Year Students** | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | ***M*** | ***SD*** | ***n*** | ***M*** | ***SD*** | ***n*** | ***t*** | ***df*** | ***p*** |
| Number of words before main verb | 3.76 | .123 | 102 | 4.15 | .124 | 102 | -2.24 | 202 | <.05 |

Second, syntactic similarity is measured by Coh-metrix as an index of syntactic complexity with the assumption that more complex sentences have less uniform and inconsistent constructions. The mean scores of 1st (M=.111 SD=.028) and 4th year students (M=.121 SD=.033) in this syntactic complexity index of Coh-metrix were also different, these mean differences were found to be statistically significant as displayed in table 4.14 below;

**Table 4.14.** *Results of independent samples t-test for 'syntactic similarity' by curricular level*

| | **1st Year Students** | | | **4th Year Students** | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | ***M*** | ***SD*** | ***n*** | ***M*** | ***SD*** | ***n*** | ***t*** | ***df*** | ***p*** |
| Syntactic Similarity | .111 | .028 | 102 | .121 | .033 | 102 | -2.36 | 202 | <.05 |

It is noteworthy to note the fact that Coh-metrix measures syntactic similarity in a different manner than the other two syntactic complexity indices. That is to say, concerning syntactic similarity, the lower the number is, the less similar the structures are, which indicates a greater variety of syntactic structures used in an essay. As can be seen in the Table 4.14 above, 1st year students achieved less similarity meaning more complexity than 4th year students. Therefore, we have to say that among three syntactic complexity measures only in syntactic similarity 1st year students excelled the 4th year students.

Third, Coh-metrix provides noun phrase (NP) density and the mean number of modifiers per NP as a syntactic complexity index. A statistically significant mean difference was found in this index as well. The fourth year students used higher number of modifiers per NP than 1st year students and this difference was found statistically significant (p<.001) as can be seen in the table 4.15 below;

**Table 4.15.** *Results of independent samples t-test for mean number of modifiers per NP by curricular level*

| | **1st Year Students** | | | **4th Year Students** | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | ***M*** | ***SD*** | ***n*** | ***M*** | ***SD*** | ***n*** | ***t*** | ***df*** | ***p*** |
| Mean Number of Modifiers | .577 | .121 | 102 | .636 | .136 | 102 | -3.25 | 202 | <.001 |

These results indicate that in terms of syntactic complexity the mean scores are different on a statistically significant scale between 1st and 4th year students in all of the measures of syntactic complexity; that is, number of words coming before the main verb, syntactic similarity and number of modifiers per NP. However, only in two of the syntactic complexity measures 4th year students excelled the 1st year students. In syntactic similarity measure, on the other hand, 1st year students outdid the 4th year students, which means 1st year students displayed more variety in their syntactic diversity adding more to their syntactic complexity.

### 4.3.4. Differences in lexical diversity

We used lexical diversity indices reported by Coh-metrix which are more sophisticated, reliable than traditional measures like TTR and free from text length effect. They, namely, are the *Measure of Textual Lexical Diversity* (MTLD: McCarthy and Jarvis, 2010) and *VocD* (Malvern, Richards, Chipere and Duran, 2004). In two measures we found statistically significant differences between 1st and 4th year students' compositions in terms of lexical diversity. The mean score in both indices were different and the 4th year students excelled the 1st year students in both indices, however only in VocD the difference was statistically significant.

In MTLD, the first and fourth year students differ in terms of mean numbers, however, this difference was not statistically significant according to the independent

samples t-test results. MTLD mean score of 4th year students was 71.517 with a SD of 15.67 while the first year students' mean score for the same index was 68.373 with a SD of 14.27.

**Table 4.16.** *Results of independent samples t-test for 'MTLD' by curricular level*

| | **1st Year Students** | | | **4th Year Students** | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | ***M*** | ***SD*** | ***n*** | ***M*** | ***SD*** | ***n*** | ***t*** | ***df*** | ***p*** |
| MTLD | 68.37 | 14.27 | 102 | 71.51 | 15.67 | 102 | -1.49 | 202 | >.05 |

In VocD, the 4th year students mean scores (M=80.59 SD=14.89) was higher than that of 1st year students (M= 75.70 SD=15.46), additionally this difference was statistically significant as demonstrated in the Table 4.17 below. These results altogether indicate lexical diversity increases along with the curricular level, however this difference is statistically significant only in VocD measure, not in MTLD measure.

**Table 4.17.** *Results of independent samples t-test for 'VocD' by curricular level*

| | **1st Year Students** | | | **4th Year Students** | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | ***M*** | ***SD*** | ***n*** | ***M*** | ***SD*** | ***n*** | ***t*** | ***df*** | ***p*** |
| VocD | 75.70 | 15.46 | 102 | 80.59 | 14.89 | 102 | -2.23 | 202 | <.05 |

## 4.4. Inter-correlations between variables and variances explained

### 4.4.1. Correlations of syntactic complexity, lexical diversity and text length with writing quality

In this section, we aim to answer our 3rd research question which was about the relationships of syntactic complexity, lexical diversity and text length with writing quality scores. Syntactic complexity and lexical diversity indicators were derived from the automated text analyzing software Coh-metrix. We used three Coh-Metrix indicators for syntactic complexity; they are namely; syntactic similarity, number of modifiers per NP and number of words coming before the main verb. Coh-metrix also provided two measures for lexical diversity, they namely are; MTLD and VocD. Text length was calculated through the number of total words appearing in each composition and lastly, the writing quality scores were obtained through the means of three independent scorers' grades. A Pearson product-moment correlation coefficient was computed to assess the relationships among all these variables as can be seen in Table 4.18**.**

As our biggest concern and dependent variable has been writing quality scores, we will begin reporting the relationships of 6 independent variables with writing quality scores. Firstly, the highest association between writing quality was found to be with text length. Though weak, there is a positive and statistically significant relationship between text length and writing quality scores ($r(204)=.379$, $p<0.01$). As this finding suggests, the quality scores of student compositions tend to increase along with the text length. Although this association ($r=.379$) is statistically weak, we could conclude that human scorers are likely to assign higher scores to longer compositions.

**Table 4.18.** *Results of Pearson correlations coefficients among seven variables*

| | **Syntactic Similarity** | **Number of Modifiers** | **Number of Word Before Main verb** | **MTLD** | **VocD** | **Text Length** | **Writing Quality** |
|---|---|---|---|---|---|---|---|
| Syntactic Similarity | 1 | | | | | | |
| Number of Modifiers | -.219** | 1 | | | | | |
| Number of Word Before Main verb | -.417** | .383** | 1 | | | | |
| MTLD | -.222** | .246** | .223** | 1 | | | |
| VocD | -.139* | .192** | .156* | .815** | 1 | | |
| Text Length | .038 | .155* | .170* | .011 | .053 | 1 | |
| Writing Quality | .092 | .141* | .110 | .088 | .177* | .449** | 1 |

*** Correlation is significant at the 0.01 level (2-tailed)*
**Correlation is significant at the 0.05 level (2-tailed)*

When it comes to syntactic complexity and its relationship with writing quality, among three Coh-metrix indices, only mean number of modifiers displayed a statistically significant, yet weak, correlation with writing quality scores ($r(204)=141$, $p<0.01$). This finding suggests that the more modifiers are used, the higher scores tent to be assigned to the compositions by human raters. Other two syntactic complexity indices (e.g; syntactic similarity and number of words coming before main verb) could only produce very weak and non-significant positive correlations with writing quality scores with r scores of .092 and .110 respectively.

As for lexical diversity and its relationship with writing quality scores, VocD measure provided by Coh-metrix yielded a very weak, yet positive and statistically

significant correlation with writing quality scores ($r(204)=.177$, $p<0.05$), which indicates that as lexical diversity measured through VocD index increases, the writing quality scores also tend to very slightly increase. The other lexical diversity measure MTLD, however, did not produce a meaningful association with writing quality scores with a very weak positive r of .088.

To sum up the independent variables' relationships with writing quality scores, it is likely to conclude that text length had the strongest significant and positive relationship was with writing quality followed respectively by syntactic complexity (observed in 1 out of 3 indices) and lexical diversity (observed in 1 out of 2 indices).

Apart from writing quality, text length also appeared as a very important variable in our analyses. Thus, this section depicts the relationship of 'writing quality scores assigned by human raters', 'syntactic complexity' and 'lexical diversity' with *text length*. There found a positive and moderately strong correlation between text length and writing scores. This finding suggested that when the text gets longer, the scores tend to get higher ($r(204)=.449$, $p<0.01$). Neither of the two lexical diversity indices (e.g. *MTLD and VocD*) provided by Coh-metrix could produce statistically significant and even moderate correlations with text length. Syntactic complexity, on the other hand, with its two Coh-metrix indices (e.g *number of words before main verb* and *the number of modifiers per NP*) yielded, though very weak, positive and statistically significant associations with text length. As number of words coming before main verbs increases, the compositions also tends to be slightly longer ($r(204)=.170$, $p<0.05$). Likewise, a similar correlation was also found between number of modifiers and text length, which states that text length tends to slightly go up along with the number of modifiers used per NP, or vice versa ($r(204)=.155$, $p<0.05$). Though expected higher, these positive and significant correlations are not surprising since both indices could contribute to the total length of the texts. In other words, it is likely to see longer texts as students use bigger number of modifiers, which is actually optional and as they use more words, which makes each sentence longer.

Another aspect uncovered by our correlation results is concerned with the computerized text processing tool Coh-metrix and its indices. As remembered, we resorted to three syntactic complexity and two lexical diversity indices measured by Coh-metrix. Our correlation results yield important considerations about the internal consistency of Coh-Metrix. As known, reliability is about the consistency of results yielded by an assessment tool. Internal consistency, as a sub category of reliability, is

related to consistent and similar results being produced when the same construct is tested through different means. When lexical diversity, for instance, was measured by two different measures, that is, with MTLD and VocD, we had a very strong positive correlation that is statistically significant between the two related measures of the same construct ($r(204)=.815$, $p<0.05$). As this correlation suggests, MTLD values tent to increase while VocD values also increase, which greatly adds to the internal consistency of Coh-metrix when it comes to measuring lexical diversity. Similarly, concerning the syntactic complexity indices of Coh-metrix, Pearson product-moment correlations were found to be statistically significant on a moderate scale and negative. For example, as the number of words coming before the main verb increases, the syntactic complexity decreases or vice versa ($r(204)=-.417$, $p<0.01$). Syntactic similarity is measured by Coh-metrix as an index of syntactic complexity with the assumption that more complex sentences have less uniform and inconsistent constructions. Therefore, the smaller the similarity, the more complex the text is. Thus the negative correlation between syntactic similarity index and other indices is, in fact, something expected and desirable for the internal consistency of Coh-Metrix.

### 4.4.2. Variance in the writing quality scores explained

We computed a hierarchal regression analysis to explore the extent of variance in the writing scores by lexical diversity and syntactic complexity indices provided by Coh-metrix as well as by text length. Table 4.19 below displays the results of hierarchical regression analysis. As the table presents, the R square of the model was found to be .206, which means that the independent variables of the regression model, altogether, explain 20.6 percent of the variance in writing quality scores of our students.

**Table 4.19.** *Hierarchical Multiple Regression Analysis with a Three-Layered Model (Dependent Variable; Writing Quality Overall Scores)*

| Model | R | R Square | Standard Error | F Model | R Square Change | F Change |
|---|---|---|---|---|---|---|
| Text Length | .449 | .202 | .422 | 51.02* | .202 | 51.02* |
| Lexical Diversity Indices | .480 | .230 | .416 | 19.95** | .029 | 3.72** |
| Syntactic Complexity Indices | .495 | .245 | .415 | 10.66** | .015 | 1.28** |

A hierarchical multiple regression analysis with a three-layered model (Dependent Variable; Writing Quality Overall Scores) was run to explore the variance in writing quality scores explained by the current study's main two independent variables (lexical variety and syntactic complexity) as well as a strong confounding variable, which is text length. Text length, as the first-entered variable to the model, explains 20.2% of the variance alone, which means that 20.2% of a writing quality score assigned by human raters for a student essay comes from text length. This unique contribution of text length to the writing quality scores was found to be on a statistically significant scale [$F(1, 202)=51.02$, $p<0.00$].

After eliminating the effect of text length as an apparently significant variable, hierarchical multiple regression analysis also explains the unique contribution of lexical variety to the overall writing quality scores. The variance explained by lexical variety to the model was found to be quite low (2.9%) and statistically significant [$F(3.200)=19.95$, $p<0.05$]. When it comes to syntactic complexity, the other independent variable of present study, it is seen that over and beyond the effect of lexical variety and text length, 1.5% of the overall writing quality scores has been explained by syntactic complexity alone. This unique contribution as well as the regression model's total contribution to the variance in writing quality scores are statistically significant [$F(6.197)=10.66$, $p<0.05$]. After all, lexical variety and syntactic complexity together with text length explains the 24.5% of the variance in writing quality scores assigned by human raters. What should be noted with great attention to make more sense of this 24.5% of variance is that lexical variety and syntactic complexity have been measured by objective and automated measures provided by a specifically designed web tool (Coh-metrix), while the writing quality scores were given by human raters following a holistic rubric. One should always remember that even by following a standardized rubric writing assessment can be subjective and many untouched factors can come into play particularly affecting the raters' perceptions.

### 4.5. Unfolding Syntactic Complexity: Embodying It as a Construct

In this subsection, we answer the 4$^{th}$ research question which was constructed with the intention of asking about the perceptions of ELT instructors related to syntactic complexity and lexical diversity in the assessment procedure in undergraduates' academic writing. After the semi-structured interviews have been transcribed and thoroughly read

several times, emerging themes and codes have been determined. The first of the themes is about embodying the concept of syntactic complexity on the minds of eight faculty members who have been scoring students' essays for about 16 years on average. To gain a deeper understating as to what syntactic complexity might be according to these interviewed instructors, we hereby divide the theme into two brad codes named associations and examples of syntactic complexity. We present the associations and examples of syntactic complexity and provide verbatim quotations to illustrate the emerging code.

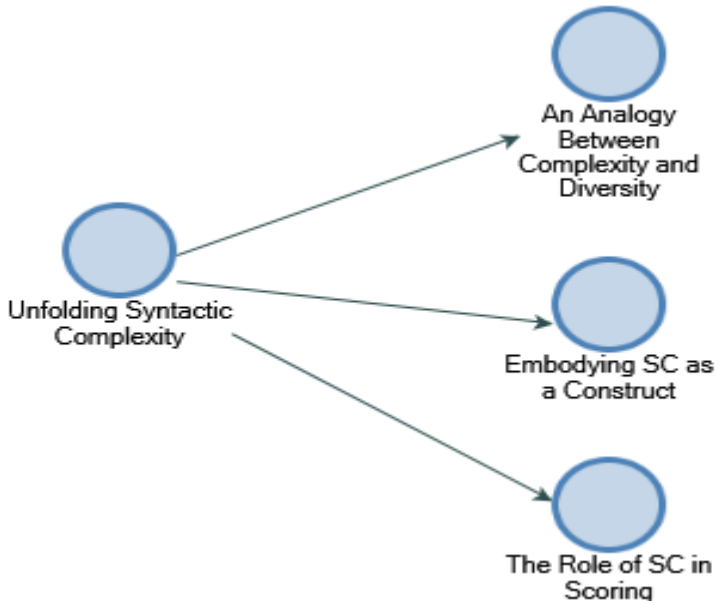



**Figure 4.1.** *The Thematic Display of 'Unfolding Syntactic Complexity'*

By associations of syntactic complexity, we mean the concepts to which the syntactic complexity refers and/or what syntactic complexity connotes in our experienced writing instructors. By examples of syntactic complexity, it is clear that we asked our instructors to exemplify syntactic complexity and we provide those examples under this code.

One of the instructors (Inst.1) thinks that variety in grammatical structures is the key element for understanding syntactic complexity. She states that the students should use a wide range of grammatical structures and try to go beyond what is common and well-known in terms of grammatical structuring. According to Inst.1, even fulfilling the necessities of essay organization alone is not enough to extract syntactic complexity:

> [the students] pay attention to the subject-verb agreement, use common and ordinary tense structures or conjunctions. They avoid using types of clauses or embedded structures, which

> naturally leads to a much simple work of written production. So, there is neither lexical variety nor syntactic variety; do not use echoic terms, antonyms and synonyms; though instructed they do not use any conjunctions apart from 'and', 'but' and 'so'. They think they can complete writing by just writing a topic sentence, a couple of ordinary sentences and a conclusion sentence, but of course this shouldn't be this way. (Inst.1)

Another instructor (Inst.2) likewise said that she associates syntactic complexity with using 'variety of structures'. She adds that the term syntactic complexity evokes an expected component in students' writings:

> Complexity in Turkish connotates something negative but in English as I can guess connotates something more positive. I think of complexity as using variety of structures. I also think that [students] should be able to use much more complex and compound structures. (Inst. 2)

Inst.2 also says "I make out of syntactic complexity the ability of a writer to manipulate the language". Through this remark of Inst.2, it is possible to deduce that syntactic complexity can be regarded as a tool of reflecting one's skills in language use.

Providing structural variety is one of the most prominent features that have been associated with syntactic complexity. One of the ways to ensure structural variety is to use clause constructions by means of conjunctions. Inst.5, as exemplified in the quotation below, states that syntactic complexity demands more than simple 'subject-verb-object' sentence construction:

> What does this term evoke? I am talking about the use of conjunctions and the transitions and about the use of clauses such as adverbial, adjective and noun clauses. To me, a complex writing includes the use of all these structures together; or else subject-verb-object ordered sentences are elementary structures. (Inst. 5)

One of the instructors (Inst.3) adds the length of the sentence as a feature of syntactic complexity. This length association is of importance in that Coh-metrix also considers length as a predictor of syntactic complexity and measures it through the number of words coming before the main verb of each sentence. Inst.3, similarly, thinks that adding clauses to the sentences is a way to prolong the sentences, however she points out that length in a sentence might not be alone for a sentence to be complex, but rather a sentence should also be grammatically accurate:

> It in fact reminds me of my own writing still. I am always criticized not to have written in a simple tone. They ask: can't you make short sentences? There is an abundance of clauses…a sentence creeps into a long one till it ends…however they are still accurate in terms of syntax and meaning. So I envisage of a writing style that involves more use of grammatical structure to get across the meaning (Inst. 3)

As can be seen from the displayed quotations, the interviewed raters have several associations on their mind regarding what syntactic complexity is. In this section as to make the concept of syntactic complexity more concrete, we asked the raters what syntactic complexity evoked on their perceptions. The most outstanding characteristics of the syntactic complexity s reported by the raters is the variety of grammatical structures, use of conjunctions, use of different clauses within sentences.

In our attempts to embody syntactic complexity, we also asked our raters to give examples to patterns or constructions which ensure syntactic complexity. In the verbatim quotations below we present the examples of syntactic complexity that our raters come across in the writing samples they score.

**Table 4.20.** *Examples that point syntactic complexity*

| **Examples that point syntactic complexity** |
|---|
| Relative Clauses |
| Reduced Relative Clauses |
| Adverbial Clauses |
| Adjective Clauses |
| Noun Clauses |
| Inversions |
| Idiomatic Expressions |
| Passive Structures |
| Phrasal Verbs |
| Embedded Structures |
| Conjunctions |

As table 4.20 above lists, there are a number of constructions that exemplify syntactic complexity. One of the instructors (Inst.2) said that *"*relative clauses, noun clauses, inverted sentences especially, inverted sentences with transitions and conjunctions" might be among the patterns that point syntactic complexity.

Inst.1 states that combining a variety of syntactic patterning can be possible by combining the clauses by conjunctions. Embedded structures and inverted sentences are among the constructions that make a writing less stereotyped, more varied and thus syntactically more complex:

> I mean they must be able to use inversion, embedded structures, and to bring a couple of sentences together using different conjunctions. For example, phrasal verbs, idiomatic expressions of course in concordance with academic writing. (Inst.1)

Inst.5 and Inst.8 express that using passive structures and using prepositions correctly is a sign of syntactic complexity. Inst.5 particularly emphasize the accurate use of conjunctions, and different types of clauses:

> For example, is there accurate use of inversions or the accurate ıse of structures like 'neither nor' and 'not only but also'. Could they convert adjective clauses to adjective phrases, or use participles in adverb phrases? These all count as complex structures for me. (Inst. 5)
>
> Of course they must be able to use clauses. Not only relative clauses but also adverbial ones and different transitions…use of passive structure…For example, I expect them to use –ing after proposition 'of' accurately and without being confused. (Inst.8)

Our instructors uncovered the links with which they associate syntactic complexity. The most outstanding association of syntactic complexity was variety. Instructors almost in unity thought that a wide range of syntactic constructions such as using clauses connected appropriate and different conjunctions is the key to understanding syntactic complexity. Our instructors also mentioned the length of sentences and the accurate use of syntactic constructions as the basis of complexity on their minds. Our interviewees in this section exemplified syntactic complexity. In other words, they clarified which kind of syntactic constructions would make a writing syntactically complex. To name those constructions: relative clauses, reduced relative clauses, adverbial clauses, adjective clauses, noun clauses, inversions, idiomatic expressions, passive structures, phrasal verbs, embedded structures, conjunctions.

### 4.5.1. The role of syntactic complexity in scoring the students' essays

In our analysis we focused on the question how syntactic complexity perceived by our instructors affects their scores. We already reported what kind of constructions would evoke syntactic complexity in our instructors' minds. As a reminder, those syntactic complexity examples were: relative clauses, reduced relative clauses, adverbial clauses, adjective clauses, noun clauses, inversions, idiomatic expressions, passive structures, phrasal verbs, embedded structures, conjunctions. Some of the instructors told that complexity in syntax is something they are looking for in their students' essays, while some other state that simple but accurate sentences would not bother them. Instructors

also think that syntactic complexity would contribute to the organization of the ideas in an essay, thus indirectly affecting the organization score as well.

Two instructors (Inst.3 and Inst.6) explain that they have high expectations from the English Language Teaching majors in the direction that these students should have a high level of language proficiency, thus reflecting this proficiency level in the form of syntactically complex sentences in their writings. These instructors state that only using simple but accurate sentences would not lead to high scores:

> We are telling them [our students] "you are going to be English teachers". So there must be a level of mastery. They must be showing us that they can use different forms and structures. If you are only using simple sentences, even if they are grammaticaly correct, you may not be able to get high grades. I expect that complexity. (Inst.3)
>
> If sentences are accurate but simple, they can not get high scores because what I expect from an ELT student is not simlicity (Inst 6)

Some other instructors (Inst.1 and Inst.2) value the students' attempts to use complex sentences even if they make mistakes in trying to do so. These instructors, as exemplified below, think that these students at least try to use what is instructed and what is expected from them as a proficient learner, so these instructors do not take points off from these mistakes resulting from the attempts of complexity:

> For example, one can excessively use "I like, I dislike… etc." on the other hand, another one tried to write many things even though made some grammatical mistakes. At least, he tries and takes a risk. This should be encouraged as well. This does not mean ever sentence should be graded high but a balance should be ensured. What I mean… the student should get the message to produce complex sentence structures (Inst. 1)
>
> It effects the score for the 'language' section. Sometimes they use the structures inaccurately, making direct transfers from Turkish. Still they are making an attempt to be complex. Even if inaccurate, [this effort] still important to me. But when used accurately, I try to score the 'language' section high. (Inst.2)

Some instructors we interviewed warn that the students' attempts to write syntactically complex sentences could lead to miscommunication, or errors of idea flow. When a conjunction or a particular structure is used improperly just for the sake of complexity and length, it can cause the sentence to be misunderstood or not to be understood at all:

> Meaning may be getting vague while trying to be complex. We may not understand the message. It should be appropriate. This a common problem we face in writing. For example, since the students knows he can get scores from using discourse markers, he can use discourse markers inappropriately (Inst.1)

> If the students make things intangible just to be complex, the writing gets intangible too. As the content disappears, of course they are graded lower. (Inst. 5)

One instructor (Inst.5) go even one step further and says that she is not even bothered by the use of simple sentences. She says that what she values the most is the correct flow of ideas without any hassles:

> Even though they don't necessarily relative clauses, even in a minimalist manner, if they can get across what they intend to, that's ok with me. I do not think why this student did not use a more complex structure. At least, he tried to get something across within the scope of his knowledge. (Inst.5)

Two instructors (Inst.2 and Inst.3) stress that syntactically complex sentences indirectly contribute to the organization of the essay. Through the complex structures, students could write well-organized and coherent essays and could display a sound and clear flow of ideas:

> The correct usage of the structures affect the organization too. For example, there are students who write very simple sentences though the content is heavy enough. I guess there is an effect up to 50% in the overall scoring. (Inst. 2)
>
> What do we expect from the students in terms of writing? We expect a certain structure of writing. For example, a thesis statement, supporting topic sentences and major and minor idea statements after each topic sentence. Therefore, we wish to see a flow of idea and we want that flow to be built by an enough number of different conjunctions and sentence types. (Inst.3)

Our instructors revealed how syntactic complexity would affect their scores. Some instructors said that they had high expectations from the ELT majors since they were considered as high proficient language learners thus they are supposed to display syntactic complexity in their writings. Our instaructors thought that syntactically complex texts are likely to have a well-organized flow. Organization score might be affected in a positive way if the syntactic devices are varied enough to connect and get across the ideas fluently. Students' attempts to use a complex grammar in their writings is valued by the instructors and generate a positive perception which is likely to bring about higher scores. Still, our instructors also exhibited a caution for the students. Our instructors warned the students to be appropriate and use syntactic constructions to the purpose. Using varied constructions just for the sake of being perceived complex, may lead to errors and confusion thus lowering the scores.

### 4.5.2. An analogy between complexity and diversity

Complexity and diversity are two highly complicated notions. In our study, we accommodate complexity to syntax, as we accommodate diversity to lexical knowledge. However, as also can be seen in several definitions we listed, the two terms are closely related, even inter-related. The mutual area that these two terms share can be named as 'variety'. Although diversity more clearly incorporates a meaning of 'being varied', complexity also has a similar meaning both in the literature definitions and in the interviews of our instructors; "As I said before we want them to use different structures, not to use only one kind of simple kind of structures. This also goes for the lexical diversity. (Inst.3)"

In the verbatim quotations above, Inst.3 visibly links the diversity and complexity notions to each other. She states that their expectations regarding the syntactic complexity depend on the range of different syntactic structures used, just likewise, she told that the lexical diversity should be ensured in the same manner.

Variety is the notion that incorporates both diversity and complexity notions. One instructor, while reciting the missing qualities of a student paper, mentioned the lack of variety both in terms of syntax and lexical knowledge; "So there is neither variety of words nor variety of grammar" (Inst.1)

### 4.6. Unfolding Lexical Diversity: Embodying It as a Construct

In the previous section, we presented the associations and examples of syntactic complexity. In this section, similarly, we attempt to embody the term lexical diversity, what it is to the instructors and from what kind of linguistic constructions or uses our instructors deduce lexical diversity. In our semi-structured interview, we first asked our instructors to envisage lexical diversity. We aim to get some deeper insights into the quest of what lexical diversity could mean and how it is considered and tracked by the instructors in students' writings.

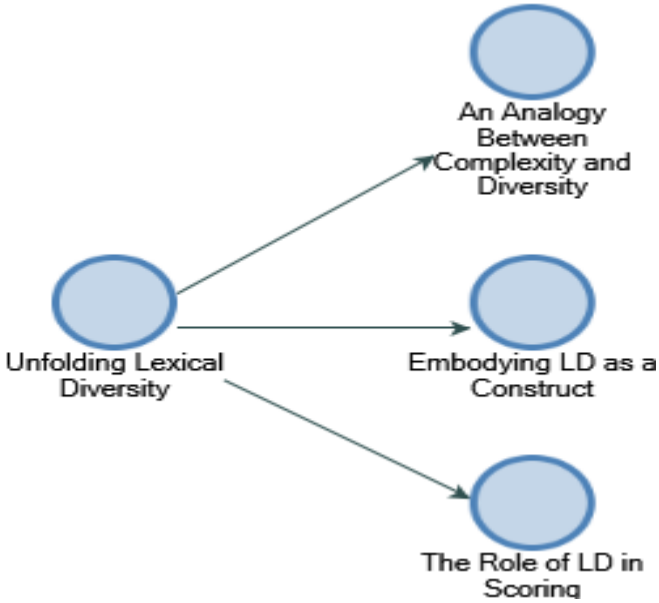



**Figure 4.2.** *The Thematic Display of 'Unfolding Lexical Diversity'*

One of the instructors (Inst.1) explained the concepts of vocabulary depth and breadth, claiming that vocabulary size, i.e. the mere number of words known to a learner, could not reflect much of the real knowledge. She especially focused on the depth dimension of the vocabulary knowledge and implied that knowing multiple meanings a word carried could signal the actual vocabulary knowledge. Therefore, from the quotation below, it is clear that lexical diversity can have much to do with depth knowledge:

> As I teach lexicology here, I know there are different terms covering vocabulary size. Vocabulary size alone is a shallow term. There are depth and breadth dimensions. Depth is associated with the several meanings of the same word and their usage depending on the context. (Inst.1)

As it is in depth knowledge, lexical diversity prominently evokes variety in use, too. Several instructors state that they demand different word types that belong to different word classes. They add that words that are infrequent in use, on the condition that they are used properly according to the topic and context, are concerned with lexical diversity. The instructors consider these types of infrequent words as advanced words:

> There can be complex words that are varied more advanced and upper-level. I think of something close the usage of such words. (Inst.6)
>
> If they have a large vocabulary on a specific subject, we are telling the students to display it in their writing. That's why we demand that they use both different word types and different words as much as possible (Inst.3)

In conceptualization of lexical diversity, what some instructors often refer to is the accurate use of words that semantically fit in the structural context (i.e the accurate use of parts of speech). Beside the accuracy, going beyond the '*commonly used words'* also prevails as an idea of lexical diversity:

> It reminds me of the difficulty level of the words. For example, does the student correctly use words apart from the commonly used ones (Inst.7)
>
> For example, are the words appropriate for the content? Or could the student use the words correctly, did they use adjective instead of adverbs, or adjectives instead of nouns? (Inst.5)

When it comes to the examples of lexical diversity, our instructors reported some vocabulary usage patterns as listed in the Table 4.21 below:

**Table 4.21.** *Examples that point lexical diversity*

| **Examples that point lexical diversity** |
|---|
| Noun forms |
| Synonyms |
| Antonyms |
| Phrasal verbs |
| Collocations |

Our instructors complain most about the monotonous usage of common words in student essays. The repetitive and frequent word use is seen contrary to lexical diversity since, as the name implies, lexical diversity is closely related with the wide range of words, wide both in meaning and number:

> It is about their ability to use advanced words with relation to content. For example, they may be always writing 'thing'. They should be for instance using 'reason' or 'compose' instead of 'thing'. This is diversity namely using the synonym. Appropriate usage of noun forms Is also a part of lexical diversity (Inst.2)
>
> There are academic synonymies, antonymies, collocations, idiomatic expressions and chunks. These are important. For instance, when we demand a cause and effect analysis essay, we want to see those causes and effects to be written. They should'nt always use 'first cause, or second cause'. They should be using different words like 'impact', 'influence'. (Inst.1)

Two instructors (Inst.3 and Inst.4) exemplified lexical diversity by pointing to color usage as a modifier. They reported that students who display lexical diversity in their writings could use synonyms of common colors and could write '*the shades of blue*' (Inst.3) while for example describing a wall. R4 proposed that instead of writing red all

the time to talk about a red flower, a student should use '*crimson or bloody red'* to diversify their lexical usage. Similarly, Inst.6 points out to the importance of using synonymy to exemplify lexical diversity and said: '*mesela important demez de significant der.*

Another instructor (Inst.5) also highlighted the importance of modifying the actions by using adverbs correctly:

> hmm especially their usage of adverbs is very important, their accurate and appropriate use… for example, these can be the structures they face in reading lessons, chunks, phrasal verbs. The use of these structures mean lexical diversity to me (Inst. 5)

We already reported the correct use of part of speech but in close relation with this, our instructors also stress the importance of variation in 'parts of speech' condition of the same word. In other words, our instructors state that a student should know the different versions or derivatives of the words. It is noteworthy to remember that this idea is very close to the depth dimension of vocabulary knowledge:

> I can say [lexical diversity is in the writings] which consists of advanced level vocabulary and perhaps noun forms. Noun forms of most verbs are accepted more advanced. Therefore, the use of less common vocabulary (Inst.2)
>
> Like the use of "you frustrate me" instead of "you make me frustrated". There is a difference here…[the students] should have a command of different forms of the same word like noun, verb and adjective (Inst.8)

In this section, our instructors unveiled the associations of lexical diversity to embody it as a construct and then gave examples of lexical diversity patterns. As the name implies, for a text to be lexically diverse, a diverse range of vocabulary should be used. The words should be broad in number and meaning since our instructors mentioned multiple facets of the vocabulary knowledge. This means that only knowing the meaning of word is not enough alone, but rather a student should know the multiple meanings and forms a word could carry. Using infrequent words, noun forms, synonyms, antonyms, phrasal verbs, and collocations is the key to understand lexical diversity.

### 4.6.1. The role of lexical diversity in scoring the students' essays

In our analysis, we found out that lexical diversity plays a role in scoring student essays. Our instructors reported that when they consider their students' writing as lexically diverse, they tend to give high scores. How lexical diversity reveals itself was exemplified in the previous sections. To remind, lexical diversity in a student text

manifests itself through noun forms, synonyms, antonyms, phrasal verbs, collocations and infrequent word usage.

Two instructors for example, state that using only the same words instead of providing varied items has a negative effect on the content score of the essay. As the content of essay is concerned with ideas related to the writing topic, the words chosen to express these ideas simultaneously gain importance. As the instructors expresses, wide range of vocabulary use can reflect a wide range of writing idea units, thus may even affect the content of the writing apart from the language features:

> […lexical diversity] I think has an effect on the language score. If the language is correct, of course this will have a positive effect on the content. There is a difference between one student always saying 'thing or cause' to express an idea. However if they use sometimes 'thing' and sometimes 'cause', this variety I suppose affect the score for the content. (Inst.2)
>
> If the students can use vocabulary effectively and properly, this sure will impact the content. I say to myself "how well he did wrote and expressed himself". Or else, they keep writing "thing, thing, thing, it is a thing, it is a thing" Or they use lots of relative clause. This shows me their vocabulary is inadequate and this sure has an impact on the content.(Inst.6)

One instructor explains that she takes great pleasure in reading lexically diverse written texts. When defined by the instructor, these texts have rare and uncommon words that are used accurately and properly. She says, although she does not take points off from frequent and ordinary words, that she believes her score intends to go up when infrequent words are faced:

> I still do not lower their score if the simple usage is also accurate however, I enjoy scoring the papers in which there are different words rather than the papers where the students constantly use the same things, very simple and commonly known words. This definitely has a positive effect in terms of high score. (Inst.5)

One instructor specifically highlights the accuracy of the word usage. She explains that the word can be different and infrequent but may not be appropriate to be used in that particular context. When used improperly, these words can lead to miscommunication

> I take [lexical diversity] into consideration however if the words are inaccurate, it isn't worthy. Because when the meaning is inaccurately conveyed, there is miscommunication or lack of communication. The word is good, different but inappropriate for that specific context…then it is called 'wrong word usage' (Inst. 7)

On the other hand, another instructor states that the students' attempt to use varied vocabulary is of importance to her. She says if the students look up at thesaurus to find synonyms or antonyms, and even if the usage is not semantically correct, she values this

attempt and does not take points off. The instructor also explains that she suggests the students look the new words up in the dictionary for a meaning check:

> This is one of the things I give extra points, because you understand that he got curious and looked up. An effort…For example, he may go and resort to Thesaurus, but still uses the word inaccurately. I do not lower his score, I tell him not to use every new word just randomly but do not lower the score since he shows a serious amount of effort (Inst.5)

Concerning lexical diversity's role in the overall scores, our instructors mention variety and appropriate use. Using a wide range of words is an expected criterion for higher scores. Looking up at thesaurus and dictionaries to explore synonym and antonyms or collocations to be lexically diverse is valued a lot by the instructors. As long as the students use the correct word in the correct place in the context, the instructors tend to give high scores.

### 4.7. General Outlook of the Scoring Procedure

#### 4.7.1. How do our instructors start scoring and continue?

Our qualitative analysis also unfolded the process that flow in the scoring procedure taken by our human instructors. The general approach to scoring adopted by our instructors is the central issue which is addressed hereby. Our analysis and the quotations given verbatim here display that our instructors seek what they teach in students 'papers concerning writing in their classes. What our instructors also report about the process is that they are tightly stick to the rubric and criteria which were agreed upon. The quotations below clearly illustrate that our instructors expect the students' writings follow the genre rules which they made clear in the classes. Among the rules to be sought in student papers, beside many others depending on the text type, 'thesis statement' sentences are undoubtedly the most significant ones. Out of the instructors' general approach, we can also deduce the way they view writing in English or the way the writing is instructed in the program. Writing an essay in English seems like filling up a form, without neglecting the 'absolute musts' such as thesis statement sentences, topic sentences, the structure and order of paragraphs and minor/major details and examples that support the main idea;

> First I read the whole essay. Then I read again based on our criteria. For example if I should look into the thesis statement [according to the criteria], I once again control the thesis statement. Similarly, then I check the developmental paragraphs, topic sentences, major and minor details (Inst.1)

> Our general approach depends on the genre type but we have different criteria for scoring thesis statement, topic sentences along with language, content, organization and coherence. Thesis statements and topic sentences are very important. (Inst.2)

We, in our analysis, also found out that after our instructors read the student papers once with an intention of broad scanning, they read the papers again. The first reading without paying the utmost attention intends to gain a general view of the writing. In their first reading, the instructors sometimes take some notes to consider afterwards, or they just pinpoint the parts they regard important. They do so with or without the rubric in hand;

> First I read the essay without following any criteria but just to wholly understand it, and to get a first impression. Then I, for more detail and with the help of the criteria, read thoroughly again (Inst.5)
>
> While scoring, I read the papers twice. Firs I skim the essay. Then, [in second reading] I write the points of scoring to the edges of the related parts following the criteria (Inst.3)

The second or sometimes the third round of reading aims to fully apply the rubric criteria and involves the actual stages of scoring;

> First I score content and organization, then grammar, vocabulary, mechanics and punctuation. I once more read for editing. I look over one paper for three times (Inst.7)
>
> …. Then I slowly and along with the criteria read once again (Inst.7)

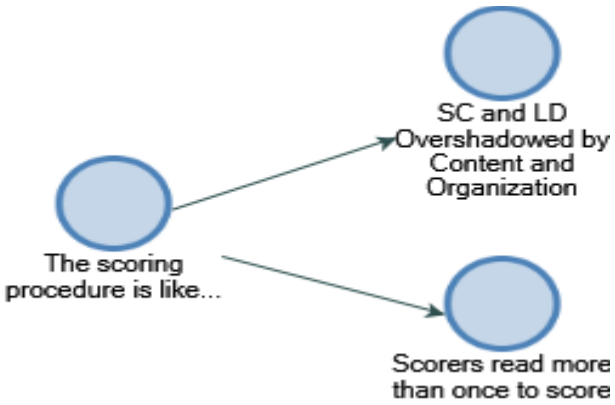



**Figure 4.3**. *The Thematic Display of 'The General Outlook of Scoring Process'*

### 4.7.2. Content and organizational patterns overshadow syntactic complexity and lexical diversity

In all writing assessment rubrics, it is quite likely to see references to the content and organization of the writing. Rubrics have separate divisions for the evaluation of content and organization. Content refers to writing topic and the ideas and examples to

support the main writing topic. Relevance of ideas and development of those ideas is, for example, a closely related issue with content. As for organization, it is concerned with the flow or display patterns of the ideas and topic. The writing devices used in the developmental path of the ideas are among the keys to a well organization in writing. While content must be coherent, the organization is supposed to be cohesive. Some rubrics evidently refer to coherence and cohesion. In our study, likewise, many several instructors state that they attach great significance to content and organization dimensions of students' writings. Our instructors, as they reported in interviews, make an order of importance on their minds while reading student papers and in this order of importance, content and organization come first, leaving SC and LD behind. One instructor (Inst.7) told that "an essay written with a good command of English can make me suppose that the content is also well developed, thus at the very beginning I divide these dimension from each other". The below given verbatim quotations exemplify the point;

> I generally start scoring the content. I love scoring with a focus on content and organization. Because the mechanic part of the writing can affect me negatively (Inst.7)
>
> First of all, I look into the content. And then I look into our expectations. For example, is there what there should be in an opinion essay? I review again like this and lastly I look into grammar, spelling, and punctuation. (ınst.5)

Inst.5, one of our most experienced instructor stated that complexity notion embodies itself as a complexity of content for herself. She told that she widely considers how well developed the thoughts are in student writings along with the necessities of the text type;

> When told complexity, I mean I more often look into the complexity of the content rather than the structural complexity… how the ideas develop and how well they flow…For example, you give a picture and tell the students to write a descriptive essay, however the final work is not descriptive at all (Inst.5)

The same instructor, which is Inst.5, clarifies that even well-developed and coherent ideas may not be enough if those ideas are not connected to each other and thus if there are logical gaps. The quotation below shows that R5 prefers the student writings to be correctly and neatly organized with a smooth flow of ideas around the main topic; However, even if the student has good ideas, he might have logical flaws while displaying those ideas. I think I first look into the organization. Then comes content (Inst.5)

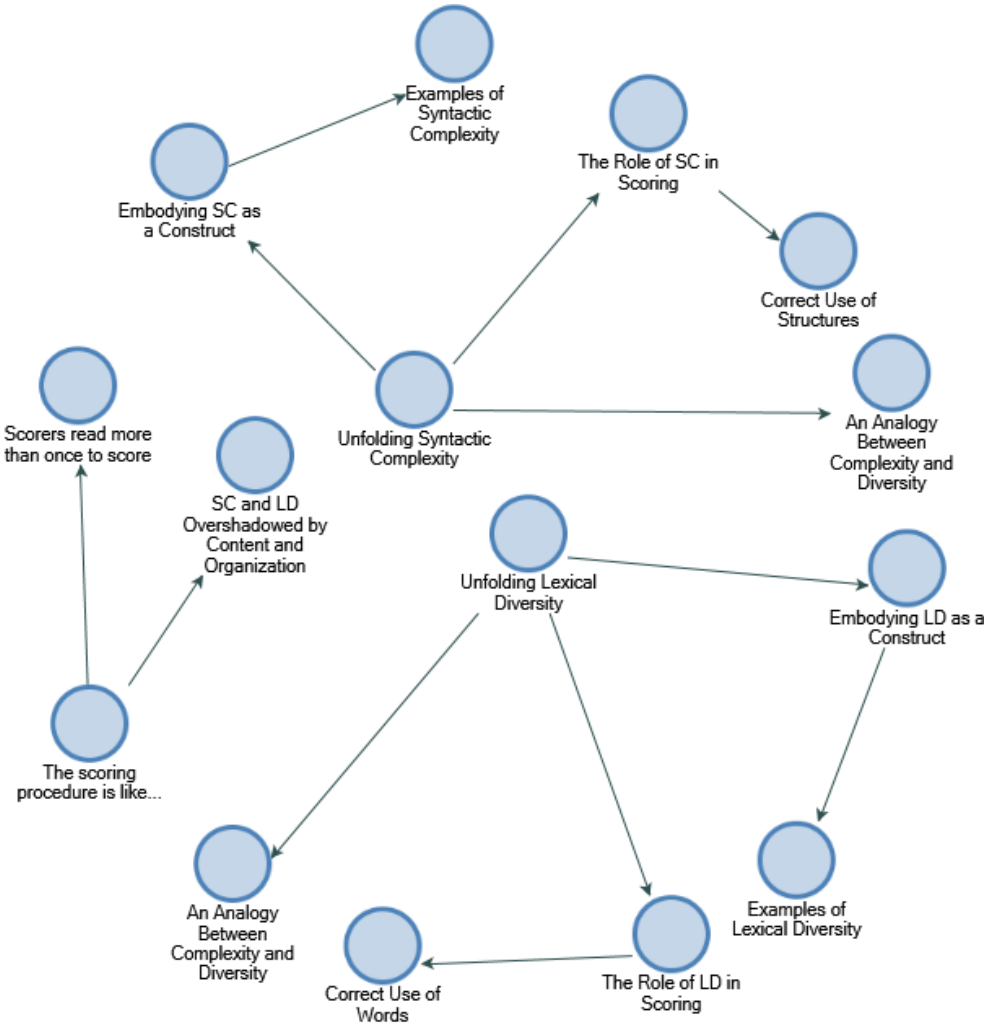



**Figure 4.4**. *The Thematic Display of 'Qualitative Results Summary'*

As displayed in Figure 4.1 above, our quantitative findings uncovered the contents of syntactic complexity and lexical diversity as constructs that have part in foreign language writing assessment which is carried out by human scorers. What the findings illustrate is the examples of SC and LD, that is to say, to human scorers' view. The findings also show a relationship between SC and LD in the perception of our scorers as well as the similarities and differences between two constructs.

Additionally, the same findings also put forward insights into the scoring procedure and how it is handled by human scorers. The findings also, as exemplified verbatim and in detail in previous sections, pose the perceived importance of content and organization in the process of assessment of student writings in foreign language.

# CHAPTER 5

## 5. DISCUSSION, CONCLUSION AND IMPLICATIONS

This sections provides an indicative summary of the study with a short view of methodology and with an emphasis on our quantitative and qualitative findings. The section also includes a discussion of findings along with inferences of conclusion, pedagogical implications for practitioners and further research suggestions for other scholars.

### 5.1. Summary of the Study

Our current study chiefly aims to unfold the ELT majors' written compositions and explore the relationships and variances at play. 204 ELT majors wrote mainly one page long compositions on a pre-determined topic. Half of the participants were first year students while the other half were fourth year students. The purpose of this sampling was to divide the students on a language proficiency basis.

The variables the relationships of one another is examined in students' compositions are namely; syntactic complexity, lexical diversity and writing quality. Syntactic complexity and lexical diversity were extracted from Coh-Metrix – an automated text processing tool-in terms of 5 different indices. As for writing quality, scores were obtained by two scoring procedures conducted by our raters with a 6-month-time lapse. One of the premises of our study is to explore the perceptions of experienced instructors regarding syntactic complexity and lexical diversity. With a qualitative research paradigm, we asked the isntructors what they recall out of these variables and how they conceptualize it.

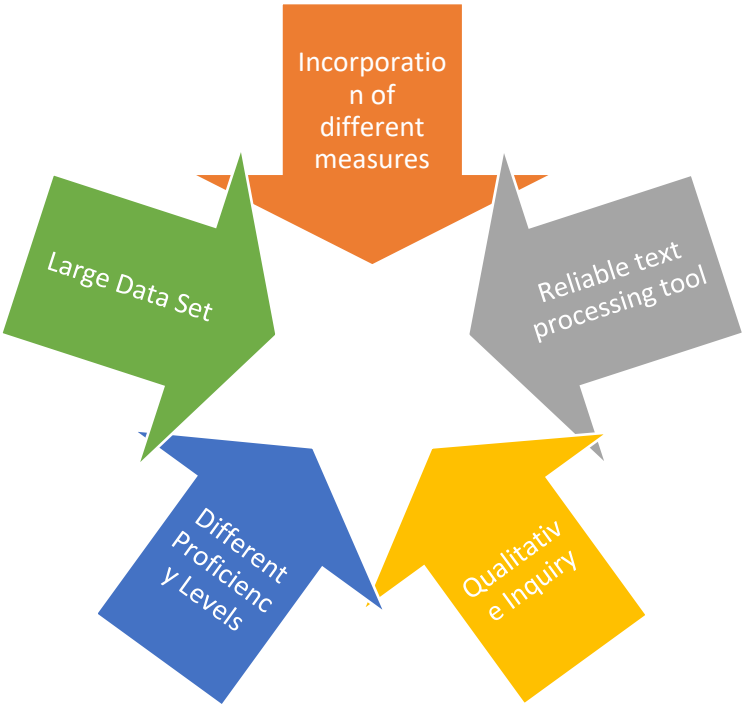



**Figure 5.1.** *The Thematic Display of Research Input*

The collected data, i.e. student compositions, were first descriptively analyzed in terms of word count and word range. This descriptive analysis posed 'text length' as a confounding variable, which was found statistically significant in further analysis and in qualitative findings. After this descriptive analysis, the student compositions were uploaded on Coh-Metrix. It offered numerical and concrete data concerning syntactic complexity and lexical diversity, which would otherwise be so abstract. The numerical data regarding syntactic complexity and lexical diversity as well as text length were tested with paired and independent samples t-tests to see the curricular level differences. Afterwards, the results were correlated among each other for the relationships. Lastly, for the variance explanations, a three-layered multiple regression test was run.

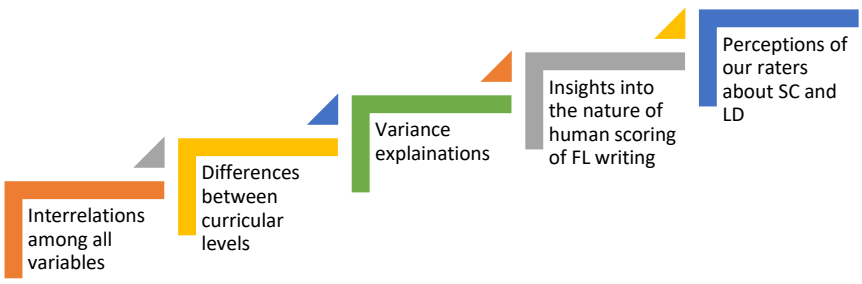



**Figure 5.2.** *The Thematic Display of Research Output*

After statistical tests have been carried out, we obtained the interrelations between all variables, the differences between curricular levels, the variance explained by the dependent variables. As a result of the qualitative analysis we had insights into the nature of human scoring of FL student writing as well as the perceptions of our instructors about syntactic complexity and lexical diversity. Figure 5.2 above displays the categories of findings we obtained at the end of our analysis.

## 5.2. Uncovering Syntactic Complexity and Lexical Diversity

Specialists on various fields have been popularly investigating complexity and diversity schemes since 1990's and endeavoring to define these concepts, which is far from being simple. Defining complexity and diversity in Second Language Acquisition field also attracted much of interest starting in the same years and continuing afterwards (Bulte and Housen, 2012, 2014; Forster and Skehan, 1996; Laufer and Nation, 1995; Lennon, 1990; Malvern et.al., 2004; Olinghouse and Wilson, 2013; Ortega, 2003). In the studies cited so far and in the upcoming lines, it is possible to see that there is not a consensus on the content of complexity and diversity in foreign language output. Among the suggestions cultivated through years are "using a wide range of structures and vocabulary" (Lennon, 1990, p. 390), "progressively more elaborate language and a greater variety of syntactic patterning" (Forster and Skehan, 1996, p.303). As can be seen, concepts of complexity and diversity in language output have largely intertwined, which means that diversity incorporates diversity while diversity incorporates complexity as well. The intertwined nature of these two highly abstract concepts have also confirmed in our qualitative analysis since some scorer participants of the present study also associated complexity and diversity. The common point where complexity and diversity meet is variety. That concept of variety involves both complexity and diversity has been verified in several research studies which attempted to define these concepts (Bulte and Housen ,2012, p.22; Rescher, 1998, p.17; Wolfe-Quintero et al., 1998) as well as in our results. We also presented similar qualitative findings and hereby state that in order for a piece of student writing to be syntactically complex, it should involve a wide range of syntactic patterns and a variety of different vocabulary items for it to be lexically diverse. Our study produced compatible findings with the 'global view' of complexity suggested by Bulte and Housen (2012). In that approach to linguistic complexity, the number, range, width, or repertoire of both grammatical and lexical items known to the student forms the central

point and they are not two separated poles, rather closely interrelated factors in understanding complexity.

Some previous researchers put forward that syntactic complexity should contain a number of simple constructs together with complex ones. In other words, a writer can produce a syntactically complex writing when he uses a balanced harmony of simple and complicated sturctures (Pallotti, 2015; Wolfe-Quintero et al., 1998). As previously reported in qualitative finding sections of this study, some experienced scorers highlighted that some students' attempts of complexity brought in misunderstandings and miscommunication in their writing. However, it does not necessarily mean that students should not try to ensure syntactic complexity in their writings. Likewise, another dimension about syntactic complexity both confirmed by previous studies and by our study is that syntactically complex writings are one of the requirements that teachers and scorers expect from high proficient learners (Bulte and Housen, 2014, p.46).

As for the approaches to understanding and defining lexical diversity, our study proposed -in the light of qualitative findings- that lexical diversity in FL student writing is primarily composed of using as many different and less known words as possible. Abundance of words appearing in a student text was found to be the key to understanding lexical diversity. This conclusion was affirmed by a bulk of previous research. Different names were given to lexical diversity so far; among them were 'lexical variation' (Engber, 1995), 'lexical density' (O'Loughlin,1995), "a combination of lexical variation and lexical sophistication" (Laufer, 2003, p.24), and 'lexical richness' as coined by Daller, von Haut and Treffers-Daller, 2003). However, all of these different characterization depends on the abundance of words.

### 5.3. Issues of Syntactic Complexity with Regards to Scoring, Scorers and Indices

Syntactic complexity has been popularly regarded as one of the trivets on which the assessment of FL writing and writing proficiency is based. One of the most prominent findings of the current study is that while the generic language proficiency arises, the students' writings become syntactically more complex. This finding has parallels with several previous research which suggested more proficient learners with more time and exposure to language could write more complex pieces of writing (Stockwell and Harrington, 2003; Stockwell, 2005; Mazgutova and Kormos, 2015; Norris and Ortega, 2009; Johansson and Geisler, 2011; Vyatkina, Hirschmann and Golcher, 2015) In

literature, it has been suggested that possessing the knowledge of more complex and particular grammatical structures might enable the learners to produce more complex ideas and peculiar expressions (Beers and Nagy, 2009).

Text length has been largely associated with writing quality and assessment. In our study as well, text length appeared as a significant variable that had a play in writing quality scores of our participants. However, contrary to our findings, some research found that more proficient leaners could pack more complex ideas into smaller sentences, thus producing smaller or shorter texts (Becker, 2010). On the other hand, Bi and Jiang (2020) rather more recently considered text length as an indicator of syntactic complexity and found out that text length together with complex nominals per clause, and clauses per T-unit as the best predictors of human judgements of 410 narratives of Chinese EFL learners. Therefore, it is possible to claim that text length in terms of syntactic complexity has an ambiguous nature as in our study we found out a moderate positive correlation between text length and writing quality scores.

As for the relationship of syntactic complexity and writing quality scores, our study which was carried out in a foreign language (FL) context could only pose weak correlations between syntactic complexity and writing quality. This finding contradicts with a number of previous research in literature. On the other hand, we should remember that comparing the studies on complexity issues needs much attention partly due to a lack of uniformity in the complexity measures and more importantly due to lack of a clear definition of the complexity construct (Bulte and Housen, 2014).

In line with our study, "nominalizations, attributive adjectives, and prepositional phrases" (Beers and Nagy, 2009, p. 187) were found to be visible in evaluating syntactic complexity of written pieces. Likewise, we also found -though very weak- a positive correlation between number of modifiers (as an index of SC) and writing quality.

In a seminal work of research synthesis, Ortega (2003) concluded that in syntactic complexity and writing relationship research which was conducted in second language (ESL) settings, participants generated more complex writings compared to those in the studies conducted in FL instructional settings One reason for this could be the differences between FL and ESL instructional settings. As suggested by Ortega (2003), in FL learning environments learners might not have as an old experience of learning a language as in ESL settings, which may be hindering the fast development of learners in FL settings.

Another reason for the weak correlations between syntactic complexity and writing quality might be the individual beliefs and approaches of human scorers to complexity in writing. As can be understood from our participating scorers' remarks, some demand and seek for syntactic complexity from their students as some do not and value the simplicity and accuracy more. Moreover, general impressions of human scorers, even if they follow a standardized criterion, are more prone to detect some organizational and content issues of writing. Human scorers might be overlooking the details and delicate signs of syntactic complexity. On the other hand, automated text processing tools like Coh-Metrix in our case, can well detect and calculate syntactic complexity in a computerized certainty. Therefore, we need to highlight that the weak and low correlations are among the overall scores of human scorers and several individual indices drawn from a computerized text processing tool. In addition, human scorers might have different expectations from their students' writings in terms of the number and nature of examples given, or the genre specific rules to be followed. Whereas, automated text processing tool do not hold any judgements or expectations, but rather only calculates syntactic complexity based on a number of pre-ordered indices.

### 5.4. Issues of Lexical Diversity with Regards to Scoring, Scorers and Indices

Several studies in the literature already assert that lexical diversity could track the learners' overall language proficiency and change across different proficiency groups. These studies started from comparisons of native groups, native/non-native groups, in ESL groups and lastly in FL learner groups. For example, Olinghouse and Leaird (2009) compared the lexical diversity scores of native English language learners of different curricular levels in a public school in US. Some earlier studies which compared lexical diversity of written texts were conducted between native and non-native groups of English learners (Linnarud, 1986; Harley and King, 1989). Others were conducted in short term (Bulte and Housen, 2014) or long term (Mazgutova and Kormos, 2015) ESL language programs and with learners of English of different L1 backgrounds (Jarvis, 2002, Yu, 2009). In all of these studies, lexical diversity was found to be developing over time and with more exposure to language through instruction. Likewise, our study produced similar findings in that our 4$^{th}$ year students wrote essays which were lexically more diverse than those of our 1$^{st}$ year students.

Another finding of our study was that syntactic complexity and lexical diversity were likely to be overlooked by human raters in overall assessment. Human raters, unlike automated tools designed to calculate these features, could miss the indices of syntactic complexity and lexical diversity, but instead, focus on organization and content of student writing. When essays of same length and topic were rated by a group of raters, they found lexical errors as the most irritating error, however, they also stated that they lowered their students grade more if the content and organization in writing is poor (Santos, 1988). That is to say, as our study also confirms, human raters consider content and organization more important and gives higher scores to well-developed content and organization even if there are some minor or major syntactic or lexical errors. We, therefore, can claim that syntactic and lexical accuracy rather than diversity is more visible to the eye of raters.

As for the lexical diversity and its relationship with overall FL writing quality scores, our study showed only a weak and positive correlation, though statistically significant, with Vocab-D measure of lexical diversity and overall quality scores. The other index of lexical diversity (MTLD) could not yield any statistically significant correlation. These findings accorded with several previous studies. In literature, there were studies which produced statistically significant and positive correlations between LD and FL writing quality as well as the studies which did not. For example, as for predicting overall writing quality, D-value exerted a weak and non-significant correlation in Bulte and Housen (2014). Likewise, Engber (1995) also put forward a non-significant and low correlation with writing quality scores (r=.23), which means that "percentage of lexical words has little, if any, relationship to quality" (p.148). Similarly, in a study with English learners of different L1 backgrounds, Jarvis (2002) presented, though moderate, a significant and positive correlation only between Swedish students' lexical diversity and writing scores. The same study, however, showed statistically non-significant and low correlations between lexical diversity and writing scores of American and Finnish students.

There were, of course, previous studies which contradicted our findings. In other words, a number of studies found a positive and moderate or strong correlations between lexical diversity and writing quality scores. However, the methodology of each research study was different. For example, Crossley et.al.,(2010) broadly characterized lexical diversity as a knowledge "breadth of lexical knowledge, depth of lexical knowledge and the accessibility to core lexical items" (p.1). These three broad categories were measured

through 10 different incidences provided by Coh-Metrix and the findings produced a positive correlation (r=.66) between these broad categories of lexical knowledge and writing quality scores assigned to 240 essays.

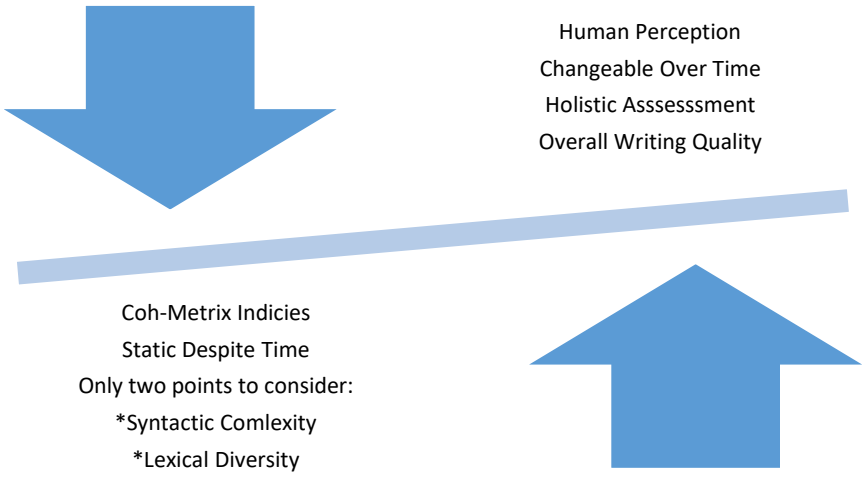



**Figure 5.3.** *Major Differences in the Nature of Comparison*

The biggest challenge in our study, contrary to the studies that contradict our findings, was that we tried to operationalize lexical diversity and syntactic complexity as independent constructs. We tried to build up well-established boundaries to embody these knowledge types, making them concrete and measurable by a computational tool. The reason behind this was that we would compare the findings with human scorers' scores and perceptions. It is already likely to numerically measure these knowledge types in writing, however, when you would compare them with human perception and human scorers' grades, they are most likely to be weakly correlated (if any) since the two parties have different qualities in nature. As illustrated in Figure 5.3 above, on one hand, there is Coh-metrix's different indices which were each formulated to measure lexical diversity and syntactic complexity, and on the other hand there are overall scores assigned to student writing by human raters through holistic scoring.

## 5.5. Issues Regarding Text Length in FL Writing Scoring

Text length is associated with the number of sentences, number of words and the length of each sentence and word used. In our study, text length was calculated based on the total number of individual words used in each student essay. Therefore, text length

can be regarded to be correlating with syntactic complexity and lexical diversity since the longer sentences, the more syntactically complex they might be. In our study, we facilitated from three Coh-Metrix indices to measure syntactic complexity, and two of them were indirectly about the length of individual sentences, and thus about the total text length. These two measures were namely 'Number of Modifiers Used Per Noun Phrase' and 'Number of Words Coming Before the Main Verb'. Likewise, these two Coh-Metrix measure, though weakly, correlated with text length on a positive and statistically significant scale. Number of words and the word length were found to be among the five strongest predictors of FL writing quality by Ferris (1994). Additionally, Friginal et. al. (2014) posed that longer texts with bigger number of words displayed more diverse vocabulary.

First, text length in our study was the variable which produced the strongest correlation with human scorers. We found a moderately strong and positive correlation which was statistically significant between text length and writing quality scores assigned by human raters. This finding is likely to be arising from that it is comparatively easier for human scorers to detect and evaluate text length. As our participating scorers stated, scorers might read the student essays more than once to evaluate it from several respects and one of these respects could possibly be the text length since it can be caught even with a glimpse of eye. Similarly, Jarvis et. al. (2003) found out that text length positively correlated with all 21 linguistic features of 160 ESL and 178 EFL student essays which were assigned high scores by human raters.

Second, text length as SC and LD, increased across different proficiency groups as put forward by Ferris (1994) and across different curricular levels in our study. Our fourth-year students with more time and exposure to language outdid the first-year students in text length as well as in indices of SC and LD. The difference regarding the text length between curricular levels was found to be statistically significant in independent samples t-tests. Thus, our study confirmed that FL learners could write longer texts when their generic language proficiency increases.

Lastly, text length, in our study as a confounding variable, also explained the variance in writing quality scores on a significant scale. Both alone and together with SC and LD on three-faceted model, text length explained the 20% and 24% of the variance respectively. Mellor (2011) also yielded similar findings in his study. Mellor (2011) wrote that "lexical diversity together with text length can more accurately predict essay quality

than either feature alone in this set of essays" (Mellor, 2011, p.9). Essay length, however, was found superior over lexical diversity indices in predicting essay quality.

### 5.6. Conclusions and Pedagogical Implications

With regard to the findings of the current study, a number of pedagogical implications for the education specialists and teachers are suggested in this section. The implications of the present study are concerning the scoring procedures applied for the FL student writing and the scorers themselves. Traditionally, academic writing in second language has been analyzed in two main types; articles and dissertations published for academic purposes and university students' writings. Homework papers, essays, and written tests are among the sorts of student writings (Hinkel, 2002). Since the number of student academic written work prevails in number, it is meaningful to analyze these scripts in terms of linguistic features and more importantly to explore how these features are related to scoring of these scripts.

First of all, our findings state that neither syntactic complexity nor lexical diversity- as individual aspects of writing- could be significant from the scorers' point of view. That is to say, human scorers regard content and organization features of a text as an inseparable part of the evaluation and they do not draw clear boundaries among these linguistic and rhetoric features of a text. Thus, in the light of our findings and the literature, we should note that human scorers must pay equal attention to the parts of the rubric they follow. Human judges should not prioritize one aspect of assessment over the other. The degree of variety in grammatical structures, the proper selection of vocabulary as well as adequate organizational support and appropriate rhetorical structure altogether are counted among the requirements of a qualified text (ETS, 2000). MELAB particular highlights that in order for an essay to be highly rated, the essay topic should be "richly and fully developed" and it should pose "a flexible use of a wide range of syntactic structures" (MELAB Technical Manual, 1994, p.7) as well as a large diversity and a proper use of words. As seen, in the criteria of standardized language tests, linguistic and rhetoric features of a text are on a balanced distribution.

Secondly, it is possible for the human scorers to utilize some automated language processing tools. Coh-metrix is one of those computerized written text processing tools which evaluates uploaded texts in terms of a number of textual and linguistic features. There are a wide array of studies that confirmed the reliability and validity of Coh-metrix.

Human scorers could use a tool as such when they have difficulty in distinguishing some linguistic features. As our findings suggest, human scorers are likely to miss syntactic complexity and lexical diversity indicators since they do not have the necessary time and attention as an automated tool. As a result, they only consult to text length and obvious mistakes. Likewise, as Santos (1988) and Davidson (1991) pointed out, the perceptions of human scorers and thus their scores are much affected by topic development and grammatical accuracy. Therefore, what we recommend for the human scorers is to train themselves to catch more delicate linguistic features like complexity and diversity, otherwise use automated tools that does the same thing with precision.

Thirdly, the teachers should be trained about the different approaches and ways of scoring a student paper. The scorers in our study must be resorting to holistic assessment in FL writing of our students. In consideration to scoring procedure, it is wise to conclude from our findings that scorers should pay due attention to holistic scoring. Crossley et. al., (2010) proposed three ways to quantify FL writing quality, one of which was 'primary trait' scoring. In this scoring approach, scorers primarily mark "rhetorical situations (e.g. the purpose, audience and assignment)" (p.284). What we observed during interviews with our scorers was that they performed a 'primary trait' scoring, therefore taking the content and organization into the upmost consideration. In a recent study Casal and Lu (2021) benefited from an targeted instruction for syntactic complexity in a sex week long academic writing course. The researchers exploited "explicit instruction of linguistic concepts, group and individual analysis of corpora and sample texts, and works and reflective discussions regarding personal writing projects" (p.99). As a result, the study concludes that targeted instruction for syntactic complexity results in increased awareness regarding the use and perception of complexity in writing.

Fourth, teachers should be mindful of the reliability of their scoring. As we performed an inter-rater and intra-rater reliability checks it was possible to see to which degree the scorers were consistent across two different holistic scoring procedures and among themselves. The similar applications of reliability checks should be frequently conducted. White (1984) pointed out that "holistic scoring is able to achieve acceptably high reliability" (p.403). Six different approaches were proposed by White (1984) to ensure the reliability of holistic scoring. For example, White (1984) suggested that raters should often come together to compare their scorings of the same paper and discuss to resolve possible discrepancies. Our findings also propose that scorers should put down

the rate of consistency among each other to determine the most reliable scorers. What White (1984) and our qualitative findings mutually recommend is that scorers must try out the criteria to rectify their expectations and to make required adaptations if necessary.

Our study, in consideration to its findings and the related literature, proposes that scorers should pre-determine their expectations and try to match them both with the criteria they follow and with their partners. Weigle (2002) wrote, for example, "a certain script might be given a 4 on a holistic scale by one rater because of its rhetorical features (content, organization, development), while another rater might give the same script a 4 because of its linguistic features (control of grammar and vocabulary)" (p. 114). Thus, the teachers and scorers teaching the same classes should frequently meet each other to discuss and decide which features (or linguistic) contribute more to the writing quality rated by human judges.

### 5.7. Suggestions for Further Research

This study aims to investigate the relationship with syntactic complexity, lexical diversity and FL writing quality scores assigned by human judges by uncovering the correlations and the extent to which syntactic complexity and lexical diversity account for the variance in FL writing quality. Second, the present study aims to find out if syntactic complexity, lexical diversity and writing quality scores of FL students change across different proficiency levels. Third, we aim to see the extent to which ELT instructors are aware of SC and LD in their scoring procedures. The upmost difficulty in designing such a study is to embody the somewhat abstract constructs like syntactic complexity and lexical diversity. Thus, the researchers should note that they use an automated tool to make these abstract constructs concrete in forms of numeric and measurable data. Additionally, there is an e-rater engine functioning in the body of ETS which can assign a total writing quality score to the uploaded papers. If future studies could access to these kind of automated engines, the comparisons between human judges and the tools would be more meaningful.

When we tracking the developmental path of syntactic complexity and lexical diversity, we compare the 1$^{st}$ year and 4$^{th}$ year university students studying in an ELT major. We accept the curricular level as the base of proficiency as the time and exposure to language might differ in a two-year time lapse. It would be plausible if future researchers could use a standardized test to their participants to determine proficiency

levels. What is more, it would be wise to conduct studies to include lower proficiency groups to depict a clearer picture of the developmental path.

We conducted a semi-structured interview with 9 scorers who have been frequently scoring student essays. The number of scorers could be increased to sort them based on their experiences and expectations from an academic student paper. Our student papers were scored by two independent human scorers; one with over 30 years of experience of scoring and near-native proficiency and the other a native American English instructor who have been working in Turkey for eight years at the time of study. The inconsistencies were resolved by a third rater who is also a native American English instructor. In this way, future researchers could probe more into the characteristics of human scorers, which is most likely to affect the scoring procedure.

## REFERENCES


Ai, H., & Lu, X. (2013). A corpus-based comparison of syntactic complexity in NNS and NS university students' writing. *Automatic treatment and analysis of learner corpus data*, *59*.

Becker, A. (2010). Distinguishing linguistic and discourse features in ESL students' written performance. *Modern Journal of Applied Linguistics*, *2*, 406-424.

Beers, S. F., & Nagy, W. E. (2009). Syntactic complexity as a predictor of adolescent writing quality: Which measures? Which genre?. *Reading and Writing*, *22*(2), 185-200.

Bi, P., & Jiang, J. (2020). Syntactic complexity in assessing young adolescent EFL learners' writings: Syntactic elaboration and diversity. *System*, *91*, 102248.

Biber, D., Gray, B., & Poonpon, K. (2011). Should we use characteristics of conversation to measure grammatical complexity in L2 writing development?. *Tesol Quarterly*, *45*(1), 5-35.

Bulté, B., & Housen, A. (2012). Defining and operationalising L2 complexity. *Dimensions of L2 performance and proficiency: Complexity, accuracy and fluency in SLA*, *32*, 21.

Bulté, B., & Housen, A. (2014). Conceptualizing and measuring short-term changes in L2 writing complexity. *Journal of Second Language Writing*, *26*, 42-65.

Casal, J. E., & Lu, X. (2021). 'Maybe complicated is a better word': Second-language English graduate student responses to syntactic complexity in a genre-based academic writing course. *International Journal of English for Academic Purposes: Research and Practice*, *2021*(Spring), 95-115.

Cohen, A. D. (1994). *Assessing language ability in the classroom.* Boston, MA: Heinle & Heinle Publishers.

Coirier, P. (1996). Composing argumentative texts: Cognitive and/or textual complexity. In G. Rijlaarsdam, H. van den Bergh, & M. Couzijn (Eds.), Theories, models and methodology in writing research (pp. 317–338). Amsterdam: Amsterdam University Press.

Covington, M. A., He, C., Brown, C., Naci, L., & Brown, J. (2006). How complex is that sentence? A proposed revision of the Rosenberg and Abbeduto D-Level Scale.

Creswell, J. W. (2002). Educational research: Planning, conducting, and evaluating quantitative. Prentice Hall

Crossley, S. A., & McNamara, D. S. (2009). Computationally assessing lexical differences in L1 and L2 writing. *Journal of Second Language Writing*, **18**, 119-135.

Crossley, S. A., & McNamara, D. S. (2010). Cohesion, coherence, and expert evaluations of writing proficiency. In S. Ohlsson & R. Catrambone (Eds.), *Proceedings of the 32nd Annual Conference of the Cognitive Science Society* (pp. 984-989). Austin, TX: Cognitive Science Society.

Crossley, S. A., & McNamara, D. S. (2011). Understanding expert ratings of essay quality: Coh-Metrix analyses of first and second language writing. *International*

*Journal of Continuing Engineering Education and Life Long Learning*, *21*(2-3), 170-191.

Crossley, S. A., & McNamara, D. S. (2012). Predicting second language writing proficiency: The roles of cohesion and linguistic sophistication. *Journal of Research in Reading*, *35*(2), 115-135.

Crossley, S. A., Salsbury, T., & McNamara, D. S. (2009). Measuring second language lexical growth using hypernymic relationships. *Language Learning*, **59**, 307-334.

Crossley, S. A., Salsbury, T., McNamara, D. S., & Jarvis, S. (2011). What is lexical proficiency? Some answers from computational models of speech data. *TESOL Quarterly*, *45*(1), 182-193.

Crossley, S. A., Weston, J., McLain Sullivan, S. T., & McNamara, D. S. (2011). The development of writing proficiency as a function of grade level: A linguistic analysis. *Written Communication*, *28*(3), 282-311.

Crowhurst, M. (1980). Syntactic complexity and teachers' quality ratings of narrations and arguments.Research in the Teaching of English, 14, 223–231

Crowhurst, M. (1983). Syntactic complexity and writing quality: A review. Canadian Journal of Education, 8, 1–16.

Daller, H., R. Van Hout, and J. Treffers-Daller. (2003). 'Lexical richness in the spontaneous speech of bilinguals,' Applied Linguistics 24: 197–222.

Deane, P. (2013). On the relation between automated essay scoring and modern views of the writing construct. *Assessing Writing*, *18*, 7-24.

deBoer, F. (2014). Evaluating the comparability of two measures of lexical diversity. *System*, *47*, 139-145.

Donovan, C. A., & Smolkin, L. B. (2006). Children's understanding of genre and writing development. In C. A. MacArthur, S. Graham, & J. Fitzgerald (Eds.), Handbook of writing research (pp. 131–143). New York: The Guilford Press

Educational Testing Service. (2005). TOEFL iBT Scores Educational Testing Service. (2000). *TOEFL 2000-2001, information bulletin for computer-based testing.* Princeton, NJ: Author.

Ellis, R. & Yuan, F. 2004. The effects of planning on fluency, complexity, and accuracy in second language narrative writing. *Studies in Second Language Acquisition* 26: 59–84.

Ellis, R., & Barkhuizen, G. (2005). *Analysing learner language*. Oxford: OUP.

Engber, C. A. (1995). The relationship of lexical proficiency to the quality of ESL compositions. *Journal o f Second Language Writing*, 4(2), 139-155.

Evans, J.D. (1996) Straightforward statistics for the behavioral sciences. Pacific Grove, CA: Brooks/Cole Publishing

Ferris, D. R. (1994). Lexical and syntactic features of ESL writing by students at different levels of L2 proficiency. *Tesol Quarterly*, *28*(2), 414-420.

Ferris, D.R. 1994. Rhetorical strategies in student persuasive writing: Differences between native and non-native English speakers. *Research in the Teaching of English* 28(1): 45–65

Foster, P., & Skehan, P. (1996). The influence of planning and task type on second language performance. *Studies in Second language acquisition*, *18*(3), 299-323.

Friginal, E., Li, M., & Weigle, S. C. (2014). Revisiting multiple profiles of learner compositions: A comparison of highly rated NS and NNS essays. *Journal of Second Language Writing*, *23*, 1-16.

Geiser, S., & Studley, R. (2001, October). *UC and the SAT: Predictive validity and differential impact of the SAT I and SAT II at the University of California*. Paper presented at the Meeting of the Board of Admissions and Relations with Schools of the University of California.

Graesser, A. C., McNamara, D. S., & Kulikowich, J. M. (2011). Coh-Metrix: Providing multilevel analyses of text characteristics. Educational Researcher, 40(5), 223–234.

Grobe, G. (1981). Syntactic maturity, mechanics, and vocabulary as predictors of quality ratings.Research in the Teaching of English, 15(1), 75–85

Guo, L., Crossley, S. A., & McNamara, D. S. (2013). Predicting human judgments of essay quality in both integrated and independent second language writing samples: A comparison study. *Assessing Writing*, *18*(3), 218-238.

Halliday, M. A. K., & Hassan, R. (1985). Language, context and text: Aspect of language in a social semiotic perspective. Geelong, Australia: Deakin University Press.

Harley, B., & King, M. L. (1989). Verb lexis in the written compositions of young L2 learners. *Studies in Second Language Acquisition, 11*, 415-439.

Hinkel, E. (2002). *Second language writers' text: Linguistic and rhetorical features*. Routledge.

Huot, B. (1990). Reliability, validity, and holistic scoring: What we know and what we need to know. *College Composition and Communication, 41*, 201-213.

IELTS. 2007. International English Language Testing System Handbook 2007.

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

Jarvis, S., Grant, L., Bikowski, D., & Ferris, D. (2003). Exploring multiple profiles of highly rated learner compositions. *Journal of Second Language Writing*, *12*(4), 377-403.

Jenkins, J., Johnson, E., & Hileman, J. (2004). When is reading also writing: Sources ofindividual differences on the new reading performance assessments. *Scientific Studies in Reading, 8*, 125-151

Johansson, C., & Geisler, C. (2011). Syntactic aspects of the writing of Swedish L2 learners of English. *Language and Computers-Studies in Practical Linguistics*, *73*(1), 139.

Kroll, B.,(1990) What does time buy? ESL student performance on home versus class compositions+ In B, Kroll Ed, Second language writing pp.140–154 New York: Cambridge University Press

Kyle, K., Crossley, S., & Verspoor, M. (2021). Measuring longitudinal writing development using indices of syntactic complexity and sophistication. *Studies in Second Language Acquisition*, *43*(4), 781-812.

Larsen-Freeman, D. 1978. An ESL index of development. *TESOL Quarterly* 12: 439–448.

Laufer, B. (2003). Vocabulary acquisition in a second language: Do learners really acquire most vocabulary by reading? Some empirical evidence. *Canadian Modern Language Review*, *59*(4), 567-587.

Laufer, B. and P. Nation. (1995). 'Vocabulary size and use – lexical richness in L2 written production,' Applied Linguistics 16: 307–22.

Leki, I., & Carson, J. G. (1994). Students' perceptions of EAP writing instruction and writing needs across the disciplines. *Tesol Quarterly*, *28*(1), 81-101.

Lennon, P. (1990). Investigating fluency in EFL: A quantitative approach. *Language learning*, *40*(3), 387-417.

Linnarud, M. (1986). *Lexis in composition: A performance analysis o f Swedish learners' written English.* Malmo, Sweden: Liber Forlag Malmo.

Lu, X. (2011). A corpus-based evaluation of syntactic complexity measures as indices of college-level ESL writers' language development. *Tesol Quarterly*, 36-62.

Lu, X. 2010. Automatic analysis of syntactic complexity in second language writing. *International Journal of Corpus Linguistics* 15(4): 474–496.

Maamuujav, U., Olson, C. B., & Chung, H. (2021). Syntactic and lexical features of adolescent L2 students' academic writing. *Journal of Second Language Writing*, *53*, 100822.

Malvern, D. D., Richards, B. J., Chipere, N., & Durán, P. (2004). Lexical diversity and language development. *Houndmills, Hampshire, UK: Palgrave Macmillan*.

Mazgutova, D., & Kormos, J. (2015). Syntactic and lexical development in an intensive English for Academic Purposes programme. *Journal of Second Language Writing*, *29*, 3-15.

McCarthy, P. M., & Jarvis, S. (2007). vocd: A theoretical and empirical evaluation. *Language Testing*, *24*(4), 459-488.

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

McNamara, D. S., Crossley, S. A., & McCarthy, P. M. (2010). Linguistic features of writing quality. *Written Communication*, *27*, 57-86.

McNamara, D. S., Graesser, A. C., McCarthy, P. M., & Cai, Z. (2014). *Automated evaluation of text and discourse with Coh-Metrix*. Cambridge University Press.

Meara, P. 2005. 'Lexical frequency profiles: A Monte Carlo analysis,' Applied Linguistics 26:32–47. *MELAB technical manual.* (1996). Ann Arbor, MI: University of Michigan Press.

Norris, J. M., & Ortega, L. (2009). Towards an organic approach to investigating CAF in instructed SLA: the case of complexity. Applied Linguistics, 30, 555–578.

Northwest Regional Educational Laboratory (2004). 6+1 trait writing – about. Retrieved December 5, 2016, from http://www.nwet.org/assessment/about.php?odelay=1&d=1.

Olinghouse, N. G. and J. Wilson. (2013). 'The relationship between vocabulary and writing quality in three genres,' Reading and Writing 26: 45–65

Olinghouse, N. G., & Graham, S. (2009). The relationship between the discourse knowledge and the writing performance of elementary-grade students. Journal of Educational Psychology, 101(1), 37–50.

Ong, J. (2013a). Discovery of ideas in second language writing task environment. System, 41, 529–542

Ong, J., & Zhang, L. J. (2010). Effects of task complexity on fluency and lexical complexity in EFL students' argumentative writing. Journal of Second Language Writing, 19, 218–233.

Ong, J., & Zhang, L. J. (2013). Effects of manipulation of cognitive processes on English-as- a-foreign-language (EFL) writers' text quality. TESOL Quarterly, 47, 375–398.

Ortega, L. (2003). Syntactic complexity measures and their relationship to L2 proficiency: A research synthesis of college-level L2 writing. Applied Linguistics, 24, 492–518

Ortega, L. (2012). Interlanguage complexity: A construct in search of theoretical renewal. In B. Kortmann & B. Szmrecsanyi (Eds.), Linguistic complexity: Second language acquisition, indigenization, contact (pp. 127–155). Berlin: De Gruyter.

Ortega, L. (2015). Syntactic complexity in L2 writing: Progress and expansion. *Journal of Second Language Writing*, *29*, 82-94.

Ortega, L. 2000. Understanding syntactic complexity: The measurement of change in the syntax of instructed L2 Spanish learners. PhD dissertation, University of Hawaii.

Pallotti, G. (2015). A simple view of linguistic complexity. *Second Language Research*, *31*(1), 117-134.

Park, J.H. (2017) Syntactic complexity as a predictor of second language writing proficiency and writing quality (Doctoral Dissertation) Available from ProQuest Dissertations & Theses Global database. (UMI No. 10248579)

Ravid, D. (2005). Emergence of linguistic complexity in later language development: Evidence from expository text construction. In *Perspectives on language and language development* (pp. 337-355). Springer, Boston, MA.

Santos, T. (1988). Professors' Reactions to the Academic Writing of Nonnative-Speaking Students. *Tesol Quarterly*, *22*(1), 69-90.

Seidman, I. (2012). *Interviewing as qualitative research: A guide for researchers in education and the social sciences*. Teachers college press.

Skehan, P. (2003). Task-based instruction. *Language Teaching, 36*, 1–14.

Stockwell, G. (2005). Syntactical and lexical development in NNS-NNS asynchronous CMC. *The JALT CALL Journal*, *1*(3), 33-49.

Stockwell, G. & Harrington, M. (2003). The incidental development of L2 proficiency in NS-NNS email interactions. *CALICO Journal* 20(2): 337–359.

Treffers-Daller, J., Parslow, P., & Williams, S. (2016). Back to basics: how measures of lexical diversity can help discriminate between CEFR levels. *Applied Linguistics*, amw009.

University of Michigan. 2003. Michigan English Language Assessment Battery Technical Manual 2003.

Vyatkina, N., Hirschmann, H., & Golcher, F. (2015). Syntactic modification at early stages of L2 German writing development: A longitudinal learner corpus study. *Journal of Second Language Writing, 29*, 28-50.

Way, D. P., Joiner, E. G., & Seaman, M. A. (2000). Writing in the secondary foreign language classroom: The effects of prompts and tasks on novice learners of French. *The Modern Language Journal, 84*(2), 171-184.

Weber, R. P. (1990). *Basic content analysis* (2nd ed.). Thousand Oaks, CA: Sage Publications

Weigle, S. C. (2002). *Assessing writing*. Ernst Klett Sprachen.

White, E. M. (1984). Holisticism. *College Composition and Communication 35*, 400-409.

Wolfe-Quintero, K., Inagaki, S., & Kim, H.-Y. (1998). Second language development in writing: Measures of fluency, accuracy, and complexity. Honolulu, HI: University of Hawaii Press.

Writing Study Group of the NCTE Executive Committee (2004). NCTE beliefs about the teaching of writing. Urbana, IL: National Council of Teachers of English. Retrieved on December 5, 2016, from htp://www.ncte.org/about/over/positions/category/write/118876.htm.

Yu, G. (2009). Lexical diversity in writing and speaking task performances.*Applied Linguistics, 31*(2), 236-259.

# APPENDICES

APPENDIX – 1. Test of English, IELTS and MELAB Scoring Rubrics

## TEST OF WRITTEN ENGLISH (TWE) SCORING GUIDE

Revised 2/90

Readers will assign scores based on the following scoring guide. Though examinees are asked to write on a specific topic, parts of the topic may be treated by implication. Readers should focus on what the examinee does well.

**Scores**

**6** **Demonstrates clear competence in writing on both the rhetorical and syntactic levels, though it may have occasional errors.**

A paper in this category
—effectively addresses the writing task
—is well organized and well developed
—uses clearly appropriate details to support a thesis or illustrate ideas
—displays consistent facility in the use of language
—demonstrates syntactic variety and appropriate word choice

**5** **Demonstrates competence in writing on both the rhetorical and syntactic levels, though it will probably have occasional errors.**

A paper in this category
—may address some parts of the task more effectively than others
—is generally well organized and developed
—uses details to support a thesis or illustrate an idea
—displays facility in the use of language
—demonstrates some syntactic variety and range of vocabulary

**4** **Demonstrates minimal competence in writing on both the rhetorical and syntactic levels.**

A paper in this category
—addresses the writing topic adequately but may slight parts of the task
—is adequately organized and developed
—uses some details to support a thesis or illustrate an idea
—demonstrates adequate but possibly inconsistent facility with syntax and usage
—may contain some errors that occasionally obscure meaning

**3** **Demonstrates some developing competence in writing, but it remains flawed on either the rhetorical or syntactic level, or both.**

A paper in this category may reveal one or more of the following weaknesses:
—inadequate organization or development
—inappropriate or insufficient details to support or illustrate generalizations
—a noticeably inappropriate choice of words or word forms
—an accumulation of errors in sentence structure and/or usage

**2** **Suggests incompetence in writing.**

A paper in this category is seriously flawed by one or more of the following weaknesses:
—serious disorganization or underdevelopment
—little or no detail, or irrelevant specifics
—serious and frequent errors in sentence structure or usage

BRITISH COUNCIL idp IELTS AUSTRALIA UNIVERSITY of CAMBRIDGE ESOL Examinations

## IELTS TASK 1 Writing band descriptors (public version)

| Band | Task Achievement | Coherence and Cohesion | Lexical Resource | Grammatical Range and Accuracy |
|---|---|---|---|---|
| 9 | ▪ fully satisfies all the requirements of the task<br>▪ clearly presents a fully developed response | ▪ uses cohesion in such a way that it attracts no attention<br>▪ skilfully manages paragraphing | ▪ uses a wide range of vocabulary with very natural and sophisticated control of lexical features; rare minor errors occur only as 'slips' | ▪ uses a wide range of structures with full flexibility and accuracy; rare minor errors occur only as 'slips' |
| 8 | ▪ covers all requirements of the task sufficiently<br>▪ presents, highlights and illustrates key features/bullet points clearly and appropriately | ▪ sequences information and ideas logically<br>▪ manages all aspects of cohesion well<br>▪ uses paragraphing sufficiently and appropriately | ▪ uses a wide range of vocabulary fluently and flexibly to convey precise meanings<br>▪ skilfully uses uncommon lexical items but there may be occasional inaccuracies in word choice and collocation<br>▪ produces rare errors in spelling and/or word formation | ▪ uses a wide range of structures<br>▪ the majority of sentences are error-free<br>▪ makes only very occasional errors or inappropriacies |
| 7 | ▪ covers the requirements of the task<br>▪ (Academic) presents a clear overview of main trends, differences or stages<br>▪ (General Training) presents a clear purpose, with the tone consistent and appropriate<br>▪ clearly presents and highlights key features/bullet points but could be more fully extended | ▪ logically organises information and ideas; there is clear progression throughout<br>▪ uses a range of cohesive devices appropriately although there may be some under-/over-use | ▪ uses a sufficient range of vocabulary to allow some flexibility and precision<br>▪ uses less common lexical items with some awareness of style and collocation<br>▪ may produce occasional errors in word choice, spelling and/or word formation | ▪ uses a variety of complex structures<br>▪ produces frequent error-free sentences<br>▪ has good control of grammar and punctuation but may make a few errors |
| 6 | ▪ addresses the requirements of the task<br>▪ (Academic) presents an overview with information appropriately selected<br>▪ (General Training) presents a purpose that is generally clear; there may be inconsistencies in tone<br>▪ presents and adequately highlights key features/bullet points but details may be irrelevant, inappropriate or inaccurate | ▪ arranges information and ideas coherently and there is a clear overall progression<br>▪ uses cohesive devices effectively, but cohesion within and/or between sentences may be faulty or mechanical<br>▪ may not always use referencing clearly or appropriately | ▪ uses an adequate range of vocabulary for the task<br>▪ attempts to use less common vocabulary but with some inaccuracy<br>▪ makes some errors in spelling and/or word formation, but they do not impede communication | ▪ uses a mix of simple and complex sentence forms<br>▪ makes some errors in grammar and punctuation but they rarely reduce communication |

| | | | | |
|---|---|---|---|---|
| 5 | • generally addresses the task; the format may be inappropriate in places<br>• (Academic) recounts detail mechanically with no clear overview; there may be no data to support the description<br>• (General Training) may present a purpose for the letter that is unclear at times; the tone may be variable and sometimes inappropriate<br>• presents, but inadequately covers, key features/bullet points; there may be a tendency to focus on detail | • presents information with some organisation but there may be a lack of overall progression<br>• makes inadequate, inaccurate or over-use of cohesive devices<br>• may be repetitive because of lack of referencing and substitution | • uses a limited range of vocabulary, but this is minimally adequate for the task<br>• may make noticeable errors in spelling and/or word formation that may cause some difficulty for the reader | • uses only a limited range of structures<br>• attempts complex sentences but these tend to be less accurate than simple sentences<br>• may make frequent grammatical errors and punctuation may be faulty; errors can cause some difficulty for the reader |
| 4 | • attempts to address the task but does not cover all key features/bullet points; the format may be inappropriate<br>• (General Training) fails to clearly explain the purpose of the letter; the tone may be inappropriate<br>• may confuse key features/bullet points with detail; parts may be unclear, irrelevant, repetitive or inaccurate | • presents information and ideas but these are not arranged coherently and there is no clear progression in the response<br>• uses some basic cohesive devices but these may be inaccurate or repetitive | • uses only basic vocabulary which may be used repetitively or which may be inappropriate for the task<br>• has limited control of word formation and/or spelling;<br>• errors may cause strain for the reader | • uses only a very limited range of structures with only rare use of subordinate clauses<br>• some structures are accurate but errors predominate, and punctuation is often faulty |
| 3 | • fails to address the task, which may have been completely misunderstood<br>• presents limited ideas which may be largely irrelevant/repetitive | • does not organise ideas logically<br>• may use a very limited range of cohesive devices, and those used may not indicate a logical relationship between ideas | • uses only a very limited range of words and expressions with very limited control of word formation and/or spelling<br>• errors may severely distort the message | • attempts sentence forms but errors in grammar and punctuation predominate and distort the meaning |
| 2 | • answer is barely related to the task | • has very little control of organisational features | • uses an extremely limited range of vocabulary; essentially no control of word formation and/or spelling | • cannot use sentence forms except in memorised phrases |
| 1 | • answer is completely unrelated to the task | • fails to communicate any message | • can only use a few isolated words | • cannot use sentence forms at all |
| 0 | • does not attend<br>• does not attempt the task in any way<br>• writes a totally memorised response | | | |

## MELAB Writing Scale

**97**

Topic is richly and fully developed. Flexible use of a wide range of syntactic (sentence level) structures, accurate morphological (word forms) control. Organization is appropriate and effective, and there is excellent control of connection. There is a wide range of appropriately used vocabulary. Spelling and punctuation appear error free.

**93**

Topic is fully and complexly developed. A wide range of syntactic structures are used with flexibility. Morphological control is nearly always accurate. Organization is well controlled and appropriate to the material, and the writing is well connected. Vocabulary is broad and appropriately used. Spelling and punctuation errors are not distracting.

87
Topic is well developed, with acknowledgement of its complexity. Varied syntactic structures are used with some flexibility, and there is good morphological control. Organization is controlled and generally appropriate to the material, and there are few problems with connection. Vocabulary is broad and usually used appropriately. Spelling and punctuation errors are not distracting.

83
A composition in which the topic is generally clearly and completely developed, with at least some acknowledgement of its complexity. Both simple and complex syntactic structures are generally adequately used; there is adequate morphological control. Organization is controlled and shows some appropriacy to the material, and connection is usually adequate. Vocabulary use shows some flexibility, and is usually appropriate. Spelling and punctuation errors are sometimes distracting.

77
Topic is developed clearly but not completely and without acknowledging its complexity. Both simple and complex syntactic structures are present; in some

73
Topic development is present, although limited by incompleteness, lack of clarity, or lack of focus. The topic may be treated as though it has only one dimension, or only one point of view is possible. In some "73" essays both simple and complex syntactic structures are present, but with many errors; others have accurate syntax, but are very restricted in the range of language attempted. Morphological control is inconsistent. Organization is partially controlled, while connection is often absent or unsuccessful. Vocabulary is sometimes inadequate, and sometimes inappropriately used. Spelling and punctuation errors are sometimes distracting.

67
Topic development is present but restricted, and often incomplete or unclear. Simple syntactic structures dominate, with many errors; complex syntactic structures, if present, are not controlled. Lacks morphological control. Organization, when apparent, is poorly controlled, and little or no connection is apparent. Narrow and simple vocabulary usually approximates meaning but is often inappropriately used. Spelling and punctuation errors are often distracting.

**63**

Contains little sign of topic development. Simple syntactic structures are present, but with many errors; lacks morphological control. There is little or no organization, and no connection apparent. Narrow and simple vocabulary inhibits communication, and spelling and punctuation errors often cause serious interference.

**57**

Often extremely short; contains only fragmentary communication about the topic. There is little syntactic or morphological control, and no organization or connection is apparent. Vocabulary is highly restricted and inaccurately used. Spelling is often indecipherable and punctuation is missing or appears random.

**53**

Extremely short, usually about 40 words or less; communicates nothing, and is often copied directly from the prompt. There is little sign of syntactic or morphological control, and no apparent organization or connection. Vocabulary is extremely restricted and repetitively used. Spelling is often indecipherable, and punctuation is missing or appears random.

APPENDIX-2: The sample of Informed Consent Form (fort he undergraduate students)

**Araştırma Gönüllü Katılım Formu**

Bu çalışma bir doktora araştırması olup Arş. Gör. Zafer SUSOY tarafından yürütülmektedir. Bu çalışmaya katılımınız gönüllülük esasına dayanmaktadır. Çalışmanın amacı doğrultusunda, yazdığınız kompozisyonlar aracılığıyla veriler toplanacaktır. Araştırma kapsamında toplanan veriler, sadece bilimsel amaçlar doğrultusunda kullanılacak, araştırmanın amacı dışında ya da bir başka araştırmada kullanılmayacak ve gerekmesi halinde, sizin (yazılı) izniniz olmadan başkalarıyla paylaşılmayacaktır.

**Bu çalışmaya tamamen kendi rızamla, istediğim takdirde çalışmadan ayrılabileceğimi bilerek verdiğim bilgilerin bilimsel amaçlarla kullanılmasını kabul ediyorum.**

| Öğrenci No | Öğrenci Adı Soyadı | İmza |
|---|---|---|
| 1077***5442 | TUĞBA NUR KARADAĞ | |
| 1255***7734 | PINAR KİZMAZ | |
| 1295***1406 | EZGİ YORULMAZ | |
| 1546***7970 | EMİNE DAĞ | |
| 1620***4760 | YUNUS EMRE YEŞİLIRMAK | |
| 1644***2628 | DÖNE ERYILMAZ | |
| 1767***9366 | YAKUP ÇANKAYA | |
| 2154***7332 | ONUR YÜCEL | |
| 2392***7614 | BÜŞRA TUNA | |
| 2521***6256 | FATMA NUR ÖZKAN | |
| 2915***0282 | REYHAN AKÇAY | |
| 2982***7264 | SEDA AYDIN | |
| 3078***5284 | MERVE ENİŞ | |
| 3094***1044 | SİMGE ATAL | |
| 3527***5036 | EBRU YILMAZ | |
| 3552***7206 | FATMA IRMAK | |
| 3566***3374 | TALİP KARAHAN | |
| 3604***6210 | ESRA TOPCU | |
| 3902***3462 | GİZEM DOĞANER | |
| 4138***7544 | AYŞENUR ÖZTÜRK | |
| 5056***1672 | MERVE HİLAL ALLITEKİN | |

| | | |
|---|---|---|
| 5108***5090 | EGEM İŞGÖRÜR | |
| 6174***8276 | MÜCAHİT ÖTER | |
| 9925***9178 | TOYLY ERGASHEV | |

APPENDIX-3. The Research Ethics Approval

APPENDIX -4. Writing Topic Selection Questionnaire

**Sayın Hocam,**

Bölümümüzde yürüyen doktora tezi çalışmam kapsamında, bölümümüz birinci ve dördüncü sınıf öğrencilerinden aynı konu hakkında görüşlerini bildirecekleri İngilizce bir kompozisyon yazmaları istenecektir. Yazacakları kompoziyonun konusunu belirlemede alanı ve öğrencileri iyi tanıyan siz uzmanların görüş ve desteğine ihtiyaç duyuyoruz.

Diğer sayfada 10 adet '*Görüş Kompozisyonu*' (Opinion Essay) konuları bulunmaktadır. Lütfen, bu konular arasında öğrencilerimizin rahatça yazabileceklerini düşündüğünüz 3 konuyu aşağıdaki öncelik sırasına yerleştiriniz. Bu uygunluk eşlemesinde konulara verilen harfleri kullanınız. Listelenen konular dışında başka bir öneriniz olursa, lütfen '*Diğer*' alanında belirtiniz.

**(En öncelikli) 1…..**

**(İkinci Alternatif) 2…..**

**(Üçüncü Alternatif ) 3…..**

**Diğer……**
**……**
**……**
**……**
**……**
**……**
**……**

**Sağladığınız kıymetli destekten ötürü minnettarız.**

**Arş. Gör. Zafer SUSOY**

**Opinion Essay Topics**

A. Do you agree or disagree with the following statement? With the help of technology, students nowadays can learn more information and learn it more quickly. Use specific reasons and examples to support your answer.
B. Some people think that they can learn better by themselves than with a teacher. Others think that it is always better to have a teacher. Which do you prefer? Use specific reasons to develop your essay.

C. Do you agree or disagree with the following statement? Teachers should be paid according to how much their students learn. Give specific reasons and examples to support your opinion.
D. Some people believe that university students should be required to attend classes. Others believe that going to classes should be optional for students. Which point of view do you agree with? Use specific reasons and details to explain your answer.
E. Students at universities often have a choice of places to live. They may choose to live in university dormitories, or they may choose to live in apartments in the community. Compare the advantages of living in university housing with the advantages of living in an apartment in the community. Where would you prefer to live? Give reasons for your preference.
F. In the future, students may have the choice of studying at home by using technology such as computers or television or of studying at traditional schools. Which would you prefer? Use reasons and specific details to explain your choice.

G. Discipline is an ever increasing problem in modern schools. Some people think that discipline should be the responsibility of teachers, while others think that this is the role of parents. Discuss both sides and give your opinion
H. Some people think that teachers should be able to ask disruptive children to leave the classroom. Do you think it is the best way to deal with a disruptive child in the classroom?
İ. Some schools are very strict about their school uniform and the appearance of their pupils while other schools have a very relaxed dress code. What are the advantages and disadvantages of children of having a school uniform?
J. It is thought by some that a school teacher’s role is to motivate and inspire students. However, other people believe that a teacher’s primary role is to pass on knowledge. What do you think is the role of a teacher?

APPENDIX -5. Essay Writing Template

**Name/Surname:**

**Write an essay on the following topic within an hour:** Students at universities often have a choice of places to live. They may choose to live in university dormitories, in private student residences, or they may choose to live in apartments in the town. Where would you prefer to live? Why? Give reasons for your preference

APPENDIX -6. A sample of Informed Consent Form (for instructors)

**Mülakata Katılım İçin Rıza Formu**

Bu çalışma Anadolu Üniversitesi Eğitim Bilimleri Enstitüsü'nde sürmekte olan bir doktora araştırması olup, Arş. Gör. Zafer SUSOY tarafından yürütülmektedir. Bu çalışmaya katılımınız gönüllülük esasına dayanmaktadır. Çalışmanın amacı doğrultusunda, sözlü yöneltilecek görüşme soruları aracılığıyla veriler toplanacaktır. Araştırma kapsamında toplanan veriler, sadece bilimsel amaçlar doğrultusunda kullanılacak, araştırmanın amacı dışında ya da bir başka araştırmada kullanılmayacak ve gerekmesi halinde, sizin (yazılı) izniniz olmadan başkalarıyla paylaşılmayacaktır.

***Bu çalışmaya tamamen kendi rızamla ve ses kaydı yapıldığını bilerek, verdiğim bilgilerin bilimsel amaçlarla kullanılmasını kabul ediyorum.***

**Katılımcı Ad/Soyadı:**

**Tarih:**

**İmza:**

APPENDIX -7. Construction Process of Interview Questions

Semi-structured interview questions **(First Version)**

1. Ne türlü öğrenci yazmalarıyla karşılaşıp, notlandırıyorsunuz?
    a. Ne kadar zamandır bu türlü yazıları notlandırıyorsunuz?
2. Bir yazıyı iyi olarak değerlendirirken dikkat ettiğiniz noktalar nelerdir?
3. İngilizce öğrenci yazmalarını notlandırıken kullandığınız bir yönerge var mı?
    a. Varsa eğer böyle bir yönerge, ona ne kadar sadık kalıyorsunuz?
4. Yazılarda 'yapısal karmaşıklık' (writing complexity) ifadesi size neleri çağrıştırıyor?
5. Yazılarda 'sözcük çeşitliliği' (lexical diversity) ifadesi size neleri çağrıştırıyor?
6. İyi bir yazı olarak kabul ettiğiniz yazılarda 'yapısal karmaşıklık' ve 'sözcüksel çeşitlilik' kendine ne kadar yer bulur?
7. Öğrencilerin sınıflarını ve/veya dil yeterliliklerini bilmek karmaşıklık ve çeşitlilik anlamındaki beklentilerinizi ve notunuzu etkiler mi? Nasıl?
8. Yazılarda 'yapısal karmaşıklık' ve 'sözcüksel çeşitlilik' notlandırmanızı etkiler mi?
9. Yazılarda 'yapısal karmaşıklık' ve 'sözcüksel çeşitlilik' sizce nasıl sağlanır?
    a. Yazılardaki 'yapısal karmaşıklık ve 'sözcüksel çeşitliliği' ne şekilde anlıyor, hangi aygıtlar aracılığıyla tespit ediyorsunuz?

Semi-structured interview questions **(End Version)**

1. Ne kadar zamandır İngilizce öğretmenliği adaylarının yazılarını notlandırıyorsunuz?
2. İngilizce öğretmen adaylarının yazılarını notlandırmak için nasıl bir süreç izliyorsunuz? Bu süreci basamaklar halinde anlatabilir misiniz?
3. İngilizce öğretmen adaylarının yazılarını notlandırırken nasıl bir yönerge takip ediyorunuz?
    a. Yönergeye ne kadar sadık kalıyorsunuz? Neden, açıklayınız?
4. Yazılarda 'yapısal karmaşıklık' (syntactic complexity) ifadesi size neleri çağrıştırıyor?
    a. Size syntactic complexity i işaret eden yapılara örnek verebilir misiniz?
    b. Öğrenci yazılarındaki 'yapısal karmaşıklık' seviyesi notunuzu nasıl etkiler?
5. Yazılarda 'sözcük çeşitliliği' (lexical diversity) ifadesi size neleri çağrıştırıyor?

a. a.Size lexical diversity i işaret eden yapılara örnek verebilir misiniz?
b. Öğrenci yazılarındaki ‘sözcük çeşitliliği’ seviyesi notunuzu nasıl etkiler?

6. İngilizce öğretmen adayları yazılarında beklentilerinizi ve notunuzu neler etkiliyor?
7. Öğrencilerin sınıflarını ve/veya dil yeterliliklerini bilmek yapısal karmaşıklık anlamındaki beklentilerinizi ve notunuzu etkiler mi?
8. Öğrencilerin sınıflarını ve/veya dil yeterliliklerini bilmek sözcük çeşitliliği anlamındaki beklentilerinizi ve notunuzu etkiler mi? Nasıl?
9. Eklemek istedikleriniz var mı, öğretmen adaylarının yazılarını notlandırırken karşılaştığınız güçlükler veya hissettikleriniz anlamında?

APPENDIX -8. Theme-Code Refinement Process in NVivo

**The first Version of Theme-Code Organization**

5.12.2019 13:44

Coding Summary By Code

İLk deneme

5.12.2019 13:44

| Aggregate | Classification | Coverage | Number Of Coding References | Reference Number | Coded By | Modified On |
|---|---|---|---|---|---|---|
| **Node** | | | | | | |
| **Nodes\\An Analogy Between Complexity and Diversity** | | | | | | |
| **Document** | | | | | | |
| Files\\ *** İle Görüşme R3 | | | | | | |
| No | | 0,0091 | 1 | | | |
| | | | | 1 | ZS | 5.12.2019 12:52 |

Yani hani demiştik ya dil yapılarını, farklı yapıları da kullanmalarını istiyoruz. Tek bir yapı üzerinden gitmesinler, sadece basit cümle kullanmasınlar vs. Sözcük çeşitliliğinde de aynı şekilde.

| | | | | | | |
|---|---|---|---|---|---|---|
| **Nodes\\Embodying LD as a Construct** | | | | | | |
| **Document** | | | | | | |
| Files\\ *** İle Görüşme R3 | | | | | | |
| No | | 0,0124 | 1 | | | |
| | | | | 1 | ZS | 5.12.2019 12:51 |

Yani bir konu hakkında yazıyorsanız o konu hakkında ya da o alana ait ne kadar geniş bir jargona sahipsiniz bunu göstermenizi bekliyoruz diyoruz öğrencilere. O yüzden de hem farklı kelime türlerini hem de mümkün olduğunca farklı kelimeleri kullanmalarını istiyoruz.

| | | | | | | |
|---|---|---|---|---|---|---|
| **Files\\ *** ile Görüşme R1** | | | | | | |
| No | | 0,0109 | 1 | | | |
| | | | | 1 | ZS | 5.12.2019 12:49 |

Ben hani lexicology dersi de verdiğim için özellikle kelime haznesinin genişliğinden bahsederken söylediğimiz şeyler var. Vocabulary size çok sığ bir şey. Bunun içerisinde depth ve breadth dediğimiz şeyler var. Depth dediğim şeylerde de yani aynı kelimenin birkaç anlamı olabilir, hangi contextte neyi kullanacağız,



| Aggregate | Classification | Coverage | Number Of Coding References | Reference Number | Coded By Initials | Modified On |
|---|---|---|---|---|---|---|
| **Files*** ile Görüşme R6** | | | | | | |
| No | | 0,0174 | 1 | | | |
| | | | | 1 | ZS | 5.12.2019 12:50 |

Yine simple olmayan, çeşitli kelimeler kullandığı, çeşitliliği olan, daha üst düzey, advanced düzey dediğimiz ya da burada upper-intermediate düzey olabilir. Kelimeler kullanabileceği bir şey anlıyorum yani.

| | | | | | | |
|---|---|---|---|---|---|---|
| **Files*** Transkript R7** | | | | | | |
| No | | 0,0073 | 1 | | | |
| | | | | 1 | ZS | 5.12.2019 12:55 |

sözcüklerin zorluk derecesi aklıma geliyor. İşte commonly used words dışında bir kelime kullanıyor mu, doğru kullanıyor mu?

| | | | | | | |
|---|---|---|---|---|---|---|
| **Files*** ile Görüşme R5** | | | | | | |
| No | | 0,0237 | 1 | | | |
| | | | | 1 | ZS | 5.12.2019 12:54 |

bir kere konuya uygun ve o kelimenin anlamını doğru bir yer de kullanabilmiş mi o, kelime yapısını doğru olarak kullanmış mı o, mesela adjective yerine adverb mü koymuş, noun yerine adjective mi getirmiş, tabi bunlar da çok önemli

| | | | | | | |
|---|---|---|---|---|---|---|
| **Nodes\\Embodying LD as a Construct\Examples of Lexical Diversity** | | | | | | |
| **Document** | | | | | | |
| **Files*** İle Görüşme R8** | | | | | | |
| No | | 0,0050 | 1 | | | |

| | | | |
|---|---|---|---|
| 1 | ZS | 5.12.2019 13:11 | |

You make me frustrated” yerine “You frustrate me” gibi. Yani orada farklı… Kelimenin noun halini, verb halini, adjective haline hakim olup

**Files*** İle Görüşme R3**

| | | |
|---|---|---|
| No | 0,0037 | 1 |

| | | | |
|---|---|---|---|
| 1 | ZS | 5.12.2019 13:05 | |

Duvarı tasvir edecek. İşte light blue falan demiyor. Bluenun shadeini veriyor.



| Aggregate | Classification | Coverage | Number Of Coding References | Reference Number | Coded By Initials | Modified On |
|---|---|---|---|---|---|---|
| **Files*** ile Görüşme R1** | | | | | | |
| | No | 0,0128 | 1 | | | |
| | | | | 1 | ZS | 5.12.2019 13:02 |

synonymler, antonymler, collocationlar var, idiomatic expressionlar ve chunklar var akademik anlamda. Bunlar önemli. Mesela biz cause & effect analysis essay sorduk. Orda işte causeları yazacak ya da effectleri yazacak. Hep causeda “first cause, first reason” mı desin? Farklı “impact”di “influence”dı falan gibi cümleleri de… Pardon, kelimeleri de kullanmasını bekleriz.

**Files*** Kür Transkript R2**

| | | |
|---|---|---|
| No | 0,0267 | 2 |

| | | |
|---|---|---|
| 1 | ZS | 5.12.2019 13:10 |

Yine diyebileceğim gibi ileri seviyede belki özellikle noun formların daha yoğun olduğu. Çoğu fiilin noun hali daha advanced diye falan kabul ediliyor. İşte dolayısıyla daha az karşımıza çıkan kelimeler olabilir.

| | | |
|---|---|---|
| 2 | ZS | 5.12.2019 13:01 |

O konunun çerçevesi içerisinde advanced kelimeleri de kullanabilmesiyle ilgili. Mesela sürekli “thing” falan yazılıyorlar. Mesela sürekli “thing” demektense bir seferinde “reason” desin. “Compose” desin vs. vs. Dolayısıyla bu da bir çeşitlilik. Eş anlamları da kullanmak. Noun formları düzgün kullanabilmek bir parçası.

**Files*** ile Görüşme R6**

No 0,0036 1

1 ZS 5.12.2019 13:07

Mesela important demez de significant der.

**Files*** Görüşme R5**

No 0,0213 1

1 ZS 5.12.2019 13:08

ımmm, adverb kullanımları özellikle çok önemli, doğru yerlerde kullanılması ve mesela hani reading derslerinde karşılaştıkları yapılar olabilir chunklar halinde bazı yapılar olabilir, phrasal verbler olabilir

**Files*** İle Görüşme R4**

No 0,0082 1

1 ZS 5.12.2019 13:06

Hiç üşenmeden bakıverseler red yerine farklı bir mesela şey kullanabilirler. Ne bileyim "crimson" derler. Yani ama hepsini red, red, red flowers yerine "crimson" der. İşte ne bileyim "It was blodly red" der



| Aggregate | Classification | Coverage | Number Of Coding References | Reference Number | Coded By Initials | Modified On |
|---|---|---|---|---|---|---|
| **Nodes\\Embodying SC as a Construct** | | | | | | |
| **Document** | | | | | | |
| **Files*** İle Görüşme R3** | | | | | | |
| No | | 0,0195 | 2 | | | |
| | | | | 1 | ZS | 5.12.2019 12:39 |

Kendi aslında yazım stilimi anlatıyor. Bana hep şey eleştirisi gelir mesela. Sen kısa cümle kuramıyor musun? Yani clauselar havada uçuşur… Böyle bir cümle başlar bitene

kadar böyle bir süner. Ama aslında gramer olarak doğrudur, anlamsal olarak da doğrudur.

2 ZS 5.12.2019 12:40

Hani syntactic complexity olan da sanki anlamı vermek için biraz daha dolandıran, daha fazla grammatical işte yapı kullanan… Bir yazım stili canlandı aklımda.

### Files*** ile Görüşme R1

| | | |
|---|---|---|
| No | 0,0346 | 1 |

1 ZS 5.12.2019 11:02

İşte subject-verb agreementına dikkat ediyor, işte belli başlı bildiği tensler ya da bildiği conjunctionları kullanıyor. Dolayısıyla embedded structurelar, clauselar falan hiçbir şey kullanmamaya çalışıyor. Ama bu da tabi ki sophistication anlamında öğrencinin son derece yalın bir şey yazmasına sebep oluyor. Dolayısıyla hem kelime anlamında variety yok, hem grammar anlamında variety yok, işte echoic termleri kullanmıyor; synonym, antonymleri kullanmıyor; and, but, so onlardan başka bağlaç bir sürü öğretilmesine rağmen kullanmıyor. Dolayısıyla hem paragrafta cümleleri birbirine bağlarken conjunctionları eksik kullanıyor hem de metin içerisinde paragrafları birbirine bağlarken de… Yani belli başlı cümle ya da phraseleri kullanması lazım ki bir önceki ya da sonrakine atıfta bulunsun. Bunları ne yapıyor? Gözden kaçırmış oluyor. Dolayısıyla sadece bir topic sentence yazarım, bir iki cümle yazarım bir de conclusion sentence yaparım olayı bitiririm diyor. Ama böyle bir şey olmaması gerekiyor.

### Files*** Kür Transkript R2

| | | |
|---|---|---|
| No | 0,0191 | 2 |

1 ZS 5.12.2019 12:35

Karmaşıklık deyince Türkçede daha kötü bir şeyleri çağrıştıyor ama İngilizce de syntactic complexity bildiğim için yani tahmin edebildiğim için daha olumlu bir şeyi çağrıştırıyor. Using variety of structures gibi düşünüyorum. Daha çok complex ve compound structures da kullanabilmeleri gerektiğini düşünüyorum

2 ZS 5.12.2019 12:37

syntactic complexityden bir yazarın dili kullanma becerisini anlıyorum.

| Aggregate | Classification | Coverage | Number Of Coding References | Reference Number | Coded By Initials | Modified On |
|---|---|---|---|---|---|---|
| **Files*** Görüşme R5** | | | | | | |
| No | | 0,0416 | 1 | | | |
| | | | | 1 | ZS | 5.12.2019 12:38 |

Bu ifaden neler çağrıştırıyor? Bu ifadede bağlaç kullanımları yani transitionlardan bahsediyorum conjunctionslardan bahsediyorum, clauseların kullanımlarından bahsediyorum yani adverbial clause, adjective clause,noun clause kullanımlarından bahsediyorum. Bunların hepsinin beraber doğru bir şekilde kullanılıp kullanılması benim için complex cümle bu yoksa subject verb object o basit bir cümle yapısıdır.

**Nodes\\Embodying SC as a Construct\Examples of Syntactic Complexity**

**Document**

**Files*** İle Görüşme R8**

| | | | | | | |
|---|---|---|---|---|---|---|
| No | | 0,0190 | 1 | | | |
| | | | | 1 | ZS | 5.12.2019 12:46 |

Yani tabi ki farklı clauseları kullanabilmesi gerekiyor bir şekilde. Sadece relative clause falan değil ama adverb olsun, işte ondan sonra farklı transitionlar olsun… Ondan sonra farklı… Yani bir şekilde pasive kullanabilsin. Arada bir kelimeleri değiştirsin. Kelimeleri değiştirdiği için cümlenin yapısı değişsin. Onun üzerine uğraşsın. Orada prepositionları bilsin. O prepositionların arkasından bambaşka, saçma bir şey geldiğinde de kafası karışmasın istiyorum. Yani -ing gelsin of'tan sonra ama onun kafasını karıştırmasın.

**Files*** ile Görüşme R1**

| | | | | | | |
|---|---|---|---|---|---|---|
| No | | 0,0085 | 1 | | | |
| | | | | 1 | ZS | 5.12.2019 12:44 |

Yani inversionı kullansın, işte embedded structurelar kullansın, iki, üç cümleyi değişik değişik conjunctionlar kullanarak bir arada kullanmayı becersin… Değişik işte phrasal verbler, idiomatic expressionlar tabi akademik anlamda uygun bir şekilde…

**Files*** Transkript R2**

| | | | | | | |
|---|---|---|---|---|---|---|
| No | | 0,0081 | 1 | | | |
| | | | | 1 | ZS | 5.12.2019 12:42 |

Relative clauselar, noun clauselar olabilir.Devrik cümleler olabilir. Inverted yapılar özellikle. Transitionlarla, bağlaçlarla kurulan devrik cümleler olabilir.

| Aggregate | Classification | Coverage | Number Of Coding References | Reference Number | Coded By Initials | Modified On |
|---|---|---|---|---|---|---|
| **Files*** ile Görüşme R5** | | | | | | |
| No | | 0,0397 | 1 | | | |
| | | | | 1 | ZS | 5.12.2019 12:45 |

Ya mesela devrik cümle yapılarında doğru kullanımlar var mı. Mesela neither nor yapılarını doğru kullanmış mı, not only but also yapıları…işte adverbial clauselarda var ya da adjective clauselardan adjective phraselere çevirebilmiş mi kullanabilmiş mi,adverb phraselerde participle kullanabilmiş mi, bunların hepsi benim için bi complex yapılara girer. Tabi doğru kullanıldığı sürece.

**Nodes\\The Role of LD in Scoring**

**Document**

**Files*** Transkript R2**

| | | | | | | |
|---|---|---|---|---|---|---|
| No | | 0,0195 | 1 | | | |
| | | | | 1 | ZS | 5.12.2019 13:35 |

Yine dil notuna katkısı var hem bence. Belki organizasyon olmayabilir ama notuna katkısı var ama dil doğru kullanılıyorsa onun tabi ki içeriğe de etkisi oluyor. Bir öğrencinin sürekli aynı fikri “thing” diyerek açıklamasıyla sürekli “cause” diyerek açıklamasını ya da bir yerde “cause” bir yerde “reason” kullanark açıklamsı birbirinden çok farklı. O illa ki içerik notunu da etkiliyordur.

**Files*** ile Görüşme R6**

| | | | | | | |
|---|---|---|---|---|---|---|
| No | | 0,0288 | 1 | | | |
| | | | | 1 | ZS | 5.12.2019 13:36 |

Eğer kelime kullanımı gayet etkinse öğrencinin tabi ki içeriği de etkiliyor. Ay ne kadar güzel yazmış, kendini ne güzel ifade etmiş diyorum. Eğer yoksa thing, thing, thing deyip duruyorlar. It is a thing, it is a thing… Habire relative clause kullanıyor. Mesela o bana vocabularynin yetersiz olduğunu gösteriyor. O zaman içeriği de etkiliyor.

**Files*** Transkript R7**

| | | | |
|---|---|---|---|
| No | 0,0189 | 1 | |

| | | |
|---|---|---|
| 1 | ZS | 5.12.2019 13:38 |

Bakıyorum ama dediğim gibi doğru kullanmayınca hiçbir şeye yaramıyor. Çeşitlilik sağlamış oluyor ama o çeşitlilik düzgün olmayınca yanlış ifade edilince bu sefer lack of communication ya da miscommunication oluyor. Sözcük güzel, farklı, değişik ama o bağlamda olmayınca ister istemez wrong word usage olmuş oluyor.



| Aggregate | Classification | Coverage | Number Of Coding References | Reference Number | Coded By Initials | Modified On |
|---|---|---|---|---|---|---|
| **Files*** ile Görüşme R5** | | | | | | |
| No | | 0,0300 | 1 | | | |

| | | |
|---|---|---|
| 1 | ZS | 5.12.2019 13:37 |

sürekli aynı şeylerle ya da çok basit duyulmuş kelimelerle, gündelik kullanımlarla kullandığı zaman tabi ki yine doğruysa gene puan kırmıyorum ama değişik kelimeler kullanan öğrencilerin kağıdını okumaktan büyük zevk alıyorum. Mutlaka olumlu yönde ya da puan artışı yönünde getirisi vardır.

**Files*** İle Görüşme R4**

| | | | |
|---|---|---|---|
| No | 0,0359 | 2 | |

| | | |
|---|---|---|
| 1 | ZS | 5.12.2019 13:39 |

Bir ekstra öğrenciye credit verebileceğim şey de kullandığı sözcüklerin seçimleri oluyor. Eğer öğrenci gayret sarf ederek ne bileyim eş anlamlı iki sözcüğü yan yana kullanmışsa veya mesela bir deyimsel ifade kullanmışsa veyahut bir ne bileyim işte konuya uygunsa. İyi kötü bir atasözü veya özdeyişsel bir şey katmışsa onlar da öğrenciye ekstra puan vermeme neden olan şeyler oluyor.

| | | |
|---|---|---|
| 2 | ZS | 5.12.2019 13:40 |

Yani ekstra puan vereceğim şeylerden bir tanesi işte mesela merak edip de bakmışsa çünkü onu anlıyorsunuz. Öğrencide o bir gayret… Yani mesela şöyle bir şey var benim biraz evvel örneklerini okuduğum öğrencilerden bir tanesi bakıyor thesaurusa. Ama gidiyor yanlışlarını kullanıyor. Yalnız mesela ben ondan not kırmıyorum. Diyorum ki bak her önüne geleni yazma. Git bir de Türkçesine bak çünkü bulduğun şey yanlış oluyor. Yani bunu göstermeye çalışıyorum ama ondan not kırmıyorum çünkü çok ciddi bir gayret gösteriyor.

**Nodes\\The Role of SC in Scoring**

**Document**

**Files*** İle Görüşme R3**

| | | |
|---|---|---|
| No | 0,0384 | 3 |

| | | |
|---|---|---|
| 1 | ZS | 5.12.2019 13:17 |

biz onlara diyoruz ki birincisi siz zaten İngilizce öğretmeni olacaksınız dolayısıyla böyle bir level of mastery olması lazım. Dolayısıyla farklı yapıları da kullanabildiğinizi bize göstermeniz lazım. O yüzden sadece basit cümleler kullanıyorsanız bu size gramerden sadece doğru yazdığınız için tam puan getirmez. Dolayısıyla o karmaşıklık bekliyorum.

| | | |
|---|---|---|
| 2 | ZS | 5.12.2019 13:33 |

Biz şimdi writingte özellikle ne bekliyoruz? Bir yapı bekliyoruz onlardan. İşte bir thesis statementın olacak. Onu destekleyen topic sentencelar ve her bir topic sentencei destekleyen major ve minor idealar olacak. Dolayısıyla biz düşünce akışını görmek istiyoruz zaten.

| | | |
|---|---|---|
| 3 | ZS | 5.12.2019 13:34 |

O akışı da böyle çok basit cümlelerle değil, böyle bol bol, farklı, birbirinden farklı bağlaçlar kullanarak, işte farklı cümle yapıları kurarak ama yine de kendi stillerini taşıyan bir yapı bekliyoruz.



| Aggregate | Classification | Coverage | Number Of Coding References | Reference Number | Coded By Initials | Modified On |
|---|---|---|---|---|---|---|
| **Files*** ile Görüşme R1** | | | | | | |
| No | | 0,0267 | 2 | | | |
| | | | | 1 | ZS | 5.12.2019 13:26 |

Kendini complex anlamına getireceğim derken anlam kayması oluyor. Ne dediğini anlayamayabiliyoruz. Ona da dikkat edilmesi lazım. Yani yerinde. Şimdi bu da yazmada çok karşılaştığımız bir problemdir. Mesela öğrenci bilir burada discourse markerlardan not alacağım diye yerli yersiz discourse marker kullanır.

| | | |
|---|---|---|
| 2 | ZS | 5.12.2019 13:24 |

Mesela “I like, I dislike, I like, I dislike”. Mesela o zaman hiç şey yok ama öbürü bir sürü bir şeyler yazmaya çalışmış, hata yapmış aman çizip çizip hani red ocean deriz ondan sonra grameri kötüdür diye. Hani en azından deniyor, risk alıyor. Biraz da onun da yüreklendirilmesi lazım. Ama bu da her cümleye de çok not verelim, yok yüksek not alsın değil. Yani o balancei tutturmak gerekir. Yani çocuğa da şu mesajı vereceksin. Complex cümle yapısı da lütfen üret.

**Files*** Transkript R2**

| | | |
|---|---|---|
| No | 0,0368 | 2 |

| | | |
|---|---|---|
| 1 | ZS | 5.12.2019 13:25 |

Language notunu tabi ki etkiliyor. Doğu da kullanımı yani bazen bunu biliyorlar. Kullanmaya çalışıyorlar fakat yanlış kullanıyorlar. Türkçeden direkt direct transitionlar gibi yanlış da kullanıyorlar. Yine de bir attempt olduğu için o yapıyı görmüş kullanmaya çalışıyor. Yanlış bile kullansa önemli benim için. Ama özellikle doğru da kullanıyorsa dil puanını yüksek vermeye çalışıyorum.

| | | |
|---|---|---|
| 2 | ZS | 5.12.2019 13:32 |

Organizationında bir parçası bu. Aslında paragraf içi düzenin de bir parçası. Yapıların düzgün kullanımı organizasyonu da aslında etkiliyor. İçeriğin yoğunluğunu etkilemese bile. İçerik çok dolu çok basit cümleler yazan öğrenciler de var. Ama organizasyon notun da etkiliyordur diye tahmin ediyorum. %50’ye yakın bir nota katkısı oluyordur bence.

**Files*** ile Görüşme R6**

| | | |
|---|---|---|
| No | 0,0179 | 2 |

| | | |
|---|---|---|
| 1 | ZS | 5.12.2019 13:21 |

sadece language useda kullanıyorum bunu açıkçası. Hani eğer onları düzgün kullandıysa yüksek not alıyor. Simple ama accurate ise yüksek not alamıyor.

| | | |
|---|---|---|
| 2 | ZS | 5.12.2019 13:21 |

Çünkü ELT öğrencisinden beklentim benim simple olması değil.

**Files*** İle Görüşme R4**

| | | |
|---|---|---|
| No | 0,0228 | 2 |

| | | |
|---|---|---|
| 1 | ZS | 5.12.2019 13:29 |

Eğer öğrenci kendi aklınca sofistike olmaya çalışırken her şeyi birbirine kattıysa zaten anlatılamaz hale geliyor. Content ortadan kalktığı için de tabi daha düşük not alıyor.

| Aggregate | Classification | Coverage | Number Of Coding References | Reference Number | Coded By Initials | Modified On |
|---|---|---|---|---|---|---|
| | | | | 2 | ZS | 5.12.2019 13:31 |

İlla ki relative clause kullanmasa bile anlattıklarını adam gibi yani minimalist bile anlatsa, anlattığını anlatıyorsa o beni rahatsız etmiyor. Hani bu öğrenci niye sofistike, daha sofistike bir yapı kullanmış demiyorum. Hiç olmazsa bildikleri içerisinde bir şeyler anlatmış. Ama hani bir şey… Dostlar alışverişte görsün şeklinde birbirinin içine girmiş, birbirinin ardına ne olduğu anlaşılmayan



**The Latest Version of Theme-Code Organization**

19.12.2019 14:58

Coding Summary By Code

19.12.2019 14:58

| Aggregate | Classification | Coverage | Number Of Coding References | Reference Number | Coded By Initials | Modified On |
|---|---|---|---|---|---|---|
| **Node** | | | | | | |
| **Nodes\\An Analogy Between Complexity and Diversity** | | | | | | |
| **Document** | | | | | | |
| **Files*** İle Görüşme R3** | | | | | | |
| | No | 0,0091 | 1 | | | |
| | | | | 1 | ZS | 5.12.2019 12:52 |

Yani hani demiştik ya dil yapılarını, farklı yapıları da kullanmalarını istiyoruz. Tek bir yapı üzerinden gitmesinler, sadece basit cümle kullanmasınlar vs. Sözcük çeşitliliğinde de aynı şekilde.

| | | | | | | |
|---|---|---|---|---|---|---|
| **Files*** ile Görüşme R1** | | | | | | |
| | No | 0,0027 | 1 | | | |

1 ZS 16.12.2019 12:39

Dolayısıyla hem kelime anlamında variety yok, hem grammar anlamında variety yok

**Files*** ile Görüşme R6**

No 0,0215 1

1 ZS 16.12.2019 12:37

Atıyorum daha althoughlar, ifler, whenler… Bağlaçların kullanıldığı, bir gerund ile yapılan reductionlarla birlikte… Hoş, reduction çok akademik writingte kullanılmaz ama yine de daha variety yapabileceği, simple sentence değil de… Öyle bir şey algılıyorum



| Aggregate | Classification | Coverage | Number Of Coding References | Reference Number | Coded By Initials | Modified On |
|---|---|---|---|---|---|---|

**Nodes\\Embodying LD as a Construct**

**Document**

**Files*** İle Görüşme R3**

No 0,0124 1

1 ZS 5.12.2019 12:51

Yani bir konu hakkında yazıyorsanız o konu hakkında ya da o alana ait ne kadar geniş bir jargona sahipsiniz bunu göstermenizi bekliyoruz diyoruz öğrencilere. O yüzden de hem farklı kelime türlerini hem de mümkün olduğunca farklı kelimeleri kullanmalarını istiyoruz.

**Files*** ile Görüşme R1**

No 0,0109 1

1 ZS 5.12.2019 12:49

Ben hani lexicology dersi de verdiğim için özellikle kelime haznesinin genişliğinden bahsederken söylediğimiz şeyler var. Vocabulary size çok sığ bir şey. Bunun içerisinde depth ve breadth dediğimiz şeyler var. Depth dediğim şeylerde de yani aynı kelimenin birkaç anlamı olabilir, hangi contextte neyi kullanacağız,

**Files*** ile Görüşme R6**

No 0,0174 1

1 ZS 5.12.2019 12:50

Yine simple olmayan, çeşitli kelimeler kullandığı, çeşitliliği olan, daha üst düzey, advanced düzey dediğimiz ya da burada upper-intermediate düzey olabilir. Kelimeler kullanabileceği bir şey anlıyorum yani.

**Files*** Transkript R7**

No 0,0073 1

1 ZS 5.12.2019 12:55

sözcüklerin zorluk derecesi aklıma geliyor. İşte commonly used words dışında bir kelime kullanıyor mu, doğru kullanıyor mu?

**Files*** ile Görüşme R5**

No 0,0237 1

1 ZS 5.12.2019 12:54

bir kere konuya uygun ve o kelimenin anlamını doğru bir yer de kullanabilmiş mi o, kelime yapısını doğru olarak kullanmış mı o, mesela adjective yerine adverb mü koymuş, noun yerine adjective mi getirmiş, tabi bunlar da çok önemli



| Aggregate | Classification | Coverage | Number Of Coding References | Reference Number | Coded By | Modified On |
|---|---|---|---|---|---|---|

**Nodes\\Embodying LD as a Construct\Examples of Lexical Diversity**

**Document**

**Files*** İle Görüşme R8**

No 0,0050 1

1 ZS 5.12.2019 13:11

You make me frustrated" yerine "You frustrate me" gibi. Yani orada farklı… Kelimenin noun halini, verb halini, adjective haline hakim olup

**Files*** İle Görüşme R3**

| | | |
|---|---|---|
| No | 0,0037 | 1 |

1 ZS 5.12.2019 13:05

Duvarı tasvir edecek. İşte light blue falan demiyor. Bluenun shadeini veriyor.

**Files*** ile Görüşme R1**

| | | |
|---|---|---|
| No | 0,0128 | 1 |

1 ZS 5.12.2019 13:02

synonymler, antonymler, collocationlar var, idiomatic expressionlar ve chunklar var akademik anlamda. Bunlar önemli. Mesela biz cause & effect analysis essay sorduk. Orda işte causeları yazacak ya da effectleri yazacak. Hep causeda "first cause, first reason" mı desin? Farklı "impact"di "influence"dı falan gibi cümleleri de… Pardon, kelimeleri de kullanmasını bekleriz.

**Files*** Transkript R2**

| | | |
|---|---|---|
| No | 0,0267 | 2 |

1 ZS 5.12.2019 13:10

Yine diyebileceğim gibi ileri seviyede belki özellikle noun formların daha yoğun olduğu. Çoğu fiilin noun hali daha advanced diye falan kabul ediliyor. İşte dolayısıyla daha az karşımıza çıkan kelimeler olabilir.

2 ZS 5.12.2019 13:01

O konunun çerçevesi içerisinde advanced kelimeleri de kullanabilmesiyle ilgili. Mesela sürekli "thing" falan yazılıyorlar. Mesela sürekli "thing" demektense bir seferinde "reason" desin. "Compose" desin vs. vs. Dolayısıyla bu da bir çeşitlilik. Eş anlamları da kullanmak. Noun formları düzgün kullanabilmek bir parçası.

**Files*** ile Görüşme R6**

| | | |
|---|---|---|
| No | 0,0036 | 1 |

1 ZS 5.12.2019 13:07

Mesela important demez de significant der.

19.12.2019 14:58

| Aggregate | Classification | Coverage | Number Of Coding References | Reference Number | Coded By Initials | Modified On |
|---|---|---|---|---|---|---|
| **Files*** ile Görüşme R5** | | | | | | |
| | No | 0,0213 | 1 | | | |
| | | | | 1 | ZS | 5.12.2019 13:08 |

ımmm, adverb kullanımları özellikle çok önemli, doğru yerlerde kullanılması ve mesela hani reading derslerinde karşılaştıkları yapılar olabilir chunklar halinde bazı yapılar olabilir, phrasal verbler olabilir

| | | | | | | |
|---|---|---|---|---|---|---|
| **Files*** İle Görüşme R4** | | | | | | |
| | No | 0,0082 | 1 | | | |
| | | | | 1 | ZS | 5.12.2019 13:06 |

Hiç üşenmeden bakıverseler red yerine farklı bir mesela şey kullanabilirler. Ne bileyim "crimson" derler. Yani ama hepsini red, red, red flowers yerine "crimson" der. İşte ne bileyim "It was blodly red" der

| | | | | | | |
|---|---|---|---|---|---|---|
| **Nodes\\Embodying SC as a Construct** | | | | | | |
| **Document** | | | | | | |
| **Files*** İle Görüşme R3** | | | | | | |
| | No | 0,0195 | 2 | | | |
| | | | | 1 | ZS | 5.12.2019 12:39 |

Kendi aslında yazım stilimi anlatıyor. Bana hep şey eleştirisi gelir mesela. Sen kısa cümle kuramıyor musun? Yani clauselar havada uçuşur… Böyle bir cümle başlar bitene kadar böyle bir süner. Ama aslında gramer olarak doğrudur, anlamsal olarak da doğrudur.

| | | | | | | |
|---|---|---|---|---|---|---|
| | | | | 2 | ZS | 5.12.2019 12:40 |

Hani syntactic complexity olan da sanki anlamı vermek için biraz daha dolandıran, daha fazla grammatical işte yapı kullanan… Bir yazım stili canlandı aklımda.

| | | | | | | |
|---|---|---|---|---|---|---|
| **Files*** ile Görüşme R1** | | | | | | |
| | No | 0,0346 | 1 | | | |
| | | | | 1 | ZS | 5.12.2019 11:02 |

İşte subject-verb agreementına dikkat ediyor, işte belli başlı bildiği tensler ya da bildiği conjunctionları kullanıyor. Dolayısıyla embedded structurelar, clauselar falan hiçbir şey kullanmamaya çalışıyor. Ama bu da tabi ki sophistication anlamında öğrencinin son derece yalın bir şey yazmasına sebep oluyor. Dolayısıyla hem kelime anlamında variety yok, hem grammar anlamında variety yok, işte echoic termleri kullanmıyor; synonym, antonymleri kullanmıyor; and, but, so onlardan başka bağlaç bir sürü öğretilmesine rağmen kullanmıyor. Dolayısıyla hem paragrafta cümleleri birbirine bağlarken conjunctionları eksik kullanıyor hem de metin içerisinde paragrafları birbirine bağlarken de… Yani belli başlı cümle ya da phraseleri kullanması lazım ki bir önceki ya da sonrakine atıfta bulunsun. Bunları ne yapıyor? Gözden kaçırmış oluyor. Dolayısıyla sadece bir topic sentence yazarım, bir iki cümle yazarım bir de conclusion sentence yaparım olayı bitiririm diyor. Ama böyle bir şey olmaması gerekiyor.



| Aggregate | Classification | Coverage | Number Of Coding References | Reference | Number | Coded By Initials | Modified On |
|---|---|---|---|---|---|---|---|
| **Files*** Transkript R2** | | | | | | | |
| No | | 0,0191 | 2 | | | | |
| | | | | | 1 | ZS | 5.12.2019 12:35 |

Karmaşıklık deyince Türkçede daha kötü bir şeyleri çağrıştıyor ama İngilizce de syntactic complexity bildiğim için yani tahmin edebildiğim için daha olumlu bir şeyi çağrıştırıyor. Using variety of structures gibi düşünüyorum. Daha çok complex ve compound structures da kullanabilmeleri gerektiğini düşünüyorum

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | | | | | 2 | ZS | 5.12.2019 12:37 |

syntactic complexityden bir yazarın dili kullanma becerisini anlıyorum.

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| **Files*** Görüşme R5** | | | | | | | |
| No | | 0,0416 | 1 | | | | |
| | | | | | 1 | ZS | 5.12.2019 12:38 |

Bu ifaden neler çağrıştırıyor? Bu ifadede bağlaç kullanımları yani transitionlardan bahsediyorum conjunctionslardan bahsediyorum, clauseların kullanımlarından bahsediyorum yani adverbial clause, adjective clause,noun clause kullanımlarından bahsediyorum. Bunların hepsinin beraber doğru bir şekilde kullanılıp kullanılması benim için complex cümle bu yoksa subject verb object o basit bir cümle yapısıdır.

**Nodes\\Embodying SC as a Construct\Examples of Syntactic Complexity**

**Document**

**Files*** İle Görüşme R8**

| | | |
|---|---|---|
| No | 0,0190 | 1 |

| | | |
|---|---|---|
| 1 | ZS | 5.12.2019 12:46 |

Yani tabi ki farklı clauseları kullanabilmesi gerekiyor bir şekilde. Sadece relative clause falan değil ama adverb olsun, işte ondan sonra farklı transitionlar olsun… Ondan sonra farklı… Yani bir şekilde pasive kullanabilsin. Arada bir kelimeleri değiştirsin. Kelimeleri değiştirdiği için cümlenin yapısı değişsin. Onun üzerine uğraşsın. Orada prepositionları bilsin. O prepositionların arkasından bambaşka, saçma bir şey geldiğinde de kafası karışmasın istiyorum. Yani -ing gelsin of'tan sonra ama onun kafasını karıştırmasın.

**Files*** ile Görüşme R1**

| | | |
|---|---|---|
| No | 0,0085 | 1 |

| | | |
|---|---|---|
| 1 | ZS | 5.12.2019 12:44 |

Yani inversionı kullansın, işte embedded structurelar kullansın, iki, üç cümleyi değişik değişik conjunctionlar kullanarak bir arada kullanmayı becersin… Değişik işte phrasal verbler, idiomatic expressionlar tabi akademik anlamda uygun bir şekilde…



| Aggregate | Classification | Coverage | Number Of Coding References | Reference Number | Coded By Initials | Modified On |
|---|---|---|---|---|---|---|
| **Files*** Transkript R2** | | | | | | |
| No | | 0,0081 | 1 | | | |
| | | | 1 | | ZS | 5.12.2019 12:42 |

Relative clauselar, noun clauselar olabilir.Devrik cümleler olabilir. Inverted yapılar özellikle. Transitionlarla, bağlaçlarla kurulan devrik cümleler olabilir.

**Files*** ile Görüşme R5**

| | | |
|---|---|---|
| No | 0,0397 | 1 |

| | | |
|---|---|---|
| 1 | ZS | 5.12.2019 12:45 |

Ya mesela devrik cümle yapılarında doğru kullanımlar var mı. Mesela neither nor yapılarını doğru kullanmış mı, not only but also yapıları…işte adverbial clauselarda var ya da adjective clauselardan adjective phraselere çevirebilmiş mi kullanabilmiş mi,adverb phraselerde participle kullanabilmiş mi, bunların hepsi benim için bi complex yapılara girer. Tabi doğru kullanıldığı sürece.

**Nodes\\The Role of LD in Scoring**

**Document**

**Files*** Transkript R2**

| | | |
|---|---|---|
| No | 0,0195 | 1 |

| | | |
|---|---|---|
| 1 | ZS | 5.12.2019 13:35 |

Yine dil notuna katkısı var hem bence. Belki organizasyon olmayabilir ama notuna katkısı var ama dil doğru kullanılıyorsa onun tabi ki içeriğe de etkisi oluyor. Bir öğrencinin sürekli aynı fikri "thing" diyerek açıklamasıyla sürekli "cause" diyerek açıklamasını ya da bir yerde "cause" bir yerde "reason" kullanark açıklamsı birbirinden çok farklı. O illa ki içerik notunu da etkiliyordur.

**Files*** ile Görüşme R6**

| | | |
|---|---|---|
| No | 0,0288 | 1 |

| | | |
|---|---|---|
| 1 | ZS | 5.12.2019 13:36 |

Eğer kelime kullanımı gayet etkinse öğrencinin tabi ki içeriği de etkiliyor. Ay ne kadar güzel yazmış, kendini ne güzel ifade etmiş diyorum. Eğer yoksa thing, thing, thing deyip duruyorlar. It is a thing, it is a thing… Habire relative clause kullanıyor. Mesela o bana vocabularynin yetersiz olduğunu gösteriyor. O zaman içeriği de etkiliyor.



| Aggregate | Classification | Coverage | Number Of Coding References | Reference Number | Coded By Initials | Modified On |
|---|---|---|---|---|---|---|
| **Files*** Transkript R7** | | | | | | |
| | No | 0,0189 | 1 | | | |
| | | | | 1 | ZS | 5.12.2019 13:38 |

Bakıyorum ama dediğim gibi doğru kullanmayınca hiçbir şeye yaramıyor. Çeşitlilik sağlamış oluyor ama o çeşitlilik düzgün olmayınca yanlış ifade edilince bu sefer lack of

communication ya da miscommunication oluyor. Sözcük güzel, farklı, değişik ama o bağlamda olmayınca ister istemez wrong word usage olmuş oluyor.

**Files*** ile Görüşme R5**

No 0,0300 1

1 ZS 5.12.2019 13:37

sürekli aynı şeylerle ya da çok basit duyulmuş kelimelerle, gündelik kullanımlarla kullandığı zaman tabi ki yine doğruysa gene puan kırmıyorum ama değişik kelimeler kullanan öğrencilerin kağıdını okumaktan büyük zevk alıyorum. Mutlaka olumlu yönde ya da puan artışı yönünde getirisi vardır.

**Files*** İle Görüşme R4**

No 0,0359 2

1 ZS 5.12.2019 13:39

Bir ekstra öğrenciye credit verebileceğim şey de kullandığı sözcüklerin seçimleri oluyor. Eğer öğrenci gayret sarf ederek ne bileyim eş anlamlı iki sözcüğü yan yana kullanmışsa veya mesela bir deyimsel ifade kullanmışsa veyahut bir ne bileyim işte konuya uygunsa. İyi kötü bir atasözü veya özdeyişsel bir şey katmışsa onlar da öğrenciye ekstra puan vermeme neden olan şeyler oluyor.

2 ZS 5.12.2019 13:40

Yani ekstra puan vereceğim şeylerden bir tanesi işte mesela merak edip de bakmışsa çünkü onu anlıyorsunuz. Öğrencide o bir gayret… Yani mesela şöyle bir şey var benim biraz evvel örneklerini okuduğum öğrencilerden bir tanesi bakıyor thesaurusa. Ama gidiyor yanlışlarını kullanıyor. Yalnız mesela ben ondan not kırmıyorum. Diyorum ki bak her önüne geleni yazma. Git bir de Türkçesine bak çünkü bulduğun şey yanlış oluyor. Yani bunu göstermeye çalışıyorum ama ondan not kırmıyorum çünkü çok ciddi bir gayret gösteriyor.



| Aggregate | Classification | Coverage | Number Of Coding References | Reference Number | Coded By | Modified On |
|---|---|---|---|---|---|---|

**Nodes\\The Role of LD in Scoring\Correct Use of Words**

**Document**

**Files*** Transkript R7**

| | | |
|---|---|---|
| No | 0,0189 | 1 |

1 ZS 5.12.2019 13:48

Bakıyorum ama dediğim gibi doğru kullanmayınca hiçbir şeye yaramıyor. Çeşitlilik sağlamış oluyor ama o çeşitlilik düzgün olmayınca yanlış ifade edilince bu sefer lack of communication ya da miscommunication oluyor. Sözcük güzel, farklı, değişik ama o bağlamda olmayınca ister istemez wrong word usage olmuş oluyor.

**Nodes\\The Role of SC in Scoring**

**Document**

**Files*** İle Görüşme R3**

| | | |
|---|---|---|
| No | 0,0384 | 3 |

1 ZS 5.12.2019 13:17

biz onlara diyoruz ki birincisi siz zaten İngilizce öğretmeni olacaksınız dolayısıyla böyle bir level of mastery olması lazım. Dolayısıyla farklı yapıları da kullanabildiğinizi bize göstermeniz lazım. O yüzden sadece basit cümleler kullanıyorsanız bu size gramerden sadece doğru yazdığınız için tam puan getirmez. Dolayısıyla o karmaşıklık bekliyorum.

2 ZS 5.12.2019 13:33

Biz şimdi writingte özellikle ne bekliyoruz? Bir yapı bekliyoruz onlardan. İşte bir thesis statementın olacak. Onu destekleyen topic sentencelar ve her bir topic sentencei destekleyen major ve minor idealar olacak. Dolayısıyla biz düşünce akışını görmek istiyoruz zaten.

3 ZS 5.12.2019 13:34

O akışı da böyle çok basit cümlelerle değil, böyle bol bol, farklı, birbirinden farklı bağlaçlar kullanarak, işte farklı cümle yapıları kurarak ama yine de kendi stillerini taşıyan bir yapı bekliyoruz.

**Files*** ile Görüşme R1**

| | | |
|---|---|---|
| No | 0,0267 | 2 |

1 ZS 5.12.2019 13:26

Kendini complex anlamına getireceğim derken anlam kayması oluyor. Ne dediğini anlayamayabiliyoruz. Ona da dikkat edilmesi lazım. Yani yerinde. Şimdi bu da yazmada çok karşılaştığımız bir problemdir. Mesela öğrenci bilir burada discourse markerlardan not alacağım diye yerli yersiz discourse marker kullanır.

2 ZS 5.12.2019 13:24

Mesela “I like, I dislike, I like, I dislike”. Mesela o zaman hiç şey yok ama öbürü bir sürü bir şeyler yazmaya çalışmış, hata yapmış aman çizip çizip hani red ocean deriz ondan sonra grameri kötüdür diye. Hani en azından deniyor, risk alıyor. Biraz da onun da yüreklendirilmesi lazım. Ama bu da her cümleye de çok not verelim, yok yüksek not alsın değil. Yani o balancei tutturmak gerekir. Yani çocuğa da şu mesajı vereceksin. Complex cümle yapısı da lütfen üret.



| Aggregate | Classification | Coverage | Number Of Coding References | Reference Number | Coded By Initials | Modified On |
|---|---|---|---|---|---|---|
| **Files*** Transkript R2** | | | | | | |
| No | | 0,0368 | 2 | | | |
| | | | | 1 | ZS | 5.12.2019 13:25 |

Language notunu tabi ki etkiliyor. Doğu da kullanımı yani bazen bunu biliyorlar. Kullanmaya çalışıyorlar fakat yanlış kullanıyorlar. Türkçeden direkt direct transitionlar gibi yanlış da kullanıyorlar. Yine de bir attempt olduğu için o yapıyı görmüş kullanmaya çalışıyor. Yanlış bile kullansa önemli benim için. Ama özellikle doğru da kullanıyorsa dil puanını yüksek vermeye çalışıyorum.

| | | | | | | |
|---|---|---|---|---|---|---|
| | | | | 2 | ZS | 5.12.2019 13:32 |

Organizationında bir parçası bu. Aslında paragraf içi düzenin de bir parçası. Yapıların düzgün kullanımı organizasyonu da aslında etkiliyor. İçeriğin yoğunluğunu etkilemese bile. İçerik çok dolu çok basit cümleler yazan öğrenciler de var. Ama organizasyon notun da etkiliyordur diye tahmin ediyorum. %50’ye yakın bir nota katkısı oluyordur bence.

| | | | | | | |
|---|---|---|---|---|---|---|
| **Files*** ile Görüşme R6** | | | | | | |
| No | | 0,0179 | 2 | | | |
| | | | | 1 | ZS | 5.12.2019 13:21 |

sadece language useda kullanıyorum bunu açıkçası. Hani eğer onları düzgün kullandıysa yüksek not alıyor. Simple ama accurate ise yüksek not alamıyor.

| | | | | | | |
|---|---|---|---|---|---|---|
| | | | | 2 | ZS | 5.12.2019 13:21 |

Çünkü ELT öğrencisinden beklentim benim simple olması değil.

| | | | | | | |
|---|---|---|---|---|---|---|
| **Files*** İle Görüşme R4** | | | | | | |
| No | | 0,0228 | 2 | | | |

| | | | |
|---|---|---|---|
| | 1 | ZS | 5.12.2019 13:29 |

Eğer öğrenci kendi aklınca sofistike olmaya çalışırken her şeyi birbirine kattıysa zaten anlatılamaz hale geliyor. Content ortadan kalktığı için de tabi daha düşük not alıyor.

| | | | |
|---|---|---|---|
| | 2 | ZS | 5.12.2019 13:31 |

İlla ki relative clause kullanmasa bile anlattıklarını adam gibi yani minimalist bile anlatsa, anlattığını anlatıyorsa o beni rahatsız etmiyor. Hani bu öğrenci niye sofistike, daha sofistike bir yapı kullanmış demiyorum. Hiç olmazsa bildikleri içerisinde bir şeyler anlatmış. Ama hani bir şey… Dostlar alışverişte görsün şeklinde birbirinin içine girmiş, birbirinin ardına ne olduğu anlaşılmayan



| Aggregate | Classification | Coverage | Number Of Coding References | Reference Number | Coded By Initials | Modified On |
|---|---|---|---|---|---|---|
| **Nodes\\The Role of SC in Scoring\Correct Use of Structures** | | | | | | |
| **Document** | | | | | | |
| **Files*** ile Görüşme R1** | | | | | | |
| No | | 0,0105 | 1 | | | |
| | | | 1 | | ZS | 5.12.2019 13:49 |

Kendini complex anlamına getireceğim derken anlam kayması oluyor. Ne dediğini anlayamayabiliyoruz. Ona da dikkat edilmesi lazım. Yani yerinde. Şimdi bu da yazmada çok karşılaştığımız bir problemdir. Mesela öğrenci bilir burada discourse markerlardan not alacağım diye yerli yersiz discourse marker kullanır.

| | | | | | | |
|---|---|---|---|---|---|---|
| **Files*** Kür Transkript R2** | | | | | | |
| No | | 0,0194 | 1 | | | |
| | | | 1 | | ZS | 5.12.2019 13:50 |

Language notunu tabi ki etkiliyor. Doğu da kullanımı yani bazen bunu biliyorlar. Kullanmaya çalışıyorlar fakat yanlış kullanıyorlar. Türkçeden direkt direct transitionlar gibi yanlış da kullanıyorlar. Yine de bir attempt olduğu için o yapıyı görmüş kullanmaya çalışıyor. Yanlış bile kullansa önemli benim için. Ama özellikle doğru da kullanıyorsa dil puanını yüksek vermeye çalışıyorum.

**Files*** İle Görüşme R4**

| | | |
|---|---|---|
| No | 0,0227 | 2 |

| | | |
|---|---|---|
| 1 | ZS | 5.12.2019 13:50 |

Eğer öğrenci kendi aklınca sofistike olmaya çalışırken her şeyi birbirine kattıysa zaten anlatılamaz hale geliyor. Content ortadan kalktığı için de tabi daha düşük not alıyor.

| | | |
|---|---|---|
| 2 | ZS | 5.12.2019 13:50 |

İlla ki relative clause kullanmasa bile anlattıklarını adam gibi yani minimalist bile anlatsa, anlattığını anlatıyorsa o beni rahatsız etmiyor. Hani bu öğrenci niye sofistike, daha sofistike bir yapı kullanmış demiyorum. Hiç olmazsa bildikleri içerisinde bir şeyler anlatmış. Ama hani bir şey… Dostlar alışverişte görsün şeklinde birbirinin içine girmiş, birbirinin ardına ne olduğu anlaşılmayan

**Nodes\\The scoring procedure is like...**

**Document**

**Files*** ile Görüşme R1**

| | | |
|---|---|---|
| No | 0,0126 | 1 |

| | | |
|---|---|---|
| 1 | ZS | 19.12.2019 14:39 |

Önce essayi tamamen bir kere okuyorum. Ondan sonra bizim bir kriterimiz var. O kriteri de baz alarak teker teker bakıyorum. Mesela thesis statements ise thesis statement'a tekrar dönüyorum. Sonra developmental paragraphlarda topic sentence, major/minor detail diye tekrar dönüyorum. En son işte biz gramerdir, vocabularydir, unitydir, coherence gibi ona bakıyorum.



| Aggregate | Classification | Coverage | Number Of Coding References | Reference Number | Coded By Initials | Modified On |
|---|---|---|---|---|---|---|
| **Files*** Transkript R2** | | | | | | |
| No | | 0,0273 | 1 | | | |

| | | |
|---|---|---|
| 1 | ZS | 19.12.2019 14:46 |

genel yaklaşımımız essay çeşidine göre değişecek şekilde fakat mesela thesis statement , topic sentence gibi puanlar vermek bunun yanında işte language, content, organization, coherence ile ilgili farklı puanlamalar var. Bunların seyri değişebiliyor bazen. Bazı essay çeşitlerine göre mesela argumentine göre pro-con chartta çizdiriyorum. Ona da puan

ayırıyoruz.Bir yerlerden kırıyoruz belki onu. Ama temelde language, content, organization gibi puanları farklı yerlere bölüyoruz ama ana fikir cümleleri çok önemli, thesis ve topic sentencelar.

### Files*** ile Görüşme R6

| | | |
|---|---|---|
| No | 0,0090 | 1 |

1 ZS 19.12.2019 14:50

soru sorarak contente refer ederek gidiyorum. Organizasyona bakarak. Hepsini aynı anda yapmaya çalışıyorum.

### Nodes\\The scoring procedure is like...\SC and LD Overshadowed by Content and Organization

### Document

### Files*** Transkript R7

| | | |
|---|---|---|
| No | 0,0407 | 3 |

1 ZS 16.12.2019 12:44

İlk önce genel olarak contentten başlıyorum. Content ve organizaton odaklı gitmeyi seviyorum. Çünkü ister istemez mechanic kısım beni olumsuz etkileyebiliyor.

2 ZS 19.12.2019 14:54

Önce bütün content organizatonlara not verme sonra tekrar dönüp bütün mekanik kısmı birlikte kağıtları ayırmamaya çalışıyorum. Süreci anlatabildim mi? Önce 30 kağıdın content- organizationı bir sonrakinde o mechanic dediğimiz grammar, vocabulary bir de o şekilde yapıyorum

3 ZS 16.12.2019 12:45

Beni ister istemez hoca olarak güzel yazılmış iyi bir İngilizce ile yazılmış kağıt sanki içerik de iyi yazılmış gibi hissettirdiği için ilk önce bir ayırıyorum. Kendi kafamda ayırıyorum. Önce content ve organizasyona bak sonra dili incelersin diye.

### Files*** ile Görüşme R5

| | | |
|---|---|---|
| No | 0,0224 | 1 |

1 ZS 16.12.2019 12:48

mesela önce contentine bakıyorum ondan sonra mesela opinion essay de olması gerekenler neler onları beklediysek onlar var mı diye okuyorum sonra bi kere daha dönüyorum ‘’grammar’’e, ‘’spellingin’’e, ‘’punctuationın’’a

| Aggregate | Classification | Coverage | Number Of Coding References | Reference Number | Coded By Initials | Modified On |
|---|---|---|---|---|---|---|
| **Files*** İle Görüşme R4** | | | | | | |
| | No | 0,0347 | 2 | | | |
| | | | | 1 | ZS | 16.12.2019 12:49 |

Yani çok ince ayrıntılarına bakmadan. Öğrenci işte istenen konu hakkında bir şeyler anlatabilmiş mi anlatamamış mı, ne olmuş ne bitmiş ona bakıyorum. Bazılarında ıı yani bütünsel olarak hiçbir şey anlaşılamaz hale geliyor. O zaman da oturup tek tek gramer hatalarını, mantık hatalarını işaretleyerek belli bir notlandırma sistemine gitmek lazım. Ancak bütün konu üzerinde çok güzel fikir anlatmış dahi olsa öğrencinin zihinsel süreçler içerisinde mantıksal atlamaları oluyor. Yani benim ilk baktığım şey sanıyorum organizasyon. Arkasından işte yani içerik…

| | | | | | | |
|---|---|---|---|---|---|---|
| | | | | 2 | ZS | 16.12.2019 12:51 |

Complexy de dediğin zaman yani ben yine kişisel olarak yapısal complexity'den çok content complexity'e bakıyorum. Düşüncesi ne kadar iyi gelişiyor, ne yapıyor, ne ediyor diye. Şöyle söyleyeyim. Mesela öğrenciye diyorsun ki descriptive essay yaz diyorsun. Mesela bir resim veriyorsun gelen şey… Descriptive değil.

| | | | | | | |
|---|---|---|---|---|---|---|
| **Nodes\\The scoring procedure is like...\Scorers read more than once to score Document** | | | | | | |
| **Files*** İle Görüşme R3** | | | | | | |
| | No | 0,0095 | 1 | | | |
| | | | | 1 | ZS | 19.12.2019 14:42 |

ben essayleri okurken iki kere okurum. Birincisi şöyle bir üstten okurum. Arkasından da o ilgili, bizim o ölçütlerimizin karşısına şeyleri yazarım… Puanları yazarım. İşte topic sentence şunu hak ediyor vs

| | | | | | | |
|---|---|---|---|---|---|---|
| **Files*** ile Görüşme R1** | | | | | | |
| | No | 0,0027 | 1 | | | |
| | | | | 1 | ZS | 19.12.2019 14:41 |

Yani ilk kez bir kere okuyorum baştan sona ondan sonrasında bölüm bölüm okuyorum.

**Files*** ile Görüşme R6**

| | | |
|---|---|---|
| No | 0,0046 | 1 |

1 ZS 19.12.2019 14:50

Okuyorum sonra emin değilsem tekrar bir okuma yapıyorum



| Aggregate | Classification | Coverage | Number Of Coding References | Reference Number | Coded By | Modified On |
|---|---|---|---|---|---|---|
| **Files*** Transkript R7** | | | | | | |
| No | | 0,0203 | 1 | | | |

1 ZS 19.12.2019 14:52

Ben iki veya üç kere okuyorum Zafer maalesef. Bir okumada bütün şeylere odaklanamıyorum. Pratikliği kazansam da şey yapıyorum yanıltmasın beni diye önce content ve organizasyon ağırlıklı gidip sonra grammar, vocabulary, mechanics, punctuation neyse o tarafa kayıyorum. Bir de editing amaçlı okuyorum. Bir kağıt 3 kere elden geçmiş oluyor.

**Files*** ile Görüşme R5**

| | | |
|---|---|---|
| No | 0,0286 | 1 |

1 ZS 19.12.2019 14:56

Şimdi öncelikle kağıdı bir okuyorum hiçbir kritere dayalı değil ama. Sadece ve sadece bi ne yazmış nelerden bahsetmiş diye bir detaylı okumaya geçmeden önce bi ön fikir edinmek için komple bi okuyorum ondan sonra daha detaylı daha yavaş kriteri de önüme alarak o şekilde okuyorum

APPENDIX -9: TOEFL Writing Scoring Criteria that was used in the present study

ETS TOEFL

## iBT/Next Generation TOEFL Test
## Independent Writing Rubrics (Scoring Standards)

| Score | Task Description |
|---|---|
| 5 | **An essay at this level largely accomplishes all of the following:**<br>• effectively addresses the topic and task<br>• is well organized and well developed, using clearly appropriate explanations, exemplifications, and/or details<br>• displays unity, progression, and coherence<br>• displays consistent facility in the use of language, demonstrating syntactic variety, appropriate word choice, and idiomaticity, though it may have minor lexical or grammatical errors |
| 4 | **An essay at this level largely accomplishes all of the following:**<br>• addresses the topic and task well, though some points may not be fully elaborated<br>• is generally well organized and well developed, using appropriate and sufficient explanations, exemplifications, and/or details<br>• displays unity, progression, and coherence, though it may contain occasional redundancy, digression, or unclear connections<br>• displays facility in the use of language, demonstrating syntactic variety and range of vocabulary, though it will probably have occasional noticeable minor errors in structure, word form, or use of idiomatic language that do not interfere with meaning |
| 3 | **An essay at this level is marked by one or more of the following:**<br>• addresses the topic and task using somewhat developed explanations, exemplifications, and/or details<br>• displays unity, progression, and coherence, though connection of ideas may be occasionally obscured<br>• may demonstrate inconsistent facility in sentence formation and word choice that may result in lack of clarity and occasionally obscure meaning<br>• may display accurate but limited range of syntactic structures and vocabulary |
| 2 | **An essay at this level may reveal one or more of the following weaknesses:**<br>• limited development in response to the topic and task<br>• inadequate organization or connection of ideas<br>• inappropriate or insufficient exemplifications, explanations, or details to support or illustrate generalizations in response to the task<br>• a noticeably inappropriate choice of words or word forms<br>• an accumulation of errors in sentence structure and/or usage |
| 1 | **An essay at this level is seriously flawed by one or more of the following weaknesses:**<br>• serious disorganization or underdevelopment<br>• little or no detail, or irrelevant specifics, or questionable responsiveness to the task<br>• serious and frequent errors in sentence structure or usage |
| 0 | **An essay at this level** merely copies words from the topic, rejects the topic, or is otherwise not connected to the topic, is written in a foreign language, consists of keystroke characters, or is blank. |