

**USING THE AUTOMATED WRITING
EVALUATION SYSTEM IN EFL CONTEXT: THE
EFFICIENCY OF THE SYSTEM AND THE VIEWS
OF THE TEACHERS AND THE STUDENTS**

Yüksek Lisans Tezi

Hilal YILDIZ

Eskişehir 2021

**USING THE AUTOMATED WRITING EVALUATION SYSTEM IN EFL
CONTEXT: THE EFFICIENCY OF THE SYSTEM AND THE VIEWS OF THE
TEACHERS AND THE STUDENTS**

Hilal YILDIZ

MA THESIS

Department of Foreign Language Education

MA in English Language Teaching Program

Supervisor: Assoc. Prof. Dr. Safiye İpek KURU GÖNEN

Eskişehir

Anadolu University

Graduate School of Educational Sciences

June 2021

JÜRİ VE ENSTİTÜ ONAYI

ABSTRACT

USING THE AUTOMATED WRITING EVALUATION SYSTEM IN EFL CONTEXT: THE EFFICIENCY OF THE SYSTEM AND THE VIEWS OF THE TEACHERS AND THE STUDENTS

Hilal YILDIZ

Department of Foreign Language Education

MA Program in English Language Teaching

Anadolu University, Graduate School of Educational Sciences, June 2021

Supervisor: Assoc. Prof. Dr. Safiye İpek KURU GÖNEN

In order to help teachers reduce their workload and provide students instant feedback, automated writing evaluation (AWE) tools, which score papers and provide feedback immediately, have been used. To better understand the potential and limitations of these tools, this study has four main objectives: (1) to examine how AWE scores agree with instructors' scores given for students' essays (2) to investigate the effectiveness of AWE feedback in the context of error reduction (3) to explore how students and (4) the instructors view the use of AWE tools. To achieve these objectives, a mixed-method approach was adopted. The quantitative analyses were carried out to explore the correlation between the seven instructors' scores and AWE scores. To explore the effectiveness of AWE feedback in error correction, the changes in errors of the experimental group (n=18), receiving AWE feedback, and the control group (n=20), receiving teacher feedback, were analyzed. Moreover, written reflection reports of students and semi-structured interviews conducted with both students and the instructors were analyzed using the Constant Comparison Method. The results indicated that there was a positive and highly significant correlation between AWE scores and the instructors' scores. Both AWE feedback and teacher feedback were effective in reducing errors in 11 categories, whereas the AWE feedback might be more effective in reducing errors in mechanics and usage categories. Finally, it was revealed that both students' and teachers' views on the use of AWE were positive, albeit there were minor limitations expressed.

Keywords: AWE feedback, L2 writing assessment, Teacher feedback, Views of students and teachers

ÖZET

OTOMATİK YAZMA BECERİSİ DEĞERLENDİRME SİSTEMİNİN YABANCI DİL OLARAK İNGİLİZCE ÖĞRETİLEN ORTAMLARDA KULLANILMASI: SİSTEMİN ETKİNLİĞİ VE ÖĞRETMENLER İLE ÖĞRENCİLERİN GÖRÜŞLERİ

Hilal YILDIZ

Yabancı Diller Eğitimi Anabilim Dalı

İngilizce Öğretmenliği Programı

Anadolu Üniversitesi, Eğitim Bilimleri Enstitüsü, June 2021

Danışman: Doç. Dr. Safiye İpek KURU GÖNEN

Öğretmenlerin iş yüklerini azaltmalarını ve öğrencilere anında geri bildirim sağlamalarını sağlamak için yazılanları puanlayan ve anında geri bildirim sağlayan otomatik yazma becerisi değerlendirme (OYBD) araçları kullanılmaktadır. Bu araçların potansiyelini ve sınırlamalarını daha iyi anlamak için, bu çalışmanın dört ana hedefi vardır: (1) OYBD puanlarını ile öğretmenlerin öğrencilerin yazdıklarına verdiği puanların korelasyonunu incelemek (2) hata azaltma üzerinde OYBD geri bildiriminin etkililiğini araştırmak (3) öğrencilerin ve (4) eğitimcilerin OYBD araçlarının kullanımı hakkındaki görüşlerini öğrenmek. Bu hedeflere ulaşmak için karma yöntem yaklaşımı benimsendi. Yedi öğretmenin puanları ile OYBD puanları arasındaki ilişkiyi keşfetmek için nicel analizler gerçekleştirildi. OYBD geri bildiriminin hata düzeltmedeki etkililiğini araştırmak için OYBD geri bildirimi alan deney grubu (n=18) ve öğretmen geri bildirimi alan kontrol grubunun (n=20) hatalarındaki değişimler incelenmiştir. Ayrıca öğrencilerin yazılı yansıtma raporları ve hem öğrencilerle hem de öğretmenler ile yapılan yarı yapılandırılmış görüşmeler sabit karşılaştırma yöntemi ile analiz edilmiştir. Sonuçlar OYBD puanları ile eğitimcilerin puanları arasında pozitif ve önemli bir korelasyon olduğunu gösterdi. Hem OYBD hem de öğretmen geri bildirimleri 11 kategoride hataları azaltmada etkiliyken, OYBD geri bildiriminin mekanik ve kullanım hatalarını azaltmada daha etkili olabileceği görüldü. Son olarak, küçük sınırlamaları olduğu ifade edilse de, katılımcıların OYBD kullanımına ilişkin görüşlerinin olumlu olduğu ortaya çıktı.

Anahtar Sözcükler: OYBD geri bildirimi, Yabancı dilde yazmanın değerlendirmesi, Öğretmen geri bildirimi, Öğrenci ve öğretmen görüşleri.

ACKNOWLEDGEMENTS

The completion of this thesis could not have been possible without the contributions and extensive support of my esteemed advisor, Assoc. Prof. Dr. S. İpek KURU GÖNEN. It would be an understatement to say that I am grateful and indebted to her guidance and encouragement. Her dedication, keen interest, meticulous efforts, and commitment to detail have been mainly responsible for extending my efforts to improve and complete my thesis. Having the opportunity to receive her valuable feedback and insight has always been inspiring and motivating. I must express that I am honored to have the chance to carry out this work with her.

My appreciation also extends to Assoc. Prof. Dr. Özgür YILDIRIM and Asst. Prof. Dr. Gonca SUBAŞI for their contributions and helpful suggestions regarding the data collection instruments and data analyses. I would also like to thank Assoc. Prof. Dr. Ali MERÇ, Assoc. Prof. Dr. İlknur İSTİFÇİ and Assoc. Prof. Dr. Fadime YALÇIN ARSLAN for their valuable feedback and suggestions.

I would also like to express my gratitude towards the participants of the study. I am deeply grateful to all the students and the instructors for taking their time to contribute to the study. I am also thankful to them for sharing their experiences and views with me for the sake of the study.

I am also thankful to all my friends who have always motivated me to complete this study, and I would like to express my sincerest gratitude to my dear friends Gamze EMİR and Gonca Nur ÖLMEZ for providing their everlasting help and support at every step of the process. I must also thank my dearest friend Hayriye KONÇ for her words of encouragement and constant emotional support.

Last but not least, I want to offer my warmest thanks to my mother, Zeynep YILDIZ for her enduring patience, love and faith in me. She has always been my strength and inspiration. I will be forever indebted to my mother for the understanding and support she has given me throughout my life and this process. Without her sacrifices, allowing me to pursue my dreams, I wouldn't be able to accomplish anything. My thanks should also go to all my family members for always being there whenever I need them.

Hilal Yıldız

Eskişehir 2021

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.

Hilal YILDIZ

TABLE OF CONTENTS

	<u>Page</u>
COVER PAGE	i
FINAL APPROVAL FOR THESIS	ii
ABSTRACT.....	iii
ÖZET	iv
ACKNOWLEDGEMENTS	v
TABLE OF CONTENTS	vii
LIST OF TABLES	xi
LIST OF FIGURES.....	xiv
LIST OF ABBREVIATIONS	xv
1. INTRODUCTION	1
1.1. Background to the Study	1
1.2. Statement of the Problem	4
1.3. Purpose of the Study	7
1.4. Significance of the Study	7
1.5. Limitations of the Study	9
1.6 Definition of Key Terms	9
2. LITERATURE REVIEW	12
2.1 Writing as a Language Skill.....	12
2.2. Approaches to Writing	15
2.2.1. Product approach to L2 writing.....	15
2.2.2.Process approach to L2 writing.....	17
2.2.3. Genre approach to L2 writing.....	19
2.3. Assessment and L2 Writing.....	20
2.3.1. Assessment of learning (AoL).....	21

	<u>Page</u>
2.3.2. Assessment for learning (AfL).....	22
2.3.3. The role of feedback in L2 writing assessment.....	23
2.4. Evaluation	26
2.5. Assessment and Evaluation of Writing in EFL Context.....	27
2.6. Automated Writing Evaluation in L2 Writing.....	29
2.6.1. Automated scoring and automated feedback through AWE tools	30
2.6.2. The use of AWE in L2 teaching and learning.....	32
2.7. Prior Research on Agreement between Human Raters and AWE Scores in L2 Writing.....	35
2.8. Prior Research on the Effectiveness of AWE on Improving L2 Writing.....	38
2.9. Prior Research on Students' and Teachers' Perceptions of the Use of AWE in L2 Writing.....	44
3. METHODOLOGY	49
3.1. Research Design	49
3.2. Setting and Participants	50
3.2.1. Setting	50
3.2.2. Participants	51
3.3. Data Collection Instruments	53
3.3.1. The AWE tool	53
3.3.2. Students' essays	57
3.3.3. Weekly reflection reports.....	58
3.3.4. Semi-structured interviews with students	59
3.3.5. Semi-structured interviews with teachers	60
3.4. Data Collection Procedures	60
3.5. Data Analysis	64
4. RESULTS	67

	<u>Page</u>
4.1. Overview of the Study.....	67
4.2. RQ1: The Correlation between the Holistic Scores Given for EFL Students' Essays by the Instructors, and the Correlation between the Instructors Scores and the AWE Scores	68
4.3. RQ2: The Change in EFL Students' Writing Accuracy After the Provision of AWE Formative Feedback and Teachers' Formative Feedback	73
4.3.1. The change in students' writing accuracy from the first drafts to the second drafts of the first essay	75
4.3.2. The change in students' writing accuracy from the first drafts to the second drafts of the second essay.....	79
4.3.3. The change in students' writing accuracy from the first drafts to second drafts of the third essay	81
4.4. RQ3: EFL Students' Views on the Use of AWE in Learning and Assessment of Writing in L2.....	86
4.4.1. Advantageous features of the AWE system	86
4.4.2. The effects of AWE feedback on improving writing performance.....	89
4.4.3. Positive effects of the AWE tool on gaining autonomy	92
4.4.4. Comparison of the AWE feedback and teachers' feedback	96
4.4.5. Drawbacks of using the AWE tool.....	98
4.5. RQ4: EFL Instructors' Views on the Use of AWE in Teaching and Assessment of Writing in L2.....	101
4.5.1 The challenges faced by teachers in teaching writing	102
4.5.2. Advantages of using the AWE tool	104
4.5.3. Disadvantages of the AWE tool.....	108
4.5.4. Comparison of AWE assessment and teachers' assessment of writing in L2.....	110
4.5.5. Implementing the AWE tool in EFL writing	112
5.DISCUSSION	115

	<u>Page</u>
5.1. The Correlation between the Holistic Scores Given for EFL Students' Essays by the Instructors, and the Correlation between the Instructors Scores and the AWE Scores.....	115
5.2. The Change in EFL Students' Writing Accuracy after the Provision of AWE Formative Feedback and Teachers' Formative Feedback.....	119
5.3. EFL Students' Views on the Use of AWE in Learning and Assessment of Writing in L2	122
5.4. EFL Instructors' Views on the Use of AWE in Teaching and Assessment of Writing in L2	130
6. CONCLUSION	135
6.1. Summary of the Findings	135
6.2. Conclusion.....	139
6.3. Implications	142
6.4. Suggestions for Further Research	146
REFERENCES.....	147
APPENDICES	
CURRICULUM VITAE	

LIST OF TABLES

	<u>Page</u>
Table 2.1. A summary of research on correlation between human rater scores and AWE scores.....	37
Table 3.1. Demographic information of the instructors	52
Table 3.2. Research questions and data collection instruments	53
Table 3.3. Scoring guides of Criterion and Instructors for the highest scores	56
Table 3.4. Procedures of data collection	63
Table 4.1. The correlation among the scores of instructors, the AWE system, and the external rater.....	69
Table 4.2. The correlations among AWE scores, external rater's scores and each instructor's scores	70
Table 4.3. Frequent error types in students' first drafts	74
Table 4.4. The first essay- comparison of pre-feedback error counts between groups.....	75
Table 4.5. Comparison of pre-feedback and post- feedback error counts of the students' first essays	76
Table 4.6. Comparison of post- feedback error counts of students' first essays between groups	77
Table 4.7. Sample feedback provided by teachers and the AWE system	78
Table 4.8. The second essay- comparison of pre-feedback error counts between groups	79
Table 4.9. Comparison of pre-feedback and post- feedback error counts of the students' second essays	80
Table 4.10. Comparison of post- feedback error counts of students' second	

	<u>Page</u>
essays between groups	81
Table 4.11. The third essay- comparison of pre-feedback error counts between groups	82
Table 4.12. Comparison of pre-feedback and post- feedback error counts of students' third essays	82
Table 4.13. Comparison of post- feedback error counts of students' third essays between groups	83
Table 4.14. The error categories and the error types students stated that they were able to correct easily	85
Table 4.15. The main categories related to students' views on the use of AWE in L2 writing	86
Table 4.16. The subcategories related to the advantageous features of the AWE system	87
Table 4.17. The subcategories related to the effect of AWE on improving writing performance	90
Table 4.18. The subcategories related to the positive effects of the AWE tool on gaining autonomy	92
Table 4.19. The subcategories related to the comparison of AWE feedback and teachers' feedback	96
Table 4.20. The subcategories related to the drawbacks of using the AWE tool	99
Table 4.21. The main categories emerging from the instructors' statements and viewpoints on the use of AWE in teaching and learning writing	102
Table 4.22. The subcategories related to the challenges faced by teachers in teaching writing	103
Table 4.23. The subcategories related to the advantages of using the AWE tool	105

	<u>Page</u>
Table 4.24. The subcategories related to disadvantages of the AWE tool	109
Table 4.25. The subcategories related to comparison of AWE assessment and teachers' assessment	111
Table 4.26. The subcategories related to implementing AWE in EFL writing	112
Table 6.1. The implications of the study	141
Table 6.2. The implications of the study	145

LIST OF FIGURES

	<u>Page</u>
Figure 3.1. The research design employed in the study	49
Figure 3.2. Sample feedback given on the AWE system	54
Figure 4.1. A sample score and explanations provided by the AWE tool for the certain score	72
Figure 4.2. A sample view of the revision box of the AWE tool	87
Figure 4.3. Trait feedback analysis provided by the AWE system.....	93

LIST OF ABBREVIATIONS

AWE	:Automated Writing Evaluation
CALL	:Computer Assisted Language Learning
EFL	:English as a Foreign Language
ELT	:English Language Teaching
EPP	:English Preparatory Program
ESL	:English as a Second Language
L2	:Second Language
NES	:Native English Speakers
NLP	:Natural Language Processing

1. INTRODUCTION

1.1. Background to the Study

The ability to write effectively has a profound impact on our daily lives on account of various reasons. To begin with, writing skills are seen as a fundamental way of disseminating ideas or information to an audience either inside or outside a classroom. Therefore, writing to communicate is one of the most essential skills to be attained in any social context. Furthermore, writing leads to intellectual and cognitive growth owing to its connection with thinking processes (Greenfield, 1972). Moreover, the ability to write is a vital skill for academic success since content knowledge is generally measured through writing, especially in higher education, and learners are constantly engaged in academic writing (Cai, 2013; French, 2020; Vyncke, 2012). In addition to the importance of writing for academic success, it is also widely acknowledged that writing plays an important role in our educational lives and beyond as being able to write effectively is of paramount importance in the context of one's future career (Bacha & Bahous, 2008).

However, since writing is regarded as the most complex skill among other language skills (Lerner, 1996; Lombana, 2002; Tangpermpoon, 2008), mastering writing skills in English as a foreign language (EFL) context is much more complicated and challenging compared to L1 writing (Silva, 1993). Some of the challenges that learners face within the process of writing are, language interference, writing anxiety, lack of motivation, limited knowledge of English grammar and mechanics (Ata & Tıkız, 2019; Chen, 2006; Yalçın, 2010). To illustrate, according to Ata and Tıkız (2019), the majority of the preparatory class students in Turkey struggle with language and mechanics due to L1 interference. Hence, a huge part of the feedback provided by instructors consists of corrective feedback regarding mechanics and language use. Therefore, they suggest that more attention should be paid to improving language use and mechanics aspects of writing in order to convey the message more clearly. In addition, Frodesen and Holten (2003) state that grammatical errors hinder the meaning in students' compositions in L2. Similarly, a study conducted by Srichanyachon (2014) concluded that grammar poses the greatest challenge for EFL learners to produce proficient compositions. Moreover, some other problems that EFL students encounter while writing in L2 are lack of vocabulary knowledge (Zheng & Yu, 2018) and difficulty in the organization. Briefly, all these surface-level errors (e.g., mechanics, grammar, etc.) prevent L2 writers express their

ideas effectively, thereby influencing writers' macro-level language use (e.g., cohesion, organization, etc.) negatively. As a consequence, different approaches to writing have been introduced to EFL writing context in order to help learners improve their writing in L2.

Given that writing skills promote learning and these skills serve as a gateway to academic success and wide career opportunities, both learning-to-write and writing-to-learn are vital elements of the language learning process in the EFL context. As an EFL student is expected to produce well-organized, coherent, and cohesive written texts, the approaches that are employed in classroom writing instruction that guides students to improve their writing skills have a great influence on EFL writers' performance, and the ways of implementing effective writing pedagogy lie at the heart of the research. The studies investigating the effectiveness of different approaches to writing, such as product, process, and genre approaches indicate that although each approach has its own limitations and advantages, the process approach to writing is one of the most effective approaches to writing (Pritchard & Honeycutt, 2007). Since this approach focuses on stages of writing which contribute to the improvement of language use, it is assumed to raise consciousness (Deqi, 2005; Hyland, 2003), promote learning (Hedge, 2005), and advocate learner-centredness (Jordan, 1997). Therefore, the process writing approach is employed in the majority of the EFL contexts, especially in higher education settings (Jefferson & Radhakrishnan, 2020; Zheng & Dai, 2012), and students in EFL contexts engage in numerous stages and cycles of writing for the purpose of improving their writing skills through writing practice and acquire native-like fluency in writing.

In EFL contexts, the most common way to handle the issues that students struggle with is through integrating formative corrective feedback in formative assessment which is embedded in the revision stage of the process writing approach. In contrast to rating and scoring writing performance as in summative assessment, formative assessment of writing allows learners to receive feedback which helps them improve their initial work. Therefore, teachers are suggested to provide continuous and elaborated feedback regarding writers' errors (Leki, Cumming, & Silva, 2010). Even though providing immediate and personalized feedback contributes to learners' improvement, it is a labor-intensive and time-consuming task for instructors. Therefore, while some instructors prefer to provide focused feedback on writing regarding a certain and limited number of errors, some instructors prefer to provide feedback for all the errors along with positive

and negative comments. Nonetheless, teachers are suggested to consider the type of feedback that students prefer to receive since learner's preference is an important affective factor that impacts learners' motivation directly (Sprague, 2017). As for students' preferences, studies show that students prefer all their errors to be marked or highlighted immediately. (Lee, 2008; Ryan, 2012). Additionally, studies suggest that students with higher-level proficiency can benefit from delayed feedback more compared to students with lower-level proficiency. (Evans, 2013; Mason & Bruning, 2001). Therefore, the fact that EFL students prefer immediate feedback is quite reasonable as they find writing skills exceptionally challenging due to their lack of English language proficiency. As a result, although providing delayed and focused feedback can ease the burden of teachers, the effectiveness of such feedback is debatable (Dihoff, Brosvic & Epstein, 2003; Dihoff, Brosvic, Epstein & Cook, 2004). Moreover, students' preferences are other significant factors that affect institutions' and teachers' decisions concerning the type of feedback to be provided for writing assessment.

As teachers' and students' preferences vary, much research has been conducted to determine the most effective and fulfilling feedback and assessment types for decades. Additionally, writing pedagogy experiences a shift from teacher-centered classroom practices to autonomous, learner-centered, and flexible practices since another important factor that contributes to EFL learners' writing ability is students' own efforts and endeavors. In this respect, studies show that what separates good writers from others is their ability to involve in planning and reviewing stages actively (Brown, 2007; Raimes, 1985). Given that language teaching pedagogy moves towards a combination of conventional and digital technology pedagogy, integrating writing strategies and activities for self-regulated learning is more important now than ever. However, without any guidance and direct or indirect correction, students find self-assessment frustrating and time-consuming (Nordrum, Evans & Gustafsson, 2013). Consequently, the use of computer technologies and web-based programs which can provide feedback to foster and facilitate students' learning has been in the center of interest recently.

One of the ways of mitigating the challenges encountered by instructors and students is to integrate Automated Writing Evaluation (AWE) tools into the writing instruction and assessment process. Although the initial purpose of AWE programs was evaluating written prose and providing summative assessment, recent advancements in AWE technology allow teachers to utilize AWE to facilitate process-oriented writing

instruction and to enhance formative assessment. Thanks to the immediate and diagnostic feedback that AWE programs provide, AWE programs are used by millions of students, especially in English-speaking countries (Grimes, 2008; Grimes & Warschauer, 2010; Zhang; 2020; Zhang & Zhang, 2018). Even though AWE programs are dominantly used in countries where English is the native language, some recent studies have endeavored to explore the potential of AWE tools for EFL writing context. Nevertheless, only a few studies have examined the potential of AWE for the Turkish EFL context. Thus, this study aims at determining if there is a plausible correlation between automated scores and instructors' scores in assessing writing skills, investigating the effectiveness of the use of AWE feedback, and finally, exploring teachers' and students' views regarding the use of AWE for formative assessment in EFL writing.

1.2. Statement of the Problem

Writing is a crucial language skill that is not only difficult to teach and learn (Graham & Harris, 2016; Khoii & Doroudian, 2013; Persky, Daane, & Jin, 2003) but also to evaluate (Çakır, 2010; Fleissner, 2006). Therefore, much research has been conducted in order to determine how this particular language skill can be improved (Graham, MacArthur, & Fitzgerald, 2013; Graham & Perin, 2007; Kellogg & Whiteford, 2012), how students can be provided with effective feedback (Biber, Nekrasova & Horn, 2011; Hattie & Timperley, 2007; Hyland, 2015; Jackel, Pearce, Radloff & Edwards, 2017), and the ways of providing an accurate evaluation for students' written products (Haswell, 2005; Neff-Lippman, 2012; Warschauer & Ware, 2006; White, 1994). In addition, studies have pointed out that process writing can be much more effective if students are provided with feedback regularly during the process (Parr & Timperley, 2010). However, the quality of feedback is just as important as receiving feedback. According to some researchers, (Allen, Jacovina & McNamara, 2016; Shermis & Burstein, 2013; Shute, 2008), the best way to give feedback is by providing individual and immediate feedback. Even though individual and immediate feedback are regarded as the ideal ways to give feedback, providing these types of feedback is time-consuming, and it requires knowledge and expertise (Lee, 2014; Marshall & Drummond, 2006; Wilson & Czik, 2016). Furthermore, the requirement of giving feedback progressively gets more difficult as the number of students increases. As a result, writing in a foreign language can be

frustrating for students, and also providing feedback and evaluating writing can be challenging for teachers.

With the advancements in technology, AWE, which has been used since the 1960s, and had been developed remarkably in the 1990s, provides an opportunity to evaluate students' written products based on both process and product-based approaches. Furthermore, these systems reduce teachers' burden by providing feedback to students' written products almost immediately (Foltz, Streeter, Lochbaum, & Landauer, 2013; Palermo & Thomson, 2018). Web-based AWE systems have the ability to evaluate quite a wide range of writing styles and genres by providing feedback on style requirements. Along with giving scores to written products based on their overall quality, they also provide corrective feedback on grammar, vocabulary, spelling, punctuation, and other common error types. (Burstein & Chodorow, 1999). These systems, which take into consideration almost every criterion that is used by human raters during the assessment process, provide instant and overall feedback to each written product of each student thanks to its algorithms. By doing so, the systems aim to increase the efficiency of feedback (Burstein, Chodorow & Leacock, 2004).

When it comes to assessing writing in a foreign language, the validity and reliability of the evaluation are just as significant as the efficiency of the evaluation (Drid, 2018). There are many studies on the reliability of the evaluations made by AWE systems, and many of these studies have reached positive results (Attali, 2004; Page, 2003; Shermis & Burstein, 2003). Moreover, in their study, Attali and Burstein (2005) found that the scores given by the AWE system were more reliable than the scores of human raters and that the scores obtained when the human-rater and AWE systems' evaluation were used together were close to perfect. These studies indicate that including AWE systems in evaluation increases accuracy, which is a prominent factor that needs to be taken into consideration during the evaluation process. Despite the positive results of the studies, the number of experimental studies conducted in the EFL context regarding this subject is quite limited, and it is frequently emphasized that the score and feedback consistency of the people and AWE systems should be investigated in the field of English language teaching and learning (Li, Link & Hegelheimer, 2015; Long, 2013; Mohsen & Abdulaziz, 2019; Zupanc & Bosnic, 2015). Furthermore, to the researcher's best knowledge, only one study has examined the correlation between AWE scores and instructors' scores in the Turkish EFL context (Doğan, 2013).

Even though many studies have been conducted to investigate the effectiveness of using AWE systems for evaluating written parts of international exams in laboratory settings, the use of AWE in the classroom environment has not been widely investigated (Balfour, 2013; Burstein, et al., 2004). As a result, more research that investigates the use of these systems in the classroom is needed. In addition to this, although many studies have focused on the result analysis and scores of automated writing skills assessment systems, (Attali & Burstein, 2005; Burstein & Chodorow, 1999; Enright & Quinlan, 2010; Shermis, Garvan & Diao, 2008) there is not enough research done in order to investigate the effectiveness of the use of feedback provided by these systems during process-oriented writing assessment (Chen & Cheng, 2008), and the results of existing studies are inconclusive (Li, et al., 2015). Moreover, most of the existing classroom research on AWE was conducted in English-speaking countries (Balfour, 2013; Li et al., 2015; Roscoe, Wilson, Johnson & Mayra, 2017) and to the researchers' best knowledge, comparative studies to date have not explored the effectiveness of AWE on reducing writing errors during revision stage that takes place inside the classroom in Turkish EFL context. Therefore, further studies that focus on using AWE in EFL classroom settings and improvement of students' writing skills are needed (Burstein, et al., 2004; Mohsen & Abdulaziz, 2019).

Another important factor that should be taken into consideration before AWE is implemented in classroom settings is students' and teachers' perception of the AWE programs as possible drawbacks or shortcomings might prevent teachers and institutions from employing AWE programs. Additionally, the results of the studies that explore students' and teachers' views are far from generalizable since there is a gap in the literature, especially concerning attitudes toward AWE (Roscoe et al., 2017; Scharber, Dexter & Riedel, 2008). Initial studies showed positive views towards AWE (El Ebyary & Windeatt, 2010; Khoii & Dorouidan, 2013), and revealed that AWE increases students motivation (Link, Dursun, Karakaya & Hegelheimer, 2014), reduces the time teachers allocate to assessment substantially (Wilson & Czik, 2016), and facilitates writing instruction (Warschauer, 2010; Warschauer & Grimes, 2008). Nonetheless, there is not a previous study conducted to explore teachers' views towards the use of AWE for the process-oriented writing assessment in the Turkish EFL context, and there are only a few studies exploring students' views (Bulut, 2019; Gençer, 2019). Hence, this study aims at finding out Turkish EFL teachers' and students' perceptions of AWE and their reflections

on implementing AWE in the classroom for formative assessment and evaluation purposes.

1.3. Purpose of the Study

The overall goal of this study was to develop an understanding of the role AWE might play in the Turkish EFL context in terms of formative assessment of writing. With this aim in mind, the following four research questions were asked:

- 1) Is there a relationship between AWE holistic scores and instructors' holistic scores given to students' essays?
- 2) Is there a significant difference in terms of writing accuracy between a group of students receiving AWE formative feedback with a group of students receiving teacher feedback?
- 3) How do EFL students view the use of AWE in L2 writing?
- 4) How do EFL instructors view the use of AWE in L2 writing?

1.4. Significance of the Study

This study is significant in terms of its contribution to the existing body of research knowledge by offering empirical evidence in several ways. To begin with, previous studies (Attali & Burstein, 2006; Balfour, 2013; Powers, Burstein, Chodorow, Fowles, & Kukich, 2002; Ramineni, Trapani, Williamson, Davey & Bridgeman, 2012) have examined the correlation between the scores given by expert raters and AWE programs in the high-stake testing environment. However, very little attention has been paid to exploring how well AWE scores correlate with instructors' scores in EFL contexts where the process approach to writing is employed (Stevenson, 2016). Therefore, the present study contributes to the research on the correlation between AWE scores and human scores given in the EFL classroom context.

This study proposes to investigate not only similarity between the scores obtained from an AWE system and instructors but also the impact of its feedback on students' writing skills by exploring if there is a significant difference in terms of writing accuracy of the experimental group compared to the control group after the provision of AWE formative feedback. Due to the lack of prior research on system's feedback and the impact of this feedback on accuracy in process-oriented writing assessment, little is known about the role that AWE formative corrective feedback might play in process writing

assessment, However, this study can contribute to the existing literature on the effectiveness of AWE in improving writing accuracy by reducing errors. Additionally, by examining the change in accuracy on drafts, this study can provide insight into the types of errors that AWE is more likely to correct in the EFL context.

Another significant use of the findings of this research might be to inform pedagogical decisions made by teachers and institutions as the study aims to reveal teachers' reflection on the use of AWE for classroom assessment and scoring. The existing studies that explored the effects of using AWE in a classroom setting failed to reveal both teachers' and students' views and attitudes towards AWE. However, this study proposes to provide information about the suitability of these systems in the environments where English is taught as a foreign language by taking the opinions of teachers and the students into consideration. Moreover, students' views on the use of AWE formative corrective feedback might help the teachers and students to estimate the potential of the use of AWE for autonomous and self-regulated learning in higher education contexts.

Given the impact of the current pandemic on educational systems worldwide, the study might inform teachers', students', and administrators' decisions pertaining to digital teaching and learning. As most, if not all, of the teaching, learning, assessment, and evaluation practices take place outside the classroom nowadays, students need to be more autonomous and they need to be encouraged to be more independent of teachers and classroom assessment to improve their learning. Moreover, due to the recent emergency remote teaching situation, pen-and-paper testing and traditional assessment are no longer administered in many countries. Hence, school administrators and teachers need to integrate computer-based assessment, web-based assessment, and innovative technologies to assess students and provide feedback. According to Limpo, Nunes, and Coelho (2020), although technological advancements have been utilized for teaching and learning writing, the significance of these tools are better understood due to the pandemic, thereby making the studies conducted on the use of innovative tools for improving writing even more necessary. Moreover, for the new generation of students who are also labeled as "digital natives" (Prensky, 2001), the integration of innovative tools into education can promote learning and teaching (Gilakjani, 2017; Hsu, 2016) in the age of information and technology. As an innovative tool, AWE might meet the needs of the 21st century if the implementation of AWE is supported by teachers' attempts to improve the writing

curriculum (Wilson & Roscoe, 2020). Hence, studies examining the efficacy of AWE and views of teachers and learners may provide a better understanding of the potential of using AWE in EFL writing. All in all, the present study provides insights into the pedagogical potential of AWE for EFL writing instruction, formative assessment, and evaluation.

1.5. Limitations of the Study

The empirical results reported in this thesis must be interpreted in light of a number of limitations. Initially, the scope of this study is limited to a private university in Turkey. Therefore, caution needs to be taken in considering the generalizability of the results to other contexts and educational settings. Secondly, the time frame of this study is restricted to a module which is eight weeks for students, and additional four weeks for the instructors, thereby consisting of a total of 12 weeks. Hence, a longitudinal study might be designed to explore the issues concerning the implementation of AWE in depth. Moreover, this study was conducted with a limited number of students (n=38) and instructors (n=7) due to the availability of the participants in the research context. Hence, the findings reflect the scores and ideas of a small scale of participants. Additionally, the data were collected from students who were at the intermediate level of English proficiency. Therefore, future studies might need to include other grade levels in research to generalize the findings.

1.6. Definition of Key Terms

Analytic scoring: It refers to scoring each aspect of a task separately.

Assessment: It refers to the process of gathering information to gauge comprehension, performance, and progress.

Autonomous learning: It is generally defined as the learning progress in which learners take control of their own learning by regulating and evaluating their learning independently.

Automated feedback: It refers to the type of instant feedback generated by the systems that use computer technologies.

Automated scoring: It can be defined as the act of producing scores automatically with the help of scoring algorithms

Automated Writing Evaluation (AWE): It refers to the use of software that provides instant individualized feedback and computed-generated scores for written products

Computed Assisted Language Learning (CALL): It refers to the use of computers in second or foreign language teaching and learning

English as a Foreign Language (EFL): It refers to situations in which English is learned as a foreign language

English as a Second Language (ESL): It refers to teaching English to non-native speakers of English living in English- speaking countries

English Language Teaching (ELT): It refers to the profession of teaching English and the industry related to the activity of teaching English.

Evaluation: It is generally defined as activities that are carried out to determine the quality, effectiveness, or success of educational procedures or making a judgement of one's performance.

Formative assessment: It refers to the type of assessment that is provided for the purpose of identifying the gap between current level success and desired outcomes in order to bridge the gap and promote learning.

Genre approach: It can be defined as the type of approach that is intended to raise learners' awareness of the social functions of different types of texts.

Holistic scoring: It is generally defined as the act of evaluating the overall quality based on the general impression of the performance.

L2: It refers to the second language

Metacognitive awareness: It refers to individual's comprehension of their own learning process and regulating their learning by employing necessary strategies.

Process approach: It refers to a type of approach to writing that focuses on the stages that writers go through during the writing process.

Product approach: It refers to a type of approach to writing that focuses on the quality of the final product

Self-regulated learning: It refers to the process in which learners plan, monitor, and evaluate their own learning.

Summative assessment: It is generally defined as the act of assessing students' performance based on a final outcome.

Zone of Proximal Development (ZPD): It refers to the level that students can reach thanks to guidance and scaffolding.

2. LITERATURE REVIEW

2.1. Writing as a Language Skill

According to Smyth (2004), the ability to write or compose requires thinking critically, analytically, and logically since students are to provide reasons or evidence in order to convey their message or express their views to others. Writing is much more than putting words on paper as writing is also the process of finding ideas, contemplating how to communicate ideas effectively, and setting down ideas on paper by organizing them (Linse, 2005).

According to Flower and Hayes's (1981) cognitive process model, writing involves three major elements which are task environment, long-term memory, and working memory. The *task environment* consists of elements, such as topic and audience. These elements determine the purpose of the composition. Thus, goals are set based on the information regarding these elements, and the next stages are planned accordingly. The second major component of writing is *long-term memory* as existing knowledge related to the topic and reader's expectations is essential for writing. Hence, writers need to reach a certain level of familiarity with the topic and they need to know how a certain type of written piece is constructed. In addition to long-term memory, *working memory* is responsible for a variety of cognitive processes that are significant for writing. These processes are planning, translating, and revision. While *planning* refers to deciding on the format and outline of the text and processing the cumulated information to decide what parts of the existing information will be transferred to writing, *the translation process* refers to translating plans and synthesized information into meaningful and grammatically correct sentences. Finally, *the review process* consists of two main cognitive processes which are *evaluating* and *revising*. These processes do not necessarily take place at the end of writing activity, but they occur consciously at any stage of writing. Therefore, they can lead to additional cycles of planning and translating. Although it is suggested by researchers (Ellis & Yuan, 2004; Jones & Tetroe, 1987; Raimes, 1985) that both L1 and L2 writers go through the same writing processes, writing skills and strategies that L2 writers employ are not independent of the level of proficiency (Cumming, 1989).

Writing is a complex skill since it requires writers to go through a wide range of cognitive processes that represent both lower thinking skills, such as remembering and processing, and higher thinking skills such as creating and evaluating (Bloom, 1956).

Moreover, it requires writers to attain a certain level of linguistic knowledge and skills as it comprises grammar, mechanics, vocabulary knowledge and organization skills, and it requires information attained through listening and reading activities (Harmer, 2007). Using all these skills and going through numerous cognitive processes, writing is regarded as a difficult skill to master, especially for EFL learners (Ahangari, Hejazi & Razmjou, 2014; Bai, Wang & Nie, 2020; Raissah & Aziz, 2020; Senel, 2018; Ying, 2018) owing to *interlingual errors* which occur due to students' tendency to transfer language patterns and rules from their native language and *intralingual errors* which occur due to limited knowledge of target language itself (Richards, 1974). In other words, L2 writers are more disadvantaged than L1 writers due to L1 interference in L2 writing since they tend to translate their first language to the target language (Yağız, Yiğiter & Genç, 2009). Moreover, L2 writers' insufficient knowledge of the target language cause overgeneralization of rules, limited variety in word use and grammar structures, wrong use of grammar rules, spelling, punctuation errors, inability to convey the meaning effectively, and other errors which occur due to incomplete knowledge of L2 (Kaweera, 2013; Phuket & Othman, 2015). Additionally, students' inexperience in writing in either their first language or the target language might have an effect on EFL writers' performance (Leki, 2001). Given that more experienced writers are more likely to adopt cognitive strategies and internalize knowledge (O'Brien-Moran & Soiferman, 2010), more attention needs to be paid to teaching L2 writing as a language skill and providing opportunities for L2 writers to practice (Allen, Crossley, Snow & McNamara, 2014). Nonetheless, high importance of writing skills for academic success and the importance of implementing writing in language learning programs used to be less acknowledged, especially in EFL context (Buitrigo & Diaz, 2018).

Unfortunately, in foreign language education, the main purpose of writing skills used to be seen as transcribing speech, and other language skills, especially speaking, were privileged over writing by structuralist linguists (Harklau, 2002). According to Harklau (2002), the main reason why writing skills had been neglected for so long is that it was seen as a supplementary skill to language acquisition, and writing was viewed as the last skill to be acquired by second language learners. As a result of the lack of attention and research, the nature of writing in L2 has not been explored sufficiently, whereas the number of L1 acquisition studies has grown exponentially (Lally, 2000). Additionally, Warschauer and Ware (2006) argue that the needs that countries were bound to fulfill

before the Second World War required them to be in favor of an approach that prioritizes other skills to be mastered. Hence, many other researchers (Leki, 2003; Nordin, 2017; Rivers, 1968; Sağlamel & Kayaoğlu, 2015) holds the proponents of the Audiolingualism Approach accountable for this negligence since advocates of the Audiolingualism Approach prioritized speaking skill over the other skills and treated writing as a supplementary tool to improve other skills. Nonetheless, Warschauer and Ware (2006) point out that paradigm shifts in education occur due to cultural, economic, political issues. Therefore, as the social, cultural, economic contexts changed after the Second World War, the role of writing in foreign language teaching has changed correspondingly.

Olson (1996) emphasizes the correlation between writing and social, economic, and global conditions, and he points out that when the level of literacy contributes to industrial, economic and social development, the art of writing and perception of writing are shaped by these improvements and it changes the societies' experiences as well. To illustrate, after the Second World War, the ability to write well started to be associated with power (Derrida, 1979; Halliday & Martin, 2003; Martin, 1994; Winsor, 2012), since writing in English particularly has been the medium of communication (Hamp-Lyons, 2002) and a source for cultural development and cognitive development (Kellogg, 2008; Schultz, 1991; Webster & Ammon, 1994). As a result of the trend of globalization and modernization, novel requirements have arisen, and curricula of academic fields and majors have been restructured in order to meet those needs. As writing in ESL and EFL contexts has become a vital component of restructured programs, more research and attention have been devoted to the ability to write in English and implementing new approaches to the classroom (Schultz, 2011).

Following the emergence of global efforts to implement writing instruction in language education, there has been a growing interest in the EFL writing process as studies have shown that implementing writing instruction in EFL settings fosters learning a foreign language in general. Even though the proponents of the input hypothesis claim that the only way to learn a language is the exposure to input, many others argue that receiving input is not sufficient to improve foreign language knowledge (Færch & Kasper, 1986; Pica, 1994; White, 1987). Swain (1985) suggested that input alone doesn't ensure that learners can reach the desired degree of accuracy and proficiency in L2. Swain's output hypothesis (1985) asserts that output triggers learners to pay attention to their use of language, and it results in noticing the gap between what is produced and

what should be produced. Cumming (1990) pointed out the connection between noticing and learning outcomes of writing by suggesting that cognitive processes such as choosing appropriate words, noticing the gap, and assessing the language result in language development. Similarly, when Izumi (2002) conducted a study that compared the effects of input and output on learning outcomes, he concluded while there was a lack of evidence supporting the positive effect of visual input, written output enhanced learning by leading to a deeper processing of target structure and more deep-rooted memory traces. According to Manchon and Larios (2008), writing is a complex and mentally challenging task that would require learners to set ongoing goals until the writing is finished. Therefore, producing a text is accepted as problem-solving and the achievement would enable learners to become more competent in the target language and learning more about language use.

As a result of the positive findings showing that implementing writing in EFL settings can promote language learning, writing has become a crucial part of language education all around the world (Manchon & Larios, 2008). In the light of these studies and developments, the importance of writing has recently been widely known, and contributions of writing to learning a language have been widely acknowledged by researchers (Cumming, 1990; Manchon, 2011; Manchon & Larios, 2008; Swain & Lapkin, 1995). Correspondingly, different approaches have been proposed and used for the purpose of helping students become more competent L2 writers and overcoming the challenges of this complex language skill.

2.2. Approaches to Writing

Three main approaches to writing in L2, namely, the product approach, the process approach, and the genre approach will be discussed separately in this section.

2.2.1. Product approach to L2 writing

The product approach is also known as the traditional approach (Gabrielatos, 2002; Matsuda, 2003; Tom, Morni, Metom & Joe, 2013) since the product approach has dominated the teaching of writing in classes for two centuries (Farris, 1987). The reason why this approach is called “product” is that it is concerned with the finished written product, and it mainly focuses on the result of teaching stages rather than the stages themselves (Kroll, 1990). Therefore, teachers are to only pay attention to the correctness of the completed written product in order to assess learners’ performance (Bazerman,

1988; Silva, 1990). To be considered successful in writing, learners are expected to produce error-free sentences in the target language (Bae, 2011; Mehr, 2017; Nunan, 1999). As a result, more attention is paid to sentence-level structure, whereas discourse and organization aspects of the texts are usually ignored (Tuffs, 1993). Accordingly, the goal of the curriculum must be teaching students grammar rules and syntax as linguistic knowledge is prioritized (Pincas, 1982; Silva, 1990). Moreover, students are provided with a sample text taken from a textbook or constructed by a teacher (Miri, 2014) as input including grammar rules and patterns that ought to be used when they write a new text, and students imitate sentences by using the model text (Rusinovci, 2015).

According to Badger and White (2000), this approach consists of four stages. The first stage is called familiarization, and the purpose of this stage is to provide a model for writing so that the sample leads to awareness of certain features of the text, such as the genre and grammar. In the following controlled writing stage, students imitate certain features of the model text to practice. The guided writing stage that comes next allows students to work on ordering and connecting ideas. However, the ideas are not as valued as the organization. Finally, in the free writing stage, students complete their writing task on their own by employing similar features of the model text.

The product approach is one of the most common ways to teach writing in EFL contexts, especially in traditional EFL classrooms. In EFL writing classrooms adopting process approach, teachers concentrate mainly on accuracy regarding, grammar, mechanics, punctuation and spelling (Alshakhi, 2019), and students are expected to produce texts meeting the language goals. In addition, teachers tend to mark the errors on students' final products, which are also their single draft, judge students' achievement in terms of accuracy, and give them back to students. Hence, corrective feedback focusing on grammatical accuracy is provided by the teachers. Nevertheless, students are not expected to make changes and resubmit their papers. Instead, they move on to a new writing task after their previous written product is evaluated.

Proponents of this approach suggest that students become much more aware of language forms (Tangpermpoon, 2008), they acquire linguistic knowledge which is internalized by imitation (Badger & White, 2000), and the model texts increase proficiency which leads to a decrease in errors (Tom et al., 2013). On the other hand, critics argue that students do not become much aware of their mistakes since they do not work further on their paper (Tom, et al., 2013), and it does not support students' creativity

since the ideas are not as valued as the linguistic knowledge (Nunan, 1989). Furthermore, it is believed that writing is a non-linear process (Emig, 1971; Flower & Hayes, 1981), whereas the product approach follows linear stages (Zhou, 2015). Nonetheless, the product approach is still used in EFL contexts and it is considered valuable to train students for high-stake exams (Hu & Wang, 2019).

2.2.2. Process approach to L2 writing

It is believed that the process approach emerged as a reaction to drawbacks and criticism of the product approach (Badger & White, 2000; Bijami & Raftari, 2013; Ezza, 2010; Matsuda, 2003; Murray & Hourigan, 2008; Zhou, 2015). When L1 researcher Emig (1971) examined the successful writers and their writing habits, she found out that these writers do not compose in a linear sequence. Since the nature of writing is recursive (Emig, 1971; Hyland, 2003; Kroll, 1990), expecting language learners to produce error-free compositions at their initial attempts is considered unrealistic given the fact that cognitive processes along with revision and feedback are seen as vital for improving writing skills (Brown, 2007; Susser, 1994). Emig's pioneering study in 1971 led to a radical change in L1 writing pedagogy. Subsequently, the adoption of the process approach to writing in EFL and ESL contexts was given impetus (Demirel, 2011; Kamimura, 2000; Matsuda, 2003; Miri, 2014). Zamel (1983) argued that L1 and L2 writers follow similar patterns in writing; therefore, language learners can benefit from a student-centered approach that values the cognitive aspect of writing. Consequently, the process approach has gained ascendancy over traditional approaches in many EFL and ESL contexts since the 1980s.

In EFL writing contexts, by employing the process approach, more emphasis is placed on classroom activities rather than the outcome alone. According to Susser (1994), awareness and intervention are the key components of the process approach to L2 writing. Hence, the classroom activities are designed to promote learners' awareness and teachers intervene in the writing process in order to provide feedback or comment. The process approach comprises four main stages, namely pre-writing, composing, revising, and editing (Tribble, 1996). The classroom activities suggested for the pre-writing stage are brainstorming, choosing ideas, and planning. Composing stage is also known as the drafting stage which is learners' initial attempt to put ideas into words. For the revising stage, feedback is seen as crucial since it is considered a form of input (Nordin, 2017).

The feedback can be provided by teachers or peers. The feedback and comments that are obtained from readers focus on organization and ideas as well as language use. The editing stage takes place after students make use of feedback and make the necessary changes on form and content (Zhou, 2015). These stages may be repeated, or they may overlap depending on the writers' cognitive processes and the number of drafts that students are given chance to write. To put it in a nutshell, this approach views writing as learning since the notion behind the process approach is to guide students through various stages which can lead to a gradual improvement in content quality as well as language competency. Therefore, the role of a teacher is perceived as a facilitator while students are active participants, and the approach is essentially learner-centered.

Proponents of the process approach to writing seem to believe that this approach motivates students to communicate their message since it prioritizes writing fluently (White & Arndt, 1991; William & Kaplan, 1996). According to Peregoy and Boyle (2001), pre-writing activities, such as brainstorming and idea-sharing result in communicating the message through oral discussion that allows students to be exposed to input which promotes language learning. Furthermore, collaboration, idea sharing, and individual feedback motivate students to write, revise and rewrite their drafts (Graham & Harris, 2016; Graham & Sandmel, 2011). As a result of collaborative and independent stages of writing, this approach allows learners to take more responsibility and to be more autonomous (White & Arndt, 1991). In addition to pre-writing and while-writing stages, post writing stage is believed to be utterly beneficial for the reason that some elements, such as peer review, teacher feedback and individual revision raise students' awareness of their strengths and weaknesses (Patthey-Chavez, Matsumura & Valdes, 2004; Wang, 2014).

On the other hand, some critics argue that since students need to revise and rewrite their draft, the process approach requires a considerable amount of time to be allocated for the procedures (Jubran, 2017; White, 1988). Moreover, some critics contend that more attention needs to be paid to mechanics (Graham & Sandmel, 2011), grammar, and sentence structure (Reid, 2001), whereas previous research has indicated that instruction provided in writing classroom should focus on both content and mechanics in order to improve the overall quality of writing (Baker, Gersten & Graham, 2003). Despite the criticism, the process approach is widely acknowledged in L2 writing instruction, and it

is generally accepted as one of the most utilized approaches in the EFL context since the 1980s (Atkinson, 2003; Zhou, 2015).

2.2.3. Genre approach to L2 writing

The genre approach is the third primary approach to writing instruction (Raimes, 1991). Swales (1990) defines genres as communicative tools that are utilized to achieve communicative goals. Discourse communities that use a genre share the same communicative aims and these aims identify a genre (Swales, 1990). Therefore, the genre approach to writing views writing as a social practice that aids to achieve communicative purposes in a particular context. According to Hyland (2007), the genre approach gained popularity in the 1990s as a result of the notion that the genre approach enables learners to use L2 writing for a variety of purposes in real-life settings by allowing them to understand L2 environments. Hence, students may practice writing business letters, different types of essays, stories, or other types of texts that might be useful for students to produce in certain situations.

In L2 writing contexts, to be able to produce a text, students go through three main phases which are modeling, joint-construction of a text, and independent construction of a text (Hyland, 2003; Macken-Horarik, 2002). Prior to the modeling phase, a context-building stage takes place so that learners can discover the cultural and social aspects of the target genre. For this reason, brainstorming activities, discussions, listening, and reading tasks are utilized during this stage to familiarize students with the social context. Teachers introduce sample texts that represent the features of a certain genre in the modeling phase for the purpose of drawing students' attention to language use in the target genre. During the joint-construction stage, students are guided by teachers or their group members to construct a product that is similar to the model text. Finally, the independent construction of the text phase allows students to construct their texts individually without scaffolding. Students are expected to make use of the relevant vocabulary and grammatical structures that they were exposed to during the preceding stages in the independent construction phase. Although teachers may not concentrate on every aspect of writing in each draft, they provide feedback regarding the features of a certain genre.

Due to the fact that genre approach, like product approach, leads students to imitate model texts, it has been criticized for restricting learners' independence (Dirgeyasa,

2016). Thus, it is claimed that the genre approach limits learners' creativity, (Rahman, 2011), and learners are viewed as passive since the initial phases demand students to observe and imitate (Badger & White, 2000). However, adherents of this approach believe that students become less dependent on teachers when they acquire the ability to detect certain features of a genre and replicate the same features for their construction (Tuffs, 1993). Additionally, being able to replicate certain genres might be a useful skill for students' academic life and real-life purposes, such as obtaining social achievements (Bijami & Raftari, 2013). Regardless of the approach preferred in writing, assessment is also an integral part of any approach preferred in L2 learning environments.

2.3. Assessment and L2 Writing

In order to understand the nature of assessment in L2 writing, we should have a look at what *assessment* entails and the relationship between assessment and learning. According to Assessment Reform Group (2002), assessment is an ongoing, dynamic process as teachers observe students every day and they make decisions about how they can improve or facilitate learning. Therefore, anything that students and teachers do in classrooms that results in decision-making and reflection can be interpreted as assessment. Assessment is a crucial component of the teaching and learning process since it is an indicator of learners' strengths and weaknesses. The assessment provides insight so that teachers can decide how to shape their instruction. As well as teachers, learners benefit from assessment by acknowledging their weaknesses and focusing on the issues pointed out by assessment in order to improve themselves.

Despite the growing emphasis on the prominence of assessment, the effectiveness of assessment for learning was ignored before the '90s (Boud, 1995). Historically, the assessment had been a requirement that took place at the end of the teaching process and it was mainly utilized to measure the school's success, teachers' effectiveness, and placement of learners (Baker & O'Neil, 1994). However, the role of assessment has started to undergo changes thanks to teachers and researchers suggesting that establishing quality assessment approaches must be the groundwork for other educational changes (Cizek, 1997). Brown (2004) points out that the nature of assessment keeps changing as the education system changes, and if the existing systems fail to employ promising assessment reforms, the systems must change first. Among the changes that occurred in

assessment, the paradigm shift from *assessment of learning* (AoL) to *assessment for learning* (AfL) has been central in research.

2.3.1. Assessment of learning (AoL)

Lebler (2008) identifies AoL as the conventional role of assessment stating that it measures pupils' comprehension of what has been taught. Since AoL utilizes summative use of assessment, the terms *summative assessment*, which refers to making a judgment based on the recorded evidence pertaining to performance or achievement, (Taras, 2005), and *AoL* are usually used interchangeably (Crooks, 2004; Harlen, 2005; Johnson & Jenkins, 2009; Stiggins, Arter, Chappuis & Chappuis, 2004). It is considered that AoL serves as an indicator of changes in knowledge and skills (Kapoor & Natarajan, 2012). Hence, the main purpose of this assessment is to provide proof of accomplishment to students themselves, their parents, or institutions (Earl & Katz, 2006). In other words, the AoL focuses on providing evidence for students' comprehension and learning by using scores or grades as indicators, whereas the main aim of this type of assessment is not to improve students learning by assessing their performance (Gardner, 2012).

In AoL practice, teachers play a dominant role as they evaluate students' writing against a certain criterion which mainly focuses on accuracy and they provide summative feedback, or grades afterward, whereas students are only informed about their performances (Lee, 2017). Since the information gathered thanks to assessment serves as an indicator and plays a role in decision-making processes, the validity and reliability qualities of an assessment are of the utmost importance. Therefore, teachers have an additional responsibility which is to oversee accuracy and the fairness of the assessment given that they are usually the ones who gather evidence and report what has been learned (Gardner, 2012). On the other hand, students tend to be passive about their learning, and they are involved in the assessment stage as respondents. According to Lee and Coniam (2013), in examination-oriented EFL writing contexts, which are in abundance, teachers are more likely to teach for tests, and students are less likely to be given insight into how they can improve their writing skills. Given that AoL is implemented in many EFL contexts, implementing alternative assessments in EFL contexts is challenging. Nonetheless, putting too much emphasis on AoL and neglecting the value of assessment for formative purposes might lead students and teachers to perceive assessment as being graded rather than an activity that might promote teaching and learning (Lee & Coniam,

2013). Hence, it is significant for teachers to make effort to explore different forms of assessment that can increase students' motivation and contribute to their learning experience (Lee, 2010).

2.3.2. Assessment for learning (AfL)

Hargreaves (2005) interprets the definition of *AfL* as observing students' performance bearing the learning goals in mind, using the information gathered in order to make decisions concerning the following steps, providing feedback to boost students' performance, being informed about how pupils learn, letting pupils be in charge of their own learning process and perceiving assessment as an opportunity to learn from. It is believed that AfL can promote students' learning by making use of the classroom practices and activities that provide information about students' progress and their learning needs. The type of assessment that uses the information provided thanks to the assessment in order to design and adapt future learning activities is known as formative assessment (Black, Harrison & Lee, 2004). Thus, the terms *formative assessment*, which refers to providing feedback to bridge the gap between the current level of learning and aimed level of learning, and *AfL* are usually used interchangeably (Harlen, 2007; Lee, Mak & Yuan, 2019).

To clarify the distinction between AoL and AfL, Cizek, Andrade, and Bennett (2019) point out that the main differences between these two arise from teachers' and students' roles, the time when assessment takes place, and the purposes of assessment. Before the assessment takes place, teachers are expected to establish a plan to achieve educational goals, and scaffold students' writing. Subsequently, teachers who implement AfL in their classroom make use of assessment to provide oral or written formative feedback to students which would give students insight into their own strengths and weaknesses. The evidence gathered by the assessment does not only help learners to be aware of their own capabilities and shortcomings, it also helps teachers to improve their teaching by making necessary modifications considering the evidence (Earl & Katz, 2006). According to Lee (2017), students are as involved and active as teachers in writing classrooms where AfL principles are implemented. Students engage in classroom assessment by making use of feedback provided by teachers or peers, monitoring their own learning progress, and bridging the gap between their current level of learning and their goals. Thus, the quality, consistency, and timeliness of feedback are crucial for

students to move forward. In addition, AfL aims to involve students in self-assessment and self-reflection with the intention of raising students' self-awareness and creating an autonomous learning environment (Lee, 2017). To put it in a nutshell, AfL is not only a way that teachers assess students' performances, but it also contributes to the teaching and learning process where both students and teachers are responsible and productive (Glasson, 2009).

The last two decades have seen a considerable amount of studies that provide insight into the theory and practice of AfL. Nevertheless, most of the literature focused on primary and secondary schools, and there have been limited studies conducted in the higher education context at the beginning of the 2000s (White, 2009). The main reason why relatively few studies have been undertaken in higher education settings is that assessment practices were mainly under the influence of a behaviorist framework which assumes that assessment should take place at the end of a teaching process and the main purpose of teaching is to prepare students for high-stake exams. While behaviorist learning theory is still influential in practice, a shift towards implementing principles of constructivist theory in assessment occurred as the constructivist framework highlights the value of interaction, collaboration, and engaging students in learning stages (White, 2009). Hence, incorporating formative assessment in EFL writing holds the potential for improving EFL learners' writing performance (Huang, 2012).

2.3.3. The role of feedback in L2 writing assessment

Feedback refers to the information given to the learners to notify them of their performance, or their current level of learning. (Hattie & Timperley, 2007; Narciss, 2008; Ur, 2012), and it is considered to be one of the most influential factors affecting students' learning (Hattie & Timperley, 2007; Joughin, 2009) since if feedback is not provided, action or performance would be unproductive (Laurillard, 2013). Hence, according to Laurillard (2013), the teaching and learning process should be conducted in a goal-action-feedback and revise cycle. Similarly, Hattie and Timperley (2007) suggest that the ideal learning environment mainly consists of feed up, feed back, and feed-forward stages. While the first stage informs learners about the desired level of learning and initial goals, feed back stage informs learners about their current stage of learning, and finally, the feed-forward stage allows learners to progress towards pre-set goals. Moreover, incorporating feedback in classroom assessment is also seen as an implication of

Vygotsky's zone of proximal development (ZPD) theory (1978) since feedback serves as *scaffolding*, which refers to support or guidance provided by the teachers, and it is believed that utilizing formative feedback as scaffolding promotes cognitive growth. In other words, feedback helps learners to reach the potential level of development which is also the desired outcome of the feed-forward stage in the teaching and learning process.

Although there is widespread agreement that feedback is essential for improving learning, there is no consensus on the most effective type of feedback since there are many factors that affect teachers' decisions on what type of feedback should be implemented in their own settings. In EFL writing context, *corrective feedback*, which refers to teachers attempts to minimize learners' errors (Chaudron,1988), is one of the most utilized feedback types, especially in higher education institutions, since in many modern theories, such as social constructivist theory and sociocultural theory, active involvement and participation are at the center of learning, and corrective feedback help learners to achieve self-correction and self-monitoring which foster learning through interaction. Additionally, corrective feedback is an indispensable part of L2 writing since it allows learners to notice the difference between their output and target-like output, and it stimulates a change in learner's interlanguage (Ruegg, 2010; Van Beuningen, 2010). Although early studies in L2 writing sought to determine if providing feedback has positive effects on writing performance, recent studies have investigated other aspects of corrective feedback since providing corrective feedback has been proven to be more effective than not providing any feedback for revision (Ferris & Roberts, 2001). As a result, recent studies have focused on the effectiveness of different types of feedback and teachers' and learners' preferences. Nowadays, both practitioners and scholars are more concerned with enhancing the quality, practicality of feedback as well as determining what type of corrective feedback fits a certain setting and a group of learners the best.

In EFL writing, written corrective feedback is provided for the purpose of helping learners recognize their linguistic errors by indicating the incorrect utterances explicitly or implicitly (Lightbown & Spada, 2006). Two of the main ways to indicate incorrect use of language are providing direct and indirect feedback. *Direct feedback* can be identified as the act of detecting learners' errors and correcting them instead of letting learners figure out how to correct a certain error, whereas *indirect feedback* refers to the act of locating and indicating the errors without explicitly correcting them (Fukuta, Tamura & Kawaguchi, 2019; Lee, Mak & Burns, 2015). Although the studies which investigated the

effectiveness of direct and indirect feedback may have reported different results, a large number of studies (Ferris, 2002; Lee, 1997) indicated that the long-term effects of utilizing indirect feedback are more advantageous in terms of error correction. Furthermore, when students try to figure out what the error might be, more cognitive engagement occurs. Thus, indirect feedback is also considered metalinguistic feedback. In EFL writing, the most common metalinguistic feedback forms are error codes or symbols which allow learners to understand what the error is and where it is placed.

Another point to consider when deciding upon a type of feedback to use is whether the feedback should be *focused* or *unfocused*. While unfocused feedback refers to the correction of all errors, focused feedback refers to focusing on specific features in a text. In addition, feedback can be provided for a single aspect of composition or it can focus on a variety of aspects, such as content, organization, style, and linguistic errors (Hyland, 2009). Although research shows that feedback should focus on both form and content (Ferris, 2010), in EFL school contexts, feedback primarily focuses on grammar (Lee, 2017). According to Lee (2017), the studies which suggest that feedback should target only certain features are conducted in experimental learning environments. Nonetheless, in most real L2 writing classrooms, teachers are required to focus on all the errors or a large number of aspects. Research points out that another reason why teachers tend to provide feedback on grammatical errors in isolation or along with feedback on the content is students' preferences. Many studies show that EFL learners prefer teachers to correct all of their errors and inform them about different aspects of their writing performance (Amrhein & Nassaji, 2010; Üstünbaş & Çimen, 2016).

In a nutshell, providing indirect feedback on both form and content for multiple drafts is suggested by many scholars. Moreover, it is suggested that written feedback is followed by oral feedback or teacher-student conferences so that students are given the chance to ask questions (Ferris, 2014; Weigle, 2002). However, providing feedback in this way is argued to be less practical and time-consuming (Long, 2013; Warschauer & Ware, 2006). Additionally, providing oral feedback and explanations along with written corrective feedback would require further qualitative analysis of the text, thereby requiring more time and effort on the teachers' part. Hence, as a supplementary or additional way of responding to learners' writing, computer-mediated feedback is offered, and recent studies investigate the potential of electronic/automated feedback in EFL classrooms. According to Brown and Bull (2013), one of the advantages of using

computer-assisted assessment or innovative forms assessment is that students can receive timely and repetitive feedback as these types of assessments provide feedback almost immediately and the same assessment for certain students can be received over and over again. Moreover, innovative forms of assessment might be able to provide formative feedback without the time and place constraints (Hepplestone, Holden, Irwin, Parkin & Thorpe, 2011), whereas some problems pertaining to the adequacy and quality of teacher feedback may occur due to time constraints in classroom settings (Gorgogeta & Vlachos, 2019). Nonetheless, innovative forms of assessment, which can provide immediate feedback and assess students' performance, hold the potential for decreasing students' dependence on classroom assessment. They can also promote self-regulated learning while allowing teachers to improve writing instruction and enable them to make decisions regarding writing instruction in a short time (Zeng, 2020).

2.4. Evaluation

Evaluation is defined as the process of collecting information for the purpose of clarifying to what extent teaching and learning processes meet the objectives (Buttram & Covert, 1978; Ioannou-Georgiou & Pavlou, 2003). In other words, evaluation can be utilized to gather evidence to decide if an educational process is effective or not, and necessary changes can be made to enhance learners' or teachers' performances. In addition, evaluation is considered to be the act of judging performance in order to determine the quality or value of performances (Scriven, 1991). Thus, evaluation is often associated with giving grades. Thanks to the feedback received after an evaluation, students and teachers are informed about the progress. Hence, through evaluation, teachers and students can find areas that need revision or attention so that learning can be improved.

The information can be gathered with several tools, such as tests, questionnaires, tasks, and observations. Although the terms evaluation and assessment are often considered to be synonyms, and they are often used interchangeably in the literature, evaluation differs from assessment. One of the main differences between assessment and evaluation is that the latter often involves judgment, such as a grade or a score (Hibbert, Van Deven & Ros, 2012). However, assessment rarely implies the level of quality (Cen & Ross, 2009). In addition, while assessment is more diagnostic given that it indicates the specific areas that need to be improved, evaluation tends to focus on overall

achievement in order to assign a grade. According to Baehr (2005), assessment offers feedback with a view to improving future success and the results of learning, whereas evaluation also facilitates decision-making by determining the degree of success. Hence, it is argued that evaluation is a more inclusive term that incorporates assessment (Grainger, Purnell & Zipf, 2008; Kelleghan & Stufflebeam, 2003). Moreover, it is argued that along with its aforementioned differences, evaluation shares some similarities with assessment. To illustrate, based on purpose and timing, content, and extent, evaluation can be identified as formative, summative, process-oriented, and product-oriented as well as assessment (Long, 1984; Scriven, 1996). Consequently, although evaluation and assessment are two relatively similar terms that researchers tend to use interchangeably, evaluation is usually associated with judgment, and its scope is larger than assessment. Furthermore, it should be noted that both assessment and evaluation are key components of EFL writing.

2.5. Assessment and Evaluation of Writing in EFL Context

Implementing precise methods for testing students' writing ability has been and keeps on being a focal issue in ESL/EFL writing research. Writing in a foreign language does not only require proficiency in that language but also requires writers to express their ideas in an organized way by providing sufficient and relevant details. According to Sallies (1998), writing measurement does not only evaluate grammar competence, but it should also measure levels of overall language competence that includes sociolinguistic and discourse competence which requires students to communicate their message effectively. Hence, over the past few decades, there has been much debate as to whether the direct or indirect assessment should be utilized. In addition, robust research has been conducted on validity and reliability issues of assessment and evaluation methods which include the use of holistic and analytic criteria for evaluating writing (Black, Daiker, Sommers & Stygall, 1992; McNamara, 2000).

For a writing assessment to be considered reliable, the results of the assessment must be the same when it is repeated on different occasions or when the assessment is carried out by different assessors (Hyland, 2019). As for validity, it must be ensured that the test taker's performance is evaluated without any fault and reasonably (McNamara, 2000). Since direct assessment of writing was viewed not as reliable as objective tests due to possible assessor errors, writing used to be assessed indirectly before the concept of

communicative language testing was presented, and in order to minimize the evaluation errors, writing skills used to be measured as one part of an overall test which includes other skills, such as reading, listening and speaking, or it was utilized to measure grammar (Weigle, 2016). Yet, it was argued that evaluating proficiency through writing should not limit students' communicative ability, authenticity, and validity (Bachman & Palmer, 1996; McNamara, 2000). As a result, the importance of direct measurement of writing has been emphasized, especially in the last few decades, and it has become the most common way of assessing and evaluating writing in L1 and L2 writing contexts (Kroll, 1998).

As nowadays writing tasks are generally measured by direct assessment, especially in higher education settings, the measurement heavily depends on criteria-specific rating scales. For evaluation and assessment purposes, the most common rating scales are holistic and analytical scales. Holistic rating, which is also known as "global grading" (Sadler, 2009, p.46), refers to the use of a general impression of the whole composition to determine the quality of writing (Hamp-Lyons, 1990). On the other hand, in analytical scoring, raters are required to measure different aspects of writing, such as content, organization, language usage and mechanics separately and combine the scores of each aspect to reach the final score (McNamara, 2000). Even though there is no consensus over which rating scale leads to a more reliable and valid assessment, it is emphasized that rater training is crucial before any rating scale is utilized in order to produce objective scores (McNamara, 2000). It is also suggested that writing texts can be judged by more than one assessor so that the final score can reflect the quality of writing more reliably. With that being said, it is argued that even though written texts are evaluated by trained assessors based on criteria, errors are likely to occur due to raters. To illustrate, raters' own identity and attitude might result in differences in grades, or student's identity might influence the rater's view of the quality of the written text, which is also known as the halo effect (Schaefer, 2008). In addition to inter-rater disagreements, the variation of scores given by the same rater for the same writing text might be inconsistent as well (McNamara, 2000). Briefly, although rater training and assessment of multiple raters resolve the possible shortcomings of subjective assessment of writing in L2 significantly, the rating might remain subjective.

2.6. Automated Writing Evaluation in L2 Writing

The rising eminence of English in education, business, science, and many other fields leads to English-medium instruction, especially in higher education, where the content of lessons is delivered in English, and students are expected to use the target language competently. Thus, being able to communicate in writing effectively is essential in ESL and EFL settings. In order to help students improve their writing, teachers assign writing tasks frequently, and they assess written composition by utilizing written corrective feedback. However, for teachers to respond to students writing in detail and frequently is time-consuming and labor-intensive (Ferris, 2003; Lee, 2017). In addition, it is argued that human raters' evaluation might be unreliable in cases of fatigue, lack of training and personal factors (McNamara, 2000; Weigle, 2013). Nonetheless, according to many scholars, AWE tools hold the promise to decrease instructors' burden, allow students to become autonomous and produce reliable scores (Shermis & Burstein, 2013; Warschauer & Ware, 2006; Weigle, 2013).

AWE, which is also referred to as automated essay evaluation (AEE), is defined as a computer technology that provides immediate feedback and scores written compositions (Shermis & Burstein, 2013). Nowadays, such software can be used for the purposes of writing instruction, formative assessment, and summative assessment of low-stakes or high-stakes exams (Hockly, 2019). Although most of the commercially available AWE products were released in the late 1990s, initial attempts to develop and use AWE started as early as 1966. Ellis Page, who is regarded as the pioneer of AWE, endeavored to develop the first AWE software when he was a high school teacher in the 1960s. As a former teacher, he was aware that even though it is widely assumed that learners need to write more to improve their composition skills, teachers are able to assign a limited number of essays due to the time-intensive nature of assessment (Shermis, Burstein, Higgins & Zechner, 2010). In the 1970s, with the development of the first AWE tool which was programmed by Page and his colleagues, AWE systems started to be used for scoring and research purposes. Initial research showed that although the first AWE tool was relatively primitive compared to the recent AWE tools, the system was able to produce similar scores to human raters, and surpassed the inter-rater reliability between human raters in some cases (Shermis & Burstein, 2013). In addition, in some cases, the performance of the automated rater was "indistinguishable from human raters" (Attali & Burstein, 2006, p.3). However, due to technological limitations, AWE research has not

accelerated until the advancement of microcomputers and the internet (Shermis et al., 2010).

With the advancement of information, communication, and artificial intelligence technology combined with wide access to computers and the internet, researchers have been able to sophisticate the AWE tools, and more AWE products have been designed to contribute to the assessment process of large scale and high-stakes exams such as TOEFL and GRE (Dikli, 2006). In addition to its scoring features, the system has been used to prepare students for high stake exams, such as TOEFL and GRE for nearly more than a decade (Dikli, 2006). Nowadays, AWE has become web-based and available to be used in language learning and teaching context. Hence, it is being used for assisting teachers and helping learners improve their writing performance by providing individualized feedback on multiple drafts as well as scoring purposes. As a result, the role of AWE has shifted from being a scoring engine to a tool that assists and promotes learning thanks to its formative assessment features (Chen & Cheng, 2008).

AWE's obvious advantages, such as providing instantaneous, individualized feedback on both form and content, its ability to predict accurate scores (Attali & Burstein, 2006; Elliot, 2003) has attracted the attention of testing companies, schools, educators, and researchers. However, the use of AWE in classroom settings is relatively new; therefore, the use and impact of AWE feedback still remain to be examined (Ware, 2011). Additionally, since none of these products were designed to be used in EFL contexts, the research on AWE was limited to psychometric studies, and the studies were either mostly carried out by the software developers in laboratories or they were usually conducted in ESL contexts in the USA since most of the AWE products were based in there (Warschauer & Ware, 2006). Nevertheless, given the importance of writing effectively in L2 contexts, interest in AWE's potential for formative assessment grows and more studies need to be conducted on AWE's validity, reliability, effectiveness, and perceptions towards AWE tools.

2.6.1. Automated scoring and automated feedback through AWE tools

Even though the name AWE may suggest that the software is used for making a judgment of written products, AWE tools can be utilized for both summative and formative assessment. AWE technologies are developed thanks to the contribution of a variety of disciplines, such as computer science, cognitive science, educational testing,

and linguistics since it is believed that all these disciplines reflect the processes that a writer goes through to write prose (Cotos, 2014). Additionally, AWE employs machine learning, natural language processing (NLP) and artificial intelligence to generate scores. Although AWE software is able to generate both analytical and holistic scores, most web-based AWE programs generate scores on the basis of holistic rubrics. AWE programs determine the quality of written texts based on a number of variables that fall under the categories of *trins* and *proxes*. While the former refers to intrinsic variables, the latter refers to approximations of observed variables (Page & Peterson, 1995). General categories of these intrinsic variables, which can be as many as 300 or 500, are content, grammar, style, mechanics, organization, and other categories that are included in the holistic scoring of human raters. Additionally, *proxes*, which refer to approximations (Page, 1966) measure sentence complexity, essay length, and other surface variables in order to evaluate the quality of the composition mathematically (Cotos, 2014; Dikli, 2006; Shermis & Burstein, 2003).

Furthermore, even though opponents of AWE claim that these programs are unable to analyze the content of composition and expressed their disbelief concerning AWE programs' ability to discern meaning (Ericsson, 2006), some researchers (Attali & Burstein, 2006; Elliott, 2003; Wilson, Roscoe & Ahmed, 2017) assert that AWE tools are capable of analyzing the discourse and provide qualitative feedback by employing latent semantic analysis (LSA) and NLP techniques. NLP refers to a set of computational techniques applied in texts or speech in order to detect structural, semantic, and linguistic features by coded algorithms, and provide individualized feedback on errors, cohesion, and coherence (Crossley, Allen, Snow & Mcnamara, 2015; Shermis, Burstein & Bursky, 2013). Thanks to NLP techniques, AWE tools are able to provide diagnostic feedback on specific categories, such as content and development, organization, grammar, mechanics, language use, style and focus, and meaning (Schultz, 2013). Although there is still room for improvement for the NLP techniques applied in AWE programs, researchers claim that NLP strategies might even be able to detect humor in the future thanks to program designers working on sophisticating AWE programs (Roscoe et al., 2017).

On the other hand, some AWE programs employ LSA for the purpose of assessing semantics, content, and linguistic features of texts and provide feedback accordingly. LSA refers to a system that compares a composition to a large corpus in order to identify similarities and differences (Foltz, Streeter, Lochbaum & Landauer, 2013). Assuming that

semantic similarity between words and paragraphs indicates the similarity of meaning (Zupanc & Bosnic, 2017), LSA can generate information by using reference corpora (Shermis, Burstein & Bursky, 2013). Landauer, Foltz, and Laham (1998) assert that LSA is almost as competent as two human raters in measuring meaning. Additionally, in terms of feedback, LSA informs the writer about content and organization along with mechanics, style, and other aspects of writing (Landauer, Laham & Foltz, 2003).

To put it in a nutshell, as the AWE tools are equipped with state-of-art technologies, they might be more advantageous than other standard CALL tools in terms of providing feedback (Cotos, 2010). The developments in innovative technological language assessment techniques, such as NLP were not used to be integrated into standard or traditional CALL applications, thereby making them less able to evaluate writing the way that AWE tools can. Hence, it can be stated that AWE tools might hold the potential for being more effective in teaching and learning writing compared to some other conventional tools.

2.6.2. The use of AWE in L2 teaching and learning

Although teacher feedback is widely acknowledged as extremely valuable, especially for formative assessment, due to human nature, workload, and lack of experience, teacher feedback may not be as effective as it is intended to be. Cotos (2010) identifies the potential drawbacks of teacher feedback as reflecting subjectivity, not being able to offer sufficient scaffolding during writing, lacking elaboration, limited chance to provide individualized feedback. On the other hand, potential advantages of AWE feedback are being goal-oriented, individualized, elaborated, immediate, taking student level and task difficulty into consideration, and allowing learners to self-assess their writing. Moreover, given that AWE programs generate scores based on an immense number of variables and statistics, teachers or school administrations can utilize AWE programs to obtain accurate scores, free of bias (Cotos, 2010).

In addition to the potential benefits of AWE for assessment, EFL writing instruction can be enhanced thanks to AWE. To begin with, AWE software allows teachers to monitor the progress of students thanks to recorded reports on AWE systems, thereby helping teachers making informed decisions about learners' progress and their teaching practices. In other words, teachers can use AWE reports to identify the issues that students struggle with and address these problems to improve instruction and learning outcomes

(Weigle, 2013). Additionally, teachers can tailor assignments instead of using existing prompts, thereby being able to assign curriculum-related writing tasks which stimulate students' prior knowledge of the topic. Finally, many AWE programs include resources, such as prewriting tools for planning and organizing, writers' handbooks, and online thesaurus that can contribute to the instructional efficacy of both novice and experienced teachers. Briefly, AWE can potentially facilitate teaching and supplement writing instruction (Chen & Cheng, 2008; Cotos, 2010; Grimes & Warschauer, 2008; Li et al., 2015; Shermis & Burstein, 2003).

As for the learner side of the equation, the theoretical underpinnings of AWE denote some potential advantages of using AWE for learning. Initially, in the context of formative assessment, teachers' corrective feedback is highly valued by learners, therefore lack of feedback may evoke resentment (Ferris, 2004) while excessive feedback might be perceived as criticism or personal attack (Hyland & Hyland, 2006), and in both cases, students might get demotivated. However, according to William (2011), the intended outcome of feedback is stimulating cognitive growth rather than causing an emotional reaction. Writing is a complex process since it includes other processes within itself. These processes are both cognitive and metacognitive processes, and the employment of metacognitive knowledge, awareness, and strategies is crucial for EFL writers (Teng, 2016). With regard to the role of AWE in metacognitive engagement, Zhang and Qin (2018) assert that new technologies should allow learners to go through planning, monitoring, and evaluation stages which are fundamental in the sense that the same processes are essential in any instructional setting to help learners to improve their writing proficiency (Hayes & Flower, 1980). Thus, AWE might lead to metacognitive engagement given that learners are involved in pre-writing activities and assessment stages. In other words, students receive feedback, edit, and rewrite compositions, and get informed about their weaknesses and strengths, thereby going through planning, monitoring, and self-evaluation stages. Additionally, Cotos (2015) asserts that although more studies need to be conducted to explicate the relationship between AWE and learners' metacognitive skills, existing evidence suggests positive results.

Two other noteworthy skills which can be attained by using AWE tools are self-regulated learning and autonomous learning. Both these skills are of paramount importance owing to their correlation with academic success, especially in higher education (Chen & Cheng, 2008; Yurdugül & Çetin, 2015). Being a self-regulated learner

requires being goal-oriented (Zimmerman, Bonner & Kovach, 1996). According to Cotos (2010), AWE feedback is also goal-oriented as it provides learners the opportunity to monitor their progress throughout their writing process to determine the gap between their current level and the level of writing proficiency that they desire to acquire. Moreover, by offering constant feedback and assistance AWE provides scaffolding that helps students to reach their ZPD (Cotos, 2010). As a result of being actively involved in their own learning process and constantly evaluating their progress, learners become autonomous and intrinsically motivated to take control of their learning (Attali & Powers, 2008; Cotos, 2014; Weigle, 2013). To conclude, becoming self-regulated and autonomous can be among the AWE's long-term advantages for learners, and these skills may contribute to learners' lifelong learning.

On the other hand, AWE has been criticized for ignoring the social and communicative aspects of writing and undervaluing the importance of human interaction (Chen & Cheng, 2008; Ferris & Hedgcok, 2005; Leki, 1990). In addition, the efficacy of AWE in supporting the socio-cognitive aspect of writing, and its ability to analyze meaning is called in question (Ericsson, 2006). Nonetheless, most novel computer-assisted language learning technologies integrate insights from socio-cognitive theories (Sadrolmomaleki, 2009). To illustrate, AWE tools enable online peer-feedback features to improve instruction and learning. Although the feature might be considered relatively novel, researchers emphasize the potential of online peer feedback in terms of error correction and improving the discourse aspect of writing (Yuhong, 2005). Moreover, developers of AWE programs argue that AWE tools do not intend to remove human interaction, on the contrary, AWE tools may facilitate students' interaction with their peers and teachers (Cotos, 2010).

All in all, the scoring and feedback functions of AWE seem pedagogically promising for summative assessment, formative assessment, writing instruction, and self-learning environments. However, empirical evidence is necessary to examine the context validity of AWE prior to conclude that AWE is useful for learning and teaching environments both in theory and practice. (Chen & Cheng, 2008).

2.7. Prior Research on Agreement between Human Raters and AWE Scores in L2 Writing

In an attempt to validate AWE systems' scores, a large and growing body of literature has investigated the validity of AWE scores or agreement between human raters and AWE scoring systems. What is known about AWE scores' validity and reliability is largely based upon empirical studies that were conducted in testing contexts. Hence, much of the current literature on this issue has paid particular attention to validate the efforts of AWE by comparing AWE scores to expert raters scores in high-stake exams, such as TOEFL, GRE, GMAT, and placement exams. Most of these studies have found a significantly high and positive correlation between human raters and automated raters (Attali & Burstein, 2006; Burstein & Chodorow, 1999; Powers et al., 2002; Ramineni et al., 2012; Shermis, Koch, Page, Keith & Harrington, 2002).

To illustrate, Burstein and Chodorow (1999) collected data from both native and non-native English speakers who took the Test of Written English that requires participants to write essays. After comparing the scores of expert raters and the e-rater system, it was found that correlation coefficients between e-rater scores and human raters were significantly high and positive ($r = .73$) and almost the same as the inter-rater reliability between two human raters ($r = .75$). Likewise, Powers et al. (2002) compared human scores and e-rater scores and concluded that the scores given by humans and e-rater were almost the same since the agreements between human raters on different prompts were between 68% and 94%, and agreements between human raters and the e-rater were between 48% and 93%. Based on the findings, it can be concluded that automated scores were significantly and positively correlated with human scores.

In addition, some studies yielded results that prove an agreement between AWE scorers and human raters can be more reliable than an agreement between two human raters (El Ebyary & Windeatt, 2010; Ramineni, 2013; Shermis et al., 2002; Vantage Learning, 2006). For instance, a study by Shermis et al. (2002) investigated the scores given for 300 online placement essays by both an AWE tool and expert raters. The study revealed that agreement between PEG automated assessment tool and six experts was significantly and positively correlated ($r = .83$), whereas the correlation between human raters was considerably lower ($r = .51$). Similarly, Vantage Learning Company (2006) investigated agreement rates between human scores and AWE scores. When compared to agreement rates of two experts, ranging from 58% to 61%, the Intellimetric system and

expert raters had higher agreement rates ranging from 65% to 70%. Additionally, the correlation coefficient between experts and Intellimetric was as high as .93. The results of this study show that the correlation between an AWE tool and a human rater can be higher than the correlation between two human raters' scores.

On the other hand, the research conducted in classroom settings in the native English-speaking (NES) context has found contradictory results (see Table 2.1). For instance, Wilson and Roscoe (2020) argued that the correlation coefficient between human raters and AWE programs may not be as high as it was reported in previous studies provided that raters' background and the research setting might change the results. The results of a study conducted by Wilson and Roscoe (2020) in a middle school in NES context concluded that the correlation coefficient between teachers was considerably high ($r = .98$), while the correlation between the PEG program and the teachers was lower ($r = .62$). Likewise, another study conducted by Ramineni et al. (2012) found that although the discrepancy between human raters and e-rater was small in most prompts, there were some prompts that correlation between human raters and E-rater was lower than .70. Nonetheless, Ramineni (2013) yielded positive results when a validation study was conducted in a university setting and in NES context. The study revealed that when the human-human and human-AWE system combinations are compared in terms of scores given for a set of prompts, the correlation coefficient between AWE and human raters was within .71 and .80, whereas, the correlation range was lower between two human raters ($r = .61-.76$). Furthermore, in a study, Attali and Burstein (2006) found almost perfect agreement ($r = .97$) between Criterion and human raters. Hence, it was concluded that AWE systems, such as e-rater can enable higher standardization of scores, thereby validating the AWE scores. Consequently, although some studies found contradictory results, most of the studies conducted in testing context and NES context revealed that AWE scores were in good agreement with human scores and it was interpreted that AWE scores might be valid, and they can be utilized to achieve higher inter-rater reliability.

Even though there has been an increasing amount of literature in ESL and EFL settings compared to the last decade (see Table 2.1), the generalisability of published research on AWE validity in EFL and ESL contexts is problematic. To illustrate, Coniam (2009) used BETSY software to examine correlations between human raters and BETSY software. In order to compare the scores, 900 texts were scored by human raters and the BETSY tool. The results showed that the correlation between human raters and the AWE

tool was significantly and positively high ($r = .80$). Therefore, Coniam (2009) concludes that AWE software holds the potential to be employed as a third rater. However, a study conducted by Li, Link, Ma, Yang and Hegelheimer (2014) in a similar ESL context found moderate correlations in the range of .42 and .47. Although the studies conducted in EFL settings are considerably scarce, a promising study was conducted by El Ebyary and Windeatt (2010). In this study, Criterion scores were significantly and positively correlated ($r = .83$) with one of the raters and moderately correlated ($r = .59$) with the second-rater, whereas the correlation between two human raters was lower and moderate ($r = .45$). In conclusion, the findings of these studies suggest that AWE can be positively and significantly correlated with human raters and may substitute human raters in the scoring process of high-stake exams and classroom assessment of essays. The summary of AWE research mentioned in this section can be seen below in Table 2.1.

Table 2.1. *A summary of research on correlation between human rater scores and AWE scores*

Study	AWE system	Context	Human-human agreement & correlation (r)	Human-AWE agreement & correlation (r)
Burstein & Chodorow, 1999	E-rater	Native Speakers: Test of Written English (TWE)	.75	92 % .73
Shermis et al., 2002	PEG	Native Speakers: Placement Tests	.51	.83
Powers et al., 2002	E-rater	Native Speakers : Graduate-level writing assessment (GRE)	68%- 99 %	48%- 93 % .73
Vantage Learning, 2006	Intellimetric	Native Speakers K-12	58%- 61%	65% -70% .93
Attali & Burstein, 2006	Criterion	K-12, GMAT & TOEFL	-	.97
El Ebyary & Windeatt, 2010	Criterion	EFL	.45	.83
Ramineni, et al., 2012	E-rater	TOEFL	.69-.82	.70
Ramineni, 2013	Criterion	Native Speakers: University	.61-.76	.71-.80
Li, et al., 2014	Criterion	ESL	-	.42 - .47
Wilson & Roscoe, 2020	PEG	Native Speakers K-12	.98	.62

As it can be seen above, due to the limited number of studies conducted in ESL and EFL settings, more studies are needed to investigate the potential of using AWE systems in L2 writing scoring.

2.8. Prior Research on the Effectiveness of AWE on Improving L2 Writing

Acknowledging that AWE feedback can positively affect the revision process, there have been some quantitative studies on whether AWE feedback encourages students to use formative feedback to revise their drafts. To illustrate, Waschauer and Grimes (2008) examined if AWE feedback has a positive influence on revision rates and yielded favorable findings. This classroom-based study included students in middle and high schools in NES context. After AWE programs were utilized for a year, it was reported by teachers that students were actively involved in autonomous activities thanks to AWE programs, and it was shown that the successful revision rate was 72 %. Although students were provided feedback on different aspects, such as organization, content, style, and grammar, most of the improvement was on word level, and revisions were mainly on mechanics.

Another study conducted by El Ebyary and Windeatt (2010) yielded favorable results in terms of the use of AWE feedback in the EFL context. In the study, the effects of AWE feedback on submission, revision, and overall improvement of the papers and participants' attitudes towards AWE were examined, and both quantitative and qualitative data were collected from 31 instructors and 549 students. However, the data concerning the revision, submission rates were collected from 24 undergraduate students who were studying their major in ELT. The students were asked to submit two drafts on four different topics and revise the feedback given by AWE for the first draft to write their second draft. Although pre-feedback interviews indicated that most of the students answered "occasionally" or "never" when they were asked how often they tend to write essays, the results showed that the students submitted two drafts for each prompt, thereby increasing their submission rates to 100 percent. Additionally, the results yielded that formative AWE feedback resulted in consistent growth in scores. To illustrate, results showed that in the final essays, a large proportion of students scored five out of six, whereas the majority of the students had scored three out of six at the beginning of the research. Furthermore, qualitative data supported the quantitative data given that some students expressed that AWE led to cognitive awareness since they were not aware that

some aspects of their writing were undesirable, such as repetition of words. Yet, Criterion feedback allowed them to notice additional areas of improvement and make necessary changes to see how the results may change.

Correspondingly, Chapelle, Cotos, and Lee (2015) conducted a classroom-based study that investigated students' inclination to revise and rewrite drafts by using AWE feedback. The data were collected from a group of international university students who were enrolled in an academic writing course. The students were asked to submit and revise several drafts on the Criterion system and use the formative feedback to edit their essays and correct their errors. The results substantiated the claims that AWE feedback can motivate students to revise their essays since students were able to utilize feedback to edit their paper by 70%. Additionally, revision classification revealed that the most recurrent revision type was word changes. In short, the literature pertaining to the effects of formative AWE feedback on students' tendency to revise and improve the quality of writing strongly suggests that AWE feedback can encourage students to revise and improve their writing compared to conventional feedback types.

In addition to those studies, AWE programs' ability to improve students' writing has also been investigated. Much of the previous studies on AWE feedback's impact on writing improvement have been conducted in NES context. One of the early studies investigating the effectiveness of AWE in an instructional setting was conducted by Attali (2004). The data for this study was collected from K-12 students from 6th grade through 12th grade across the country. In order to determine whether a significant change occurred in students writing competence, thousands of first and final drafts of essays were examined. As a result, it was found that students' writing performances were improved in terms of both form and content. To illustrate, most of the errors were decreased significantly, discourse elements of the essays were improved, and the lengths of essays were increased. Additionally, students were able to improve their overall scores.

Another relatively current study conducted by Palermo and Thomson (2018) in NES context found similar results. The study aimed to explore the effects of combining AWE with different instructional strategies. The participants were 6th through 8th-grade students enrolled in summer school programs across a state and students were divided into three groups based on the type of instruction they receive and the type of writing process they went through. The instruction strategy employed for one of the groups was Self-Regulated Strategy Development (SRSD) and the other two groups received

conventional instruction. In addition to different types of instruction strategies, one group received conventional feedback while the SRSD group and one more group used an AWE tool to receive feedback. The findings of the study revealed that when an AWE tool is combined with conventional writing instruction, students were able to compose essays in higher quality compared to the group receiving conventional feedback. Nonetheless, the most successful group out of three was the one that received SRSD instruction combined with AWE feedback. The results of the study also reported positive changes in students' writing quality. Briefly, it can be inferred that the results of the studies that examined the effectiveness of AWE on improving writing revealed favorable findings.

Apart from these, there also have been studies conducted in ESL and EFL contexts. To illustrate, Li, et al. (2015) carried out a study to determine how AWE feedback impacts students' writing practice and how writing accuracy changes from the first draft to the final draft as a result of AWE feedback. To be able to answer these questions, quantitative data were collected from four instructors and 70 students in an ESL context over 15 weeks. In terms of writing instruction, teachers followed a process approach and assigned students to submit their essays on an AWE tool called Criterion and the students used the AWE corrective feedback to write their final drafts. The quantitative data analysis revealed that error rates were significantly lower ($p < .01$) in final drafts and AWE corrective feedback was able to improve writing quality in many aspects, especially regarding grammar and mechanics and overall linguistic accuracy. Researchers conclude that in the light of these findings, it can be asserted that AWE feedback has the potential to improve L2 writing instruction.

In another study which was conducted by Mohsen and Abdulaziz (2019), the effects of using only AWE feedback for revision and combining AWE feedback with instructor feedback, which they define as "hybrid mode", were compared. The study was divided into two phases and in the first phase, six students submitted their essays on an AWE program to receive feedback on errors, content, organization, and style. Subsequently, students were asked to revise and edit their first drafts accordingly. However, in the second phase, when students submitted their first drafts on another topic, they were told to use AWE feedback on error categories, and further feedback regarding content, organization, and style had been provided by teachers before they wrote the second draft. Consequently, paired t-test analysis showed that there was an improvement from the first drafts to final drafts in both modes. Nonetheless, it was found that the hybrid mode,

applied in the second phase, led to a significant improvement and outweighed the scores in the first phase when only AWE feedback was utilized. All in all, the findings indicate that involving teacher feedback in process writing along with AWE feedback might result in significantly better results.

Along with the studies investigating the effects of AWE corrective feedback on students' overall improvement, there have been some studies that examined the impact of AWE feedback on various aspects of writing to determine the areas that are most likely to be improved by AWE corrective feedback. For instance, research by Long (2013) was carried out to identify the main differences between AWE feedback and teacher feedback, and it also aimed to find out whether AWE feedback brings about improvement in writing in the EFL context or not. The data for the first question were collected from 26 students in a university and for the second question, 216 students' papers on six different topics and multiple drafts of each topic were examined. The comparison of AWE feedback and teacher feedback on writing revealed that AWE feedback focuses on linguistic and stylistic issues, whereas teacher feedback focuses on meaning more. Additionally, the findings regarding the second question showed that marginal improvement occurred in four areas while minimal improvements were observed in the other areas. The areas that were improved by the students the most were confusing words, punctuation, spelling, and short sentences.

Likewise, Ranalli, Link, and Chukharev-Hudilainen (2017) conducted a study examining the Criterion tool's ability to decrease errors of college students in an ESL context. The study included 36 students in lower-level proficiency and 46 students in upper-level proficiency, and the results showed that lower-level students benefited more from AWE feedback compared to upper-level students. The researchers infer that the results may indicate that AWE feedback is more useful to improve form-based aspects of writing, therefore the implementation of AWE tools in classrooms with lower-level students, where form-focused instruction is utilized, might result in considerably better results. To illustrate better, in terms of the AWE tool's ability to decrease errors across different categories, the results indicate that errors, such as wrong words, punctuation, subject-verb agreement and missing articles were much less challenging to correct compared to run-on sentences and fragment issues. To put it in a nutshell, studies that looked into the impact of AWE feedback on accuracy across specific errors and areas of improvement in writing found that AWE can result in significant improvement in word

and sentence-level writing performance whereas teachers' additional support might be needed in terms of discourse-level construction.

Apart from these studies, some researchers have carried out comparative studies that examined the differences between control groups receiving traditional feedback and experimental groups which were provided with AWE feedback. Khoii and Doroudian (2013) carried out research to identify the differences between the effects of utilizing AWE feedback and teacher feedback in formative assessment. In order to collect data, 11 EFL students were assigned to submit seven essays on an AWE system, and 11 students in the control group handed in their essays to their teachers to receive feedback. For both groups, a process writing approach was used for ten weeks. To identify the differences between the groups, students' first drafts and final drafts were analyzed, and in order to see the differences in terms of summative assessment, a final test was conducted to collect a single product from each participant. The findings of the research showed that the control group was significantly outperformed by the experimental group based on the formative assessment of multiple drafts. However, no significant difference was found in students' final papers.

By the same token, Wang, Shang, and Briody (2013) conducted a study that investigated the effects of AWE on writing performance and found similar results. Participants of the study were 57 EFL students in a university and both qualitative and quantitative data were collected from the participants. In order to determine if there was a significant difference between the groups in terms of changes in writing accuracy, the period of the study was divided into three stages. Initially, the students in the experimental group received training on how AWE works, and both groups composed a pre-test. Subsequently, the students in the experimental group submitted their essays on the AWE program, received AWE feedback, and wrote their final drafts accordingly while the control group received teacher feedback to write their final drafts for fourteen weeks. Finally, both groups wrote an essay about the same topic as they were given in the pre-test stage. Analysis of students' first drafts and final drafts, submitted for fourteen weeks, showed that the experimental group was able to decrease their errors in their final drafts more than the control group, and there was a significant difference between the groups in terms of accuracy. Additionally, it was revealed that there was a significant difference in terms of accuracy when the experimental group's pre-feedback and post-feedback compositions were taken into consideration. Conversely, when Wilson and Czik (2016)

combined teacher feedback and AWE feedback to integrate into English Language Arts courses in a secondary school to see if there is a significant difference between the conditions of receiving combined feedback and only teacher feedback, they found no significant differences between the experimental and control group regarding the final drafts of students' essays.

In another study that was carried in the EFL context, Cheng (2017) examined the effect of AWE feedback on the writing quality of learners' reflective journals. The participants were 138 students in a university and students were divided into two groups to see the differences between the experimental group and the control group in terms of writing improvement. Both groups were taught by the same teacher; therefore, there was no difference between the groups regarding the writing instruction. Students in both groups were required to submit three entries to a web-based AWE program. However, only the students in the experimental group were allowed to use AWE feedback of the system to revise and improve their entries while the control group didn't receive any automated feedback. In order to determine the effects of AWE on writing quality, a mixed-methods approach was employed. The quantitative data were collected through a survey, AWE scores, and human scores given for students' initial and final drafts. The qualitative data were collected through semi-structured interviews, observations, and think-aloud protocols. The findings showed that although the participants in the experiment group scored lower than the participants in the control group at the beginning of the implementation, post-treatment results showed that the experimental group outperformed the control group and there was a significant difference between the experiment group and control group scores. In addition, qualitative data analysis revealed that the majority of the students found AWE feedback helpful for improving writing.

In the Turkish EFL context, a study conducted by Bulut (2019) examined the effect of an AWE tool on writing achievement. 60 students, studying at a state university, participated in the study, and 30 of the students were placed in the experimental group, whereas the other 30 students were placed in the control group. During the instruction period, the control group received teacher feedback, whereas the experimental group uploaded their essays to the AWE program and received automated feedback. A writing test at the beginning of this period and another writing test at the end of the instruction period were utilized to collect data, and these were respectively used as the pre-test and post-test data to compare the writing achievement results of the two groups. The results

of the study revealed that even though both traditional and AWE assessments led to enhanced writing achievement, the students in the experimental group got higher scores compared to the students in the control group. Hence, it can be inferred that AWE might be used to enhance teaching and learning writing in the Turkish EFL context.

Another study in the Turkish EFL context was conducted by Gençer (2019), and it investigated how automated feedback affects writing performance in terms of grammar and mechanics both in the short-term and the long-term. In order to investigate the short-term effects of automated feedback, data were collected through students' paragraphs and essays. An AWE tool was used to provide feedback to students while they were writing their drafts. In total, each student wrote eight drafts and four of these drafts were paragraphs, whereas the other four drafts were essays. While students only used automated feedback outside the classroom to write their drafts, they received teacher feedback after they submitted their drafts. The students' errors under nine categories were counted for each writing draft to explore how automated feedback affects writing in the short term in terms of grammar and mechanics accuracy. The findings showed that significant error reduction was observed in some papers and automated feedback had a positive effect on students' writing performance in the short term. Moreover, students' first paragraph drafts and last essay drafts were included in data analysis to explore the long-term effects of automated feedback. The findings revealed that automated feedback had positive effects on students' writing performance in the long term as well. Similar to the study by Bulut (2019), this study indicates AWE might improve students' writing in L2. In addition, it can be stated that AWE in the EFL writing context might promote improving the grammar and mechanics aspects of writing. Nevertheless, to be able to generalize the findings, more studies need to be conducted in the Turkish EFL context.

2.9. Prior Research on Students' and Teachers' Perceptions of the Use of AWE in L2 Writing

A growing body of literature suggests that AWE programs hold the potential to improve students' writing performance thanks to the formative feedback offered by the system, and its scoring ability. However, there are many other factors influencing students' achievement, such as students' motivation, anxiety, perception of classroom practice, classroom environment, and possibly their teachers' own views. Since affective factors also play a crucial role in students' actions and success, teachers' and students'

perceptions need to be examined prior to integrating an innovative program or approach into existing conditions in a setting.

Prior research on students' and teachers' perceptions has produced mixed findings. To illustrate, after carrying out a study on students' perceptions towards AWE in EFL context, Chen and Cheng (2008) reported that students were anxious to use AWE in the classroom. They asserted that AWE feedback should not replace teacher feedback, and they expressed a lack of trust in AWE feedback and only half of the classroom found AWE feedback helpful. Conversely, when El Ebyary and Windeatt (2010) conducted a study in a similar context with 24 ELT trainees, interviews with the trainees and post-treatment questionnaires showed that after the trainees used the AWE program over eight weeks, participants' attitudes towards AWE changed positively. Even though participants expressed their distrust and expectation of dissatisfaction with the program, post-treatment questionnaires and interviews showed that 88% of the participants were satisfied with feedback noting that automated feedback led to positive changes in their writing performance and led to increased awareness of their progress. Moreover, Wilson and Czik (2016) investigated the effects of combining teacher feedback with AWE feedback on students writing motivation. The comparative study revealed that students in the experimental group reported that they found AWE feedback motivating. The findings suggest that students may benefit from AWE feedback, combined with teacher feedback, to show persistence, especially at the initial stages, as students who have higher motivation are more likely to revise their drafts.

In a study carried out by Li et al. (2014), researchers investigated how teachers and students perceive AWE holistic scores. The data consisted of surveys and interviews conducted with three writing instructors and 14 students. After the AWE program was integrated into an ESL writing curriculum, questionnaires were distributed following the first paper and fourth paper. Consequently, instructors revealed that they agreed with the low grades given by AWE; however, they disagreed with the high grades stating that AWE scored some papers higher than they would. Conversely, it was reported that the students who had no prior experience with an AWE program expressed that they gained trust in the system which is contrasting with Chen and Cheng's (2008) findings indicating students' lack of trust in AWE systems' assessment.

Furthermore, a comprehensive study conducted by Li et al. (2015) examined how students and teachers view the use of AWE feedback. The data were collected through

interviews with 27 students and four teachers after the teachers and students used an AWE program for the writing lesson over 15 weeks. Qualitative data analysis revealed that instructors found AWE advantageous by stating that the AWE program changed the way they respond to papers positively. To illustrate, instructors noted that they saved a considerable amount of time by not providing feedback on some areas, such as grammar, mechanics, thereby allowing them to focus on content and organization aspects of writing only and saving time. Furthermore, instructors expressed that AWE would be adequate to raise learner's awareness of their sentence-level errors and allow them to separate form from content when they give feedback. In addition, it was reported that AWE feedback motivated students to use feedback and revise their writing persistently. However, instructors also shared their concerns that the long-term effects of AWE feedback might not be as successful as traditional feedback. Additionally, instructors noted that AWE feedback needs to be improved in terms of its response to content and organization. Likewise, according to the results, 18 out of 27 students shared their satisfaction with the program noting that they found AWE feedback helpful and useful for correcting errors. Taking all the findings into consideration, it can be inferred that both the teachers and the students perceived that AWE can positively affect the L2 writing practice in classrooms.

In addition to the studies examining students' perceptions, there have been some studies focusing on teachers' and principals' perceptions of AWE. To illustrate, Warschauer and Grimes (2008) integrated AWE into a high school curriculum in NES context. The sources of qualitative data were semi-structured interviews with eight teachers, three principals, and two students who used an AWE program for over a year. The interviews with teachers and principals revealed that they were all satisfied with the addition of AWE into their curriculum by noting that AWE increased students' motivation and changed their revising behavior positively. Additionally, it was noted that AWE had a positive influence on students' performance in terms of content since, reportedly, students' writing became more "insightful" and "creative" (Warschauer & Grimes, 2008, p.28). Moreover, teachers stated that one of the biggest advantages of AWE was engaging students in an autonomous learning condition and saving their time.

Another study that focused on teachers' perceptions of the use of AWE in the classroom was conducted by Link et al. (2014). The study investigated the perspectives of instructors who integrated AWE in their ESL writing lessons at a university. The data were collected through observations, individual and group interviews. The five

participants used AWE in their writing lessons for a year. An AWE program was employed to be used in writing instruction and various stages of writing, such as pre-writing, peer review, and revision. The system was also utilized for formative assessment of writing in seven classrooms so that teachers could observe students and get insights regarding the potential of AWE. The qualitative data analysis showed that the results were similar to the findings of Li et al. (2015) as instructors stated that the AWE tool was competent in finding grammatical and mechanical issues of writing by allowing instructors to save their time to focus on content-related and organizational issues. Furthermore, instructors mentioned that they found AWE programs useful for boosting learners' metalinguistic ability, reducing teachers' workload, and providing feedback on the sentence level. Nonetheless, they also noted that AWE feedback needs to be combined with teachers' feedback to provide high-quality feedback and give reliable scores.

In the Turkish EFL context, learners' opinions towards the use of AWE in L2 writing have been addressed only to a very limited extent due to lack of research. To the researcher's best knowledge, the two studies that have examined learners' opinions towards the use of AWE in L2 writing in the Turkish EFL context are the aforementioned studies conducted by Gençer (2019) and Bulut (2019). The study conducted by Gençer (2019) explored students' opinions towards the use of AWE through pre-surveys which were administered prior to students' exposure to automated feedback, and post-surveys were administered following the treatment. The survey included items measuring students' opinions towards the use of AWE feedback and their views of technology use in L2 writing. The results showed that all the participants found the AWE tool helpful and the majority of the students were satisfied with automated feedback, especially in terms of grammar and mechanics. Correspondingly, the majority of the students expressed that AWE feedback helped them to improve their writing skills. Moreover, the error categories that students reported that AWE was most helpful for were, respectively, punctuation, spelling, and grammar. All in all, the findings of the study showed that students' opinions towards the use of AWE feedback and technology in L2 writing were positive even though they were skeptical of the use of technology for learning prior to the study. Hence, it can be concluded that students' involvement with technology to promote learning might change their perception of the use of technology to enhance their learning, and it might improve their learning.

Likewise, Bulut (2019) investigated the impact of using the AWE tool on writing motivation. The data were collected by a questionnaire including items related to students' goal orientation, self-efficacy, intrinsic motivation and extrinsic motivation. The comparison of the experimental and the control groups' answers to the questionnaire revealed that, while the control group appeared to be more extrinsically motivated, the experimental group was found to be more intrinsically motivated, and their self-efficacy level increased thanks to their experience with the AWE tool. Therefore, it can be inferred that the motivation of the control group might have stemmed from the results of their writing performance while the experimental group might be motivated by their positive feelings towards the writing in L2. Moreover, it was stated that following the treatment, students became more eager to write in L2. The findings of these studies indicate that AWE might encourage students to write in L2 and allow them to become more autonomous by promoting self-regulated learning.

Although the importance of context validity is emphasized by researchers (Chen & Cheng, 2008), there is a limited number of studies investigating students' and teachers' perceptions towards the use of AWE in a classroom setting. Much of the existing literature has focused on the effectiveness of AWE assessment and feedback and its validity pertaining to the correlation between AWE scores and instructors' scores. Hence, the generalizability of the findings is problematic, and more studies need to be conducted, especially in the EFL context due to the fact that the employment of AWE and related research is already restricted in the EFL context. Furthermore, in the Turkish EFL context, the agreement between automated scores and AWE scores, the effect of automated feedback on L2 writing, and learners' and teacher's views on using AWE in EFL writing have been rarely investigated. To the researcher's best knowledge, there is no prior study examining teachers' views on the use of AWE in formative classroom assessment. Hence, previous research in the Turkish EFL context can only be considered the first step toward a more profound understanding of the role AWE might play in the Turkish EFL context. Consequently, this study aims to fill the gap in the literature by looking into the impact of AWE on writing accuracy, the correlation between AWE scores and Turkish EFL instructors' scores, and learners' and teachers' views on using AWE for teaching, learning, and improving L2 writing.

3. METHODOLOGY

The main purposes of this chapter are to describe the design of the study, introduce the setting and the participants, explain the data collection instruments utilized in the study, provide information about data collection procedures, and depict the data analysis process.

3.1. Research Design

A concurrent triangulation mixed-methods design was employed in this study. This design involved a single study including both qualitative and quantitative data collection which were conducted at the same time (Creswell, 2009). Therefore, the data collection was concurrent, and two methods were used to cross-validate, confirm and substantiate the findings within the study. First, the data were collected, and the analyses were carried out afterwards. In other words, triangulation occurred once all the results were in. Subsequently, the findings were compared, contrasted, and similarities and differences were revealed. The purpose of employing this research design was to corroborate and complement the findings generated by each method. The concurrent use of quantitative and qualitative data collection and analysis was also utilized to add to the depth and scope of the findings. Figure 3.1 illustrates the research design employed in this study.

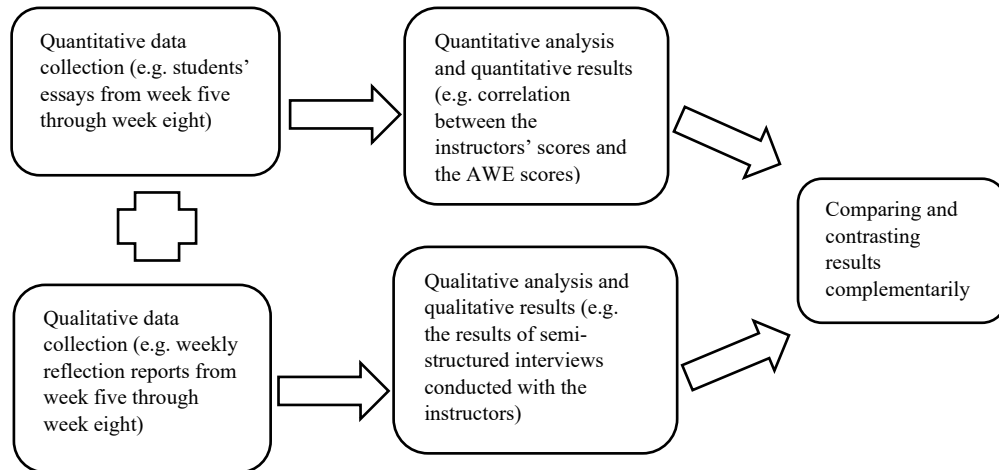


Figure 3.1. *The research design employed in the study*

Concerning the quantitative dimension of the study, correlational and comparative research designs were adopted as the first research question is related to the correlation of AWE scores and human scores, and the second one addresses how AWE feedback change writing accuracy. With regard to the qualitative dimension of the study, which

investigates the students' and the instructors' views on the use of AWE in L2 writing assessment, data were collected through written reflection reports and semi-structured interviews. Data were interpreted to reach a conclusion in terms of participants' views on the use of AWE for formative assessment in the EFL context.

3.2. Setting and Participants

3.2.1. Setting

The present study was conducted in the School of Foreign Languages Department at Antalya Bilim University in Turkey. The English Preparatory Program (EPP) of the school is compulsory for most of the departments unless students pass the proficiency test prepared by the testing unit of the school. The EPP comprises five levels which are elementary, pre-intermediate, intermediate, upper-intermediate, and pre-faculty levels. Upon entering university in the fall, students take a placement test and provided they are placed in intermediate or above, students can sit for the proficiency exam. To be exempted from the EPP of the school, students must receive 70 points or higher. Otherwise, they are required to start the first module at the level determined by the placement test. If they complete the intermediate or higher levels successfully, students get another chance to sit for the proficiency exam. Nonetheless, if they cannot get the required score from the proficiency exam, they need to continue studying in EPP for the following semester or year until they are qualified to start their education in their selected departments.

One year of the EPP includes four modules and each module lasts for eight weeks. Each week students receive 25 hours of instruction of which four hours are allocated for writing instruction and practice. In order to pass a module, students are required to obtain a grade of 70 points when all the percentages of course components are totaled up. The course components include writing tasks, speaking tasks, presentations, midterm, and final exams. While the percentages of components vary, two writing tasks account for ten percent of the overall score and another writing task is included in the final exam which comprises of 20 percent of the final exam.

With regard to writing lessons, the process approach to writing instruction is integrated and formative assessment is used for the purpose of both helping students attain the learning objectives. Accordingly, students constantly write more than one draft on a variety of topics related to curricular themes, and formative corrective feedback is provided by the course instructor for each draft. As for writing tasks, the average of the

first and second drafts is accepted as students' final grades. However, as for the writing part of the final exam, which only takes place at the end of each module, summative assessment is utilized. Since writing skills are taught both as integrated with other skills and independently in the EPP, much time and effort are devoted to writing instruction and assessment. All the class materials, tasks, and exams are prepared by the testing unit of the EPP. In addition, instructors are obliged to follow the same procedures step by step in writing instruction, and they are only allowed to choose the topics for practice from a predetermined list. Hence, writing instruction and assessment are required to be standard.

3.2.2. Participants

The participants of this study (n=38) were selected through the convenience sampling method since the participants were accessible (Dörnyei, 2007; Given, 2008) as the researcher is also an instructor at this particular school of foreign languages department. Additionally, convenience sampling was used in this study since participants participated voluntarily (Dörnyei, 2007) by signing informed consent forms (see Appendix A) at the beginning of the study. The study was carried out in the fall term of the 2019- 2020 academic year at four intermediate-level classes. Two of the classes were set as the experimental groups, including 18 participants, and the other two classes were set as the control groups, also including 20 participants. The main reason why four classes were recruited for the study is that the number of students in each classroom is limited with only 14 to 17 students per class.

AWE feedback was provided only for the students who volunteered to receive AWE feedback in the experimental group. Even though repeat students were allowed to use AWE feedback if they volunteered, their data were excluded due to a number of reasons. Initially, as repeat students were not placed in intermediate level at the placement test, their level might have differed from the students who were placed by the placement test. Secondly, repeat students' excessive exposure to traditional feedback, given previously in the research context, might have an additional effect on their views of AWE, and minimizing the recall bias would yield more accurate results. Apart from the repeat students, 52 students participated in the study at the beginning of the study. 26 of these students were in the experimental group and 26 of them were in the control group. However, due to a number of reasons, such as absenteeism, incomplete data, and not taking part in all steps of the study, the data gathered from eight students in the

experimental group and six students in the control group were removed from the analysis. Hence, only the data gathered from 18 students from the experimental group and 20 students from the control group were used. The researcher was also the course instructor of the experiment groups since using AWE technology requires prior knowledge and practice. Moreover, the experimental groups had to be guided and trained in using AWE feedback to revise and edit their essays throughout the study.

By the same token, in the qualitative part of the study, only the data that were gathered from the students who participated regularly in all steps were included in the data analysis. Therefore, written answers to weekly reflective report questions were collected from the same 18 students in the experimental groups. Moreover, from the experimental groups, 13 students who used AWE feedback consistently participated in semi-structured interviews voluntarily.

In addition to students, semi-structured interviews were conducted with seven teachers. Before the semi-structured interviews were conducted, participants were informed that recordings would be used for research purposes, and they all signed informed consent forms (see Appendix B). All the teachers were experienced, non-native EFL instructors and all the instructors reported that they had no experience with AWE technology prior to this study. Most of the teachers graduated from ELT departments with BA or MA degrees. Furthermore, all the instructors received a one-year intensive training program provided by the professional development unit of the school. This program trains instructors on lesson planning, classroom instruction, formative assessment, summative assessment, and some other educational practices. Moreover, instructors attend standardization sessions prior to each writing test. Hence, it is constantly assured by the testing unit that there is little to no discrepancy among the scores given by the instructors. The following Table 3.1. shows instructors' demographic information in detail.

Table 3.1. *Demographic information of the instructors*

Instructors	Highest degree of education	Graduation	Teaching Experience	Prior Experience with AWE
Instructor 1	MA	ELT	4 years	None
Instructor 2	MA	ELT	3 years	None
Instructor 3	Bachelor's Degree	Non-ELT	3 years	None
Instructor 4	MA	ELT	8 years	None
Instructor 5	MA	ELT	7 years	None
Instructor 6	MA	Non-ELT	5 years	None
Instructor 7	Bachelor's Degree	ELT	4 years	None

3.3. Data Collection Instruments

This study investigates the efficiency of the use of AWE in the formative assessment process, the agreement of AWE scores with human scores, and the views of teachers and students of the use of AWE in EFL writing. The data collection instruments that were used to answer the questions of this study consisted of:

- the AWE tool (Criterion)
- essays, written by students
- written reflection reports written by the students
- semi-structured interviews conducted with the students
- semi-structured interviews conducted with the teachers

Two of these instruments (the AWE tool and the students' essays) were parts of the quantitative data collection process and the remaining two instruments (the written reflection reports and the semi-structured interviews) were utilized to provide data for qualitative data analysis. In the process of deciding on data instruments and developing them, three experts at Anadolu University, ELT department were consulted. A summary of the data collection instruments utilized for each question can be seen below.

Table. 3.2. *Research questions and data collection instruments*

Research Questions	Data Collection Instruments
1) Is there a relationship between AWE holistic scores and instructors' holistic scores given to students' essays?	AWE program and students' essays
2) Is there a significant difference in terms of writing accuracy between a group of students receiving AWE formative feedback with a group of students receiving teacher feedback?	AWE program and students' essays
3) How do EFL students view the use of AWE in L2 writing?	Written reflection reports and semi-structured interviews
4) How do EFL instructors view the use of AWE in L2 writing?	Semi-structured interviews

3.3.1. The AWE tool

Criterion is a web-based AWE program that was released by ETS in 2002. Criterion utilizes NLP methods to provide comments, and it uses an e-rater scoring application to measure the overall quality of writing. Criterion is able to score essays in terms of organization, style, grammar, mechanics, usage, vocabulary use, and lexical complexity

and it provides corrective feedback on a variety of error types under these categories and more, such as formation errors, punctuation, agreement errors, wrong word, spelling errors, etc. As for style errors, the system detects the essential parts of an essay, and it reminds users to revise these parts by taking the systems' style suggestions into consideration. The system also warns users about informal language use, repetitions, and short sentences. The following Figure 3.2 shows the categories in which the system provides feedback and how this system warns users about errors.

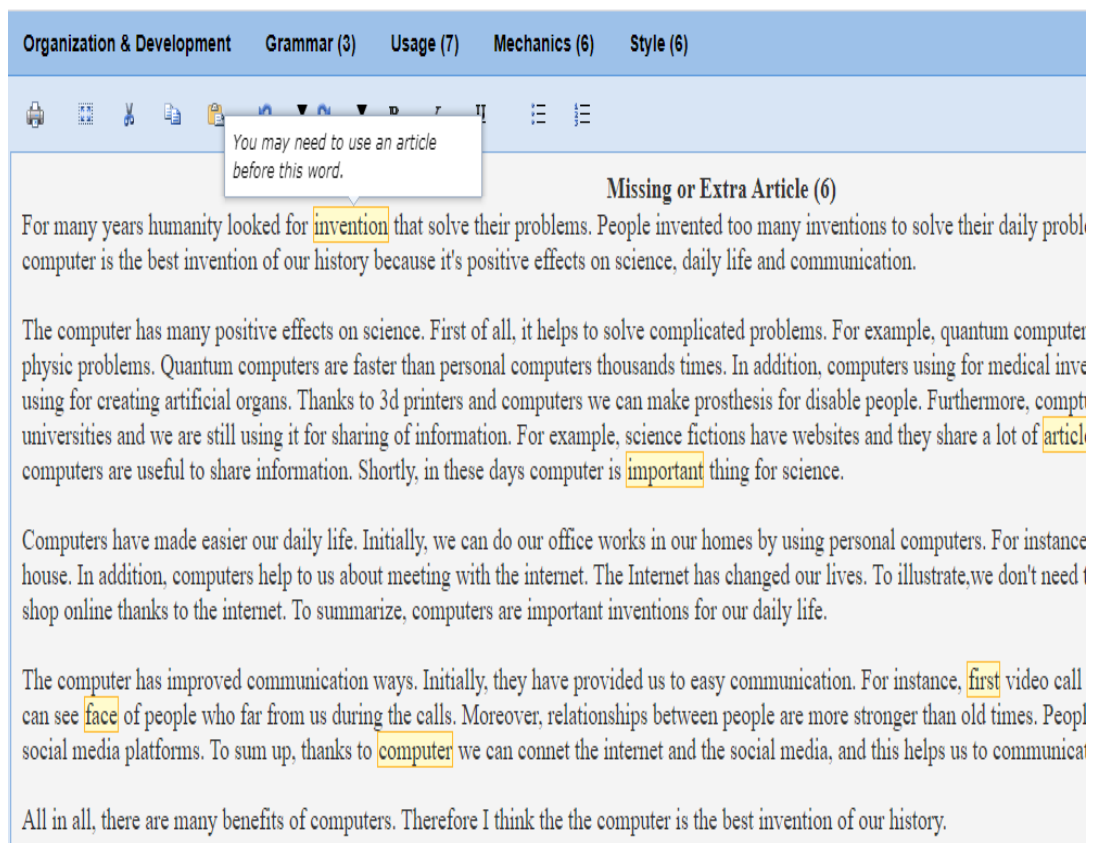


Figure 3.2. Sample feedback given on the AWE system

Being trained on a large corpus of compositions, the system can assess a variety of genres, such as narrative, expository, persuasive, cause & effect, compare & contrast, and descriptive essay genres. Furthermore, Criterion can identify discourse units within an essay which includes the main idea, supporting details, examples, thesis statement, topic sentences, and conclusion parts. In addition to the system's automated feedback, teachers and peers are able to provide feedback for students through the system. Students can view all kinds of feedback by looking into error categories, clicking on subcategories, reviewing highlighted errors within the subcategories, and viewing indirect corrective

feedback concerning the error by moving the cursor over it. In contrast to hand-written essays, students do not need to rewrite the whole essay since their essays are duplicated and displayed side by side so that students can make only the necessary changes or improvements on the essay displayed on the right tab and submit their edited essays afterward.

With regard to instructional purposes, the system enables teachers to use pre-writing and planning features in order to assist students during the stages of the process approach to writing. In addition to its use in writing instruction in school settings, the system can be utilized for self-regulated learning. To encourage learner autonomy, Criterion offers online sources such as writer's handbook and students' personalized portfolios so that pupils can track their own progress. Furthermore, Criterion provides diagnostic feedback in several languages to be accessible to a wide range of students. Moreover, the system grants the opportunity to revise and resubmit their essays ten times. Hence, students can receive immediate feedback after each attempt, and decrease their errors gradually.

The reason why Criterion was chosen to be employed in this study is that research shows that Criterion, when compared to other AWE programs and web-based AWE systems, is one of the most accurate AWE tools regarding feedback and scoring (Zupanc & Bosnic, 2015). Furthermore, in addition to existing prompts, teachers can create their own prompts. Hence, the system allows teachers to assign essays related to the curricular themes. Moreover, the error categories for which Criterion provides feedback are quite similar to error categories that instructors provide error codes for at the school in which the research was conducted. Another reason is that the holistic rubric used in the School of Foreign Languages consists of almost the same number of bands (see Appendix C), as the Criterion holistic rubric (see Appendix D), and both rubrics share similar scoring guides (see an example in Table 3.3). Finally, Criterion can be integrated into a context where process approach to writing instruction is utilized since Criterion provides formative assessment, whereas other AWE programs focus on summative assessment of writing (Dikli, 2006). Hence, the tool used in this study was found suitable to be used in the study context owing to the shared similarities, such as the rationales for the scoring guides. The following Table 3.3 shows the rationales for the highest scores in the AWE holistic rubric and the instructors' holistic rubric.

Table 3.3. *Scoring guides of Criterion and Instructors for the highest scores*

Criterion holistic scoring guide (6)	Instructors' holistic score (18-20)
Develops ideas well and uses specific, relevant details throughout the essay.	Content fulfills the requirement of the task
Is well organized with clear transitions; maintains focus	Ideas are totally relevant, well-developed, supported with relevant examples
Sustains varied sentence structure	Effective organization of ideas / effective use of cohesive devices
Exhibits many specific word choices	Minor grammatical errors- Wide range of appropriate structures
Contains little or no errors in grammar and conventions; errors do not interfere with understanding.	Minor capitalization/ Punctuation errors
Clearly states the position and effectively persuades the reader of the validity of the argument	Very effective use of target vocabulary- Wide range
	Minor spelling mistakes.

As seen in Table 3.3 above, between the highest intervals of both rubrics, there are similarities in terms of rationales given for the scores, and a comparison can be made. Owing to the similarities between the holistic rubrics, the AWE tool was found suitable for the comparison of scores between the AWE tool and the instructors' scores.

In this study, the Criterion was used as a feedback source by the experimental group. However, the viewing scores option was disabled for students since a previous study conducted by Li et al. (2015) indicated that being scored by AWE constantly might demotivate students as they might get frustrated and stressed if they try to get higher scores by numerous attempts. Thus, the scores were only displayed for the researchers and instructors who were involved in the study. Furthermore, students were allowed to use the system if they wanted to write multiple drafts outside the classroom or use the system for other lessons or personal study. In fact, the students requested the instructor to create new prompts for self-study and expressed their willingness to use the system for other purposes many times throughout the study and afterwards.

Moreover, a large body of data were obtained thanks to Criterion as it had been used for multiple purposes, such as providing feedback to the students in experimental groups and scoring randomly chosen 35 essays (five essays from each instructor's class) which belonged to the students from each instructor's class. The essays that the instructors handed to the researcher belonged to the first drafts of the final writing test of the module. The randomly chosen essays from the final writing test were uploaded to the system via the accounts which were used by instructors to develop an understanding of AWE

technology, automated feedback, and automated scoring features in particular. While the students used Criterion for four consecutive weeks including the training week, the instructors used the AWE program for three weeks. Each instructor received the resource books provided by the Criterion tool. The resource books included instructors' beginning guide, students' beginning guide, instructors' manual, students' manual, and other resource books which were easily accessible through user accounts. In addition, the instructors received individual training before they were provided with the subscriptions. This three-week training period aimed to familiarize the teachers with the AWE system by directing them to compare AWE scores and AWE feedback with their scores and feedback. For the instructors to gain insight into the program, both a student account and the teacher account were provided so that they could develop an in-depth insight and understanding of the system and take reflective notes accordingly.

3.3.2. Students' essays

Throughout the study, essays were collected from the experimental and the control groups. The essays which belonged to experimental groups were uploaded to the AWE system to receive diagnostic feedback and observing the change of accuracy from first drafts to second drafts. On the other hand, the essays which belonged to the students in the control group were assessed by the course instructors of the control groups. Subsequently, the errors on both the first drafts and the second drafts were counted by the researcher and another instructor who volunteered to help. Furthermore, apart from the essays of experimental and control groups, five assessed essays from each instructors' classroom were chosen by random sampling method to be uploaded to the AWE program so that instructors could compare their feedback and scores with the automated feedback and automated scores.

As a part of quantitative data, during the research period, students wrote two drafts for each opinion essay about the topics related to curricular themes (technology, inventions and economy). The topics were the same for both the experimental and the control groups, and the length of essays ranged between 250 and 300 words. As students received training on the AWE program and received writing instruction on writing an opinion essay during the fourth week, they wrote their essays on the fifth, sixth, and seventh weeks. For each draft, students were given fifty minutes to compose, revise and edit their essays. Even though all the students had access to their phones during the

lessons, the first and the second essays were written by hand by both the experiment and the control groups since changing the writing environment would be another variable and writing on technological devices might produce more errors, such as spelling errors (White & Ihasz, 2015). Moreover, it would not be possible to know if some errors had been autocorrected by smartphones or not as technological devices are capable of offering word suggestions and corrections (Baron, 2020). Finally, as there were not enough computers in the school, the best solution was to establish similar research environments for both groups. Hence, the essays were transformed into digital files and uploaded to the students' accounts and the students displayed the feedback in the classroom to write their second draft. By doing so, the researcher aimed to ensure that using AWE feedback to improve writing was the only independent variable of the comparative analysis. Nonetheless, the students were allowed to use AWE feedback outside the classroom, and they were allowed to discover the features of AWE outside the classroom.

While the experimental group utilized AWE diagnostic feedback to write their second drafts for each topic, the control group received error codes (see Appendix E). Both groups only received written feedback and teachers did not interfere in the revision process by providing oral feedback as it is also the school regulation to teach students using only error codes and written comments to make necessary changes on their essays. Furthermore, the course instructors of the control groups were informed that they were expected to provide students only with written corrective feedback and written comments for the sake of reliability of the findings. Nonetheless, both groups were provided with whole class feedback regarding organization and content.

3.3.3. Weekly reflection reports

Upon completing their second drafts by utilizing AWE feedback and viewing the feedback provided by the AWE system for their second drafts, the students in the treatment group were asked to write reflection reports structured by four questions (see Appendix F). Two experts in the ELT department were consulted while the questions were formed and the questions mainly focused on:

- students' views of the AWE feedback
- the type of errors that students were able to correct
- the type of errors students have difficulty in correcting
- the challenges that students face while using the AWE program

Weekly written reflection reports were chosen as data collection instruments to unveil students' experiences instantaneously as soon as they complete writing their essay by using AWE feedback so that students' critical reflection could provide information concerning the use of AWE for formative assessment of writing. Considering that students' views might change over time or they could elaborate on what they noted previously, reflection reports were collected each week. Consequently, 54 written reflection reports were included in qualitative data analysis. Due to the fact that all the students were Turkish, reflection questions were written in Turkish. By doing so, the researcher aimed to remove the language barrier and allow students to express themselves clearly. Participants were informed that written reflection reports would be used for research purposes.

3.3.4. Semi-structured interviews with students

In this study, semi-structured interviews were conducted to unfold students' views on the use of AWE for formative assessment and evaluation purposes. The reason why this particular method was chosen to collect data is that semi-structured interviews can create more opportunities for participants to share their experiences, allow further inquiry of justification and clarification, and they provide an extended exploration of the participants' views (Galletta, 2013). Moreover, semi-structured interviews allow researchers to move the interview towards the subject of inquiry to gather more relevant data and they can lead participants to elaborate on what they mention by asking follow-up questions (Flick, 2013). In the present study, the semi-structured interviews complemented written reflection reports by reporting on similar and additional topics as follows:

- the role of AWE in improving writing
- similarities and differences of AWE feedback with teacher's feedback
- advantages and disadvantages of AWE feedback
- challenges
- preferences
- future use and suggestions

The questions of the semi-structured interviews (see Appendix G) were shaped by experts' opinions and altered according to the feedback received from the experts. The interviews were conducted in participants' native language.

3.3.5. Semi-structured interviews with teachers

In addition to students, semi-structured interviews were conducted with instructors to unveil the role AWE might play in formative assessment and evaluation of EFL writing. Semi-structured interviews were utilized as they would allow the researcher to gain an in-depth understanding of participants' views and allow each participant to share their own experiences (Merriam & Tisdell, 2015). Before the semi-structured interviews were carried out, the instructors received training on the use of the AWE tool, and they were provided with the teacher and student account to allow them to explore the features of AWE during the three-week training period. The instructors were asked to take reflective notes regarding their views during this period. The questions of the semi-structured interviews (see Appendix H) were shaped by experts' opinions and altered according to the feedback received from the experts. The questions mainly focused on:

- the role AWE might play in formative assessment
- similarities and differences of AWE feedback with teacher's feedback
- advantages and disadvantages of AWE feedback
- challenges and limitations
- the preferred use of AWE feedback
- future use and suggestions

3.4. Data Collection Procedures

The data for the present study were collected in the fall term of the 2019- 2020 academic year. Before the research was initiated, all the necessary permissions were granted from the Ethical Committee of Anadolu University (see Appendix I). Moreover, a research proposal was sent to Educational Testing Service (ETS), and permission to use Criterion for this particular research was granted. Furthermore, 30 student accounts and one teacher account were provided to be used without any charges during the research period by the ETS support group. Upon receiving permissions to conduct the study, the following steps were taken in order to collect data:

- As most of the students have almost no experience in writing essays, the intermediate-level writing instruction and writing practice were focused on paragraph writing. Hence, the first three weeks were allocated to teaching opinion paragraphs. During this period students were required to write paragraphs that are

quite similar to opinion essays in terms of style and they received constant corrective feedback and written comments regarding their errors.

- During the first week, both the experimental group and the control group received writing instruction for four hours a week. The following week, both groups wrote their first opinion paragraph draft, received written corrective feedback, revised their essays, and wrote the second draft. In a similar fashion, in the third week, students wrote their first and second drafts on another topic after receiving the teacher's feedback. Consequently, all the students wrote opinion paragraphs on three different topics and received written corrective feedback three times. This step was identical for both groups as the students from the experimental group also needed to be exposed to the teacher's feedback so that they could unveil their experiences of both feedback types and provide information regarding the similarities, differences, advantages and disadvantages of these feedback types.
- In the fourth week, the requirements, and stylistic features of the genre (opinion essay) were introduced to both groups. Additionally, the AWE tool was introduced to the experimental group, tutorials provided by ETS were shared with the students and students were trained on using AWE features. Furthermore, students were provided with their usernames and passwords and encouraged to explore the AWE features in their free time.
- In the fifth week, both groups wrote their first opinion essay on the same topic. The control group received error codes and written comments to rewrite and correct their errors, whereas the experimental group utilized AWE feedback to write their second drafts. In the same week, both groups viewed the feedback provided for their second drafts as well. Written reflection questions were distributed to the experimental group as soon as they viewed the AWE feedback provided for their second drafts.
- In order to compare the change in accuracy of the experimental group's and the control group's essays, screenshots of AWE diagnostic feedback were saved and assessment results were recorded. As for the control group, the students' first and second drafts were copied after the feedback was provided and the copies were saved for further examination of error counts. In addition to the essays, written reflection reports were collected each week and they were saved for further investigation.

- In the sixth and the seventh weeks, the same steps taken in the fifth week were repeated. However, each week the topics of opinion essays were changed according to the curricular theme of that week. In this way, the students in the experimental and the control groups wrote two drafts on three different topics and the experimental group received feedback from the AWE tool six times.
- In the eighth week, the final writing test as a part of the final exam was administered. Before the writing test was administered, all the instructors participated in a standardization session conducted by the testing unit of the school.
- All the instructors who participated in the study scored the papers of their own class and provided feedback. In order to be compared with AWE scores and AWE feedback, five random compositions were chosen from the papers, collected, assessed, and evaluated by each instructor. These papers were copied and saved for the final stage of the research and they were uploaded to the AWE system via some students' accounts.
- At the end of the eighth week, students' thoughts about their experiences with the AWE tool were obtained through semi-structured interview questions.
- After all the data were gathered from the students, the researcher informed the instructors about this system and the instructors were trained by demonstrating how the AWE programs works, its capabilities, and its features. Besides, the AWE system tutorials were shared with the instructors. Additionally, instructors were provided with both the instructor account and the students' accounts to which the papers were uploaded. Furthermore, they were informed about the points that the study focuses on, and they were requested to take notes about their observations.
- Teachers were given three weeks to explore AWE features, explore AWE feedback and AWE scores by comparing them with their own feedback and scores provided for the same papers. Moreover, the instructors were allowed to use the system in any way they desired for three weeks. Additional support was provided by the researcher whenever the instructors requested.
- Finally, three weeks later, English Instructors' opinions about the system were unveiled by conducting semi-structured interviews with each of them. The detailed research design and procedures can be seen below in Table 3.4.

Table 3.4. *Procedures of data collection*

	Experimental Group	Control Group
Week 1	Introduction to writing (paragraph level)	Introduction to writing (paragraph level)
Week 2	Opinion paragraph practice - Opinion paragraph 1st draft - Corrective feedback (correction symbols) - Opinion paragraph 2nd draft	Opinion paragraph practice - Opinion paragraph 1st draft - Corrective feedback (correction symbols) - Opinion paragraph 2nd paragraph
Week 3	Opinion paragraph practice - Opinion paragraph 1st draft - Corrective feedback (correction symbols) - Opinion paragraph 2nd draft	Opinion paragraph practice - Opinion paragraph 1st draft - Corrective feedback (correction symbols) - Opinion paragraph 2nd paragraph
Week 4	Introducing the AWE tool Students' Subscription Training Introduction to Opinion Essay	Introduction to Opinion Essay
Week 5	Opinion essay practice - Opinion Essay 1st draft - Criterion Feedback - Opinion Essay 2nd draft -Reflection on AWE Feedback	Opinion essay practice - Opinion Essay 1st draft - Teacher's written feedback (Correction symbols) - Opinion Essay 2nd draft
Week 6	Opinion essay practice - Opinion Essay 1st draft - Criterion Feedback - Opinion Essay 2nd draft -Reflection on AWE Feedback	Opinion essay practice - Opinion Essay 1st draft - Teacher's written feedback (Correction symbols) - Opinion Essay 2nd draft
Week 7	Opinion essay practice - Opinion Essay 1st draft - Criterion Feedback - Opinion Essay 2nd draft -Reflection on AWE Feedback	Opinion essay practice - Opinion Essay 1st draft - Teacher's written feedback (Correction symbols) - Opinion Essay 2nd draft
Week 8	- Reflection on AWE Feedback - Comparing the error reduction and scores - Semi-structured interviews with volunteers	

Table 3.4. (Continued) Procedures of data collection

Week 9	-Introducing the AWE tool and informing the instructors about the general features of AWE -Providing them with teacher and students' accounts
Week 10	-Training instructors on AWE features -Providing the teachers with informative documents -Allowing them to explore AWE features
Week 11	-Viewing how the same essays they gave scores and feedback were assessed by the system -Letting them compare AWE scores and AWE feedback with their own scores and feedback
Week 12	-Conducting semi-structured interviews with the instructors pertaining to their views on AWE

3.5. Data Analysis

For the purpose of answering the first research question (RQ1) regarding the correlation between the AWE scores and the instructors' scores, SPSS (20) software was used, and correlation between instructors' scores and AWE scores given for the same essays were compared with the correlation between each of the seven instructors and the eighth instructor who was referred as the external rater in this study. Although such raters have been commonly referred to as human raters in the literature as most of the studies have been conducted in laboratory settings, in this study, the eighth instructor was referred to as the external rater since this study was classroom-based. The eighth instructor only participated in this study to help determine how applicable and suitable computer-generated scores might be in the particular context since to validate the scores, the essays are assessed by two instructors in the EFL higher education context. Additionally, it is a common way to investigate the correlation between AWE scores and human raters (Cohen, Ben-Simon & Hohav, 2003). To determine whether there was a statistically significant correlation between the instructors' scores and AWE scores, and between the instructors' scores and the external rater's scores, normal distribution analysis "Kolmogorov-Smirnov Normal Distribution" was applied to the variables under study. As a result of the analysis, it was found that the AWE scores did not meet the normality condition ($p < .05$). Therefore, a non-parametric Spearman analysis was utilized for the correlation analysis.

In order to analyze quantitative data for the second research question (RQ2) which aimed to explore the effectiveness of AWE feedback and teacher feedback in error

reduction, and whether there was a significant difference between these two types of feedback in terms of correcting errors, the first and second drafts of essays were treated, respectively, as pre-feedback and post-feedback. The feedback type was considered as the independent variable, and in order to examine the error reduction across drafts, Ferris's model (2006), which focuses on investigating the effectiveness of a feedback type on common errors of was adopted. In Ferris's study, the typical errors in an ESL context were included. As the current study was conducted in an EFL context, instead of including the same error categories used in Ferris's study (2006), the errors of the students taking part in the present study were taken into account. To find the common errors, the errors on the first drafts of each writing task were counted, and the most common errors were found. To ensure inter-coder reliability, the errors on 30% of the essays were categorized and counted by another experienced instructor. Miles and Huberman's (1994) formula $[(\text{consensus}) / (\text{consensus} + \text{disagreement}) \times 100]$ was used. The inter-rater reliability was high (.89) and the discrepancies were resolved between the instructors before the final categorization of errors. The frequency analysis revealed that the most common errors in the grammar category were fragment, run-on sentences, sentence structure errors, and subject-verb agreement errors. In the usage category, the most common errors were the ill-formed verbs, wrong word choice, articles, and prepositions. Finally, the common errors in the mechanics category were spelling, capitalization, and punctuation. Although singular/plural nouns and tense errors were common errors in some writing tasks, these errors weren't included in this study since the AWE tool does not detect or provide errors in these two error categories. All in all, a total of 11 error types were included in the comparative analysis. After the common error types were found, the error counts were normalized by using the formula suggested by Biber, Douglas, Conrad, and Reppen (1998) since essay length was another variation. In order to normalize the errors, the average essay length was measured by counting every word on each student's drafts. The normalized error counts were found by multiplying the error counts for each error type on each student's drafts with the number of errors on essays and dividing the number by the average word count for each writing task (error count x essay length / average word count). The error counts on both drafts of the first, the second, and the third essays were normalized to the average of, respectively, 276, 282, and 295 words.

Once the error counts were normalized, the effect of using AWE on the improvement of student writing in terms of accuracy was analyzed. Shapiro-Wilk,

skewness-kurtosis and histogram graphs were examined to determine if there was a normal distribution. Since the data did not meet the normality assumptions, the Wilcoxon signed-rank analysis was used to determine the error reduction from pre-feedback to post-feedback within each group. Moreover, Mann-Whitney U analysis was used to determine whether there was a significant difference between the groups according to the post-tests.

For the purpose of answering the third research question (RQ3), regarding the students' views on the use of AWE, written reflection reports and the complementary interviews were transcribed word by word by the researcher. Correspondingly, in order to analyze the semi-structured interviews conducted with the instructors to answer the fourth research question (RQ4), examining the instructors' views on the use of AWE tools, the same steps were taken. To analyze the data gathered to answer the third and the fourth questions, Constant Comparison Method (CCM) which has roots in grounded theory research (Glaser & Strauss 1967) was used. CCM enables the researcher to recognize the relationships between serial data (Charmaz, 2006), similar to the weekly reflection reports in this study, and it allows sustaining the connection between subsequent data and complementary data. Hence, this method was used to analyze the students' written reflection reports and all the participants' semi-structured interviews.

The constant comparison method was utilized in order to interpret emerging categories grounded in data which were the participants' statements. Initially, the data were sorted into codes. In this study, a code refers to the words or phrases representing emotions and opinions that reveal participants' views. Subsequently, the codes were compared and contrasted within and between data in order to develop emerging subcategories. Consequently, emerging subcategories were grouped into core categories. In order to decrease the subjectivity of the qualitative analysis process, another experienced rater who holds an MA degree in ELT, and pursues her Ph.D. studies in ELT, was involved in the data analysis. In order to calculate the inter-rater reliability agreement level between the raters, the point-by-point agreement formula $[(\text{the number of agreements}) / (\text{the number of agreements} + \text{the number of disagreements}) \times 100]$ suggested by Tawney and Gast (1984) was used. After 100 percent of the data were analyzed by both raters, the inter-rater reliability between the raters was found as .91. Then, raters had discussions to come to an agreement.

4. RESULTS

4.1. Overview of the Study

This study aimed to investigate the use of automated writing evaluation (AWE) for both summative and formative assessment purposes in the context of EFL writing. Hence, it was aimed to explore if there is a relationship between AWE scores with instructors' scores given for students' essays and to investigate the effectiveness of AWE feedback on error correction in an EFL higher education context. Furthermore, the opinions of students and teachers about the use of AWE in teaching and learning L2 writing were explored. In line with these goals, the following four questions were posed:

RQ1: Is there a relationship between AWE holistic scores and instructors' holistic scores given for students' essays?

RQ2: Is there a significant difference in terms of writing accuracy between a group of students receiving AWE formative feedback with a group of students receiving teacher feedback?

RQ3: How do EFL students view the use of AWE in L2 writing?

RQ4: How do EFL instructors view the use of AWE in L2 writing?

In order to examine the effectiveness of using AWE feedback and teachers' feedback in the L2 writing assessment, and especially in the revision stage, an AWE tool was integrated into the classroom assessment to provide feedback to the 18 students in the experimental group and teacher feedback was provided to a control group with 20 students. In addition to students' use of AWE, seven instructors received training on the use of AWE in writing assessment. The purpose of the training was to familiarize the instructors with the AWE system, provide them with general knowledge about the system, and the system's features. To answer the first question, which aimed to investigate the correlation between the instructors' scores and AWE scores, Spearman's rank correlation test was used. The effectiveness of both feedback types in reducing errors was analyzed with Wilcoxon signed-rank tests, and Mann-Whitney U tests were conducted to see whether there was a significant difference between the feedback types regarding their effect on improving writing accuracy. The data collected to answer the third and the fourth questions, regarding the students' and the teachers' views on AWE, were analyzed qualitatively using the Constant Comparison Method. The results are explained in detail in the following sections.

4.2. RQ1: The Correlation between the Holistic Scores Given for EFL Students'

Essays by the Instructors, and the Correlation between the Instructors Scores and the AWE Scores

In order to answer the first question, in addition to the seven teachers who actively took part in the training process and the interviews, an additional instructor volunteered to score the same essays (n=35) scored by the seven instructors. The eighth instructor who was referred to as the external rater in this study had a similar teaching background and four years of teaching and assessment experience in the context where the study was conducted. The main reason why another instructor was involved in this part of the study was to compare the correlation between instructors' scores and AWE scores with the correlation between instructors' scores and another unbiased rater. In EFL higher education settings, the most common way to ensure reliability and validity in scoring is to achieve inter-rater agreement between two instructors. Thus, by including the external rater, it was aimed to determine how applicable and suitable computer-generated scores might be in this certain context.

In order to eliminate researcher bias, five opinion essays written about the same topic were chosen randomly via random sampling method and handed to the researcher by each instructor (n=7). The essays were written by each instructor's own students and the essays were written during the final writing test which was administered at the end of the final week of the first module. Each instructor was responsible for evaluating the essays written by their own students. Nonetheless, as required by the school policy, all the instructors, (n=8, including the external rater) who volunteered in this part of the study attended a lengthy standardization session to resolve discrepancies of more than 2 points. The standardization session took place before the instructors assessed the essays and after the instructors scored the essays and copies of the randomly chosen essays were handed to the researcher and the eighth instructor. Upon receiving the essays, seven instructors started to score the essays and gave formative feedback to the students in their own writing class. On the other hand, the eighth instructor only scored the essays given by other instructors (n=35) for the purpose of exploring the correlation among the scores. In addition, all the essays were transferred into digital files and uploaded to the AWE system to receive computer-generated scores. Both the instructors and the AWE system scored the essays using holistic rubrics. The instructors used the holistic rubric designed by the testing office, and the AWE system evaluated the essays through a six-point holistic

rubric. Prior to the correlational analysis, the scores given by the instructors and the AWE tool were standardized to range from a minimum of 0 points to a maximum of 100 points.

In the analysis in which the study sought to determine whether there was a statistically significant correlation between the instructors' scores and AWE scores, the normal distribution analysis "Kolmogorov-Smirnov Normal Distribution" was applied to the variables examined. As a result, it was determined that the scores did not meet the normality assumptions ($p < .05$). Although normality transformations were applied to the scores, the distribution could not be normalized. Hence, a nonparametric Spearman analysis was used. The following Table 4.1. summarizes the results of the analysis.

Table 4.1. *The correlation among the scores of instructors, the AWE system, and the external rater*

Variables		Instructors' scores	AWE scores	External rater's scores
Instructors' scores	<i>r</i>	1	.862*	.804*
	<i>p</i>		.000	.000
	<i>N</i>	35	35	35
	$\bar{x}$	70,64		
AWE scores	<i>r</i>	.862*	1	.876*
	<i>p</i>	.000		.000
	<i>N</i>	35	35	35
	$\bar{x}$		59,76	
External rater's scores	<i>r</i>	.804*	.876*	1
	<i>p</i>	.000	.000	
	<i>N</i>	35	35	35
	$\bar{x}$			66,28

* $p < .05$

When the table was examined, it was seen that the instructors' average score was ($\bar{x} = 70,64$), the average AWE score was ($\bar{x} = 59,76$), and the average of the external rater's scores was ($\bar{x} = 66,28$). The correlation coefficient between the instructors' mean scores and AWE mean scores was $r = .862$, and the determination coefficient was $r^2 = .743$. That is, when the AWE scores increased, the instructors' scores increased as well, or when the scores of one variable decreased, the scores of the other variable decreased too. In addition, the correlation coefficient between the instructors' scores and the external rater's scores was $r = .804$, and the determination coefficient was $r^2 = .646$. Consequently, both the correlation between the instructors' scores and AWE scores and the correlation between the instructors' scores and the external rater's scores were statistically significant ($p < .05$). When the interval levels included in the correlation coefficient interpretation criteria presented by Alpar (2014) was taken into account, it was observed that there was

a high, positive, and significant correlation between the instructors' scores and the AWE scores ($r = .862, p < .05$). Moreover, the results put forward that there was a high, positive, and significant correlation between the instructors' scores and the external rater's scores ($r = .804, p < .05$). That is, as the instructors' scores increased or decreased, the AWE scores and the external rater's scores increased or decreased together as well.

In order to investigate the agreement between each instructor's individual scores and AWE scores, and the agreement between the instructors' scores and the external rater's scores, a nonparametric Spearman analysis was used. The results of the analysis can be seen below in Table 4.2.

Table 4.2. *The correlations among AWE scores, external rater's scores and each instructor's scores*

Variables		Instructors' scores	AWE scores	External rater's scores
Teacher 1	<i>r</i>	1	.825	.649
	<i>p</i>		.086	.236
	<i>n</i>	5	5	5
	$\bar{x}$	73	66,4	71
Teacher 2	<i>r</i>	1	.975*	1*
	<i>p</i>		.005	.000
	<i>n</i>	5	5	5
	$\bar{x}$	67,5	59,76	69
Teacher 3	<i>r</i>	1	.949*	.872
	<i>p</i>		.014	.054
	<i>n</i>	5	5	5
	$\bar{x}$	69	63,08	65
Teacher 4	<i>r</i>	1	.975*	1*
	<i>p</i>		.005	.000
	<i>n</i>	5	5	5
	$\bar{x}$	71	56	66
Teacher 5	<i>r</i>	1	.740	.865
	<i>p</i>		.152	.058
	<i>n</i>	5	5	5
	$\bar{x}$	74	56	63
Teacher 6	<i>r</i>	1	.866	.564
	<i>p</i>		.058	.322
	<i>n</i>	5	5	5
	$\bar{x}$	73	59,76	64
Teacher 7	<i>r</i>	1	.918*	.803
	<i>p</i>		.028	.102
	<i>n</i>	5	5	5
	$\bar{x}$	67	56,44	65

* $p < .05$

As seen in Table 4.2 above, the discrepancy between the average of instructors' scores and the AWE scores was higher than the discrepancy between the average of the instructors' scores and the external rater's scores. This might have occurred because the rubrics used by the instructors ranged from 0 to 20 and the rubric used by the system ranged from 0 to 6. Although the AWE scores were seemingly lower than both the instructors' scores and the external rater's scores, it was found that there was a high, positive, and significant correlation between the AWE scores and the second instructor ($r = .975, p < .05$), the third instructor ($r = .949, p < .05$), the fourth instructor ($r = .975, p < .05$) and the seventh instructor ($r = .918, p < .05$). This shows that both external rater's scores and AWE scores were positively and significantly correlated with other raters, and in some cases, the correlation between AWE scores and other raters was higher.

It was also seen that there were high, positive, and significant correlations between the external rater's scores and, respectively, the second instructor ($r = .1, p < .01$) and the fourth instructor ($r = .1, p < .01$). Hence, the correlations between the instructors' scores and the external rater's scores were higher than the correlations between the instructors' scores and the AWE scores in some cases. Nonetheless, it can be seen in Table 4.2 that, in general, when the correlation between the AWE scores and the instructors' scores was higher, the correlation between the instructors' scores and the external rater's scores was higher as well.

In addition to the quantitative analysis, during the third and the final week of training and observation of the system period, which took place to familiarize the instructors with the system, instructors were asked to use the student accounts they were provided with and review the feedback and scores given by the AWE system to the same essays they scored and provided feedback for. Furthermore, the instructors were asked to take reflective notes of their views on the AWE scores and feedback. During the interviews, the instructors were asked to state their views on the AWE scores and to compare the AWE scores with their scores. As a result of the interviews, it was found that six teachers found AWE scores similar to their scores, and only one instructor found the AWE scores lower than her own scores. The excerpts below indicate the instructors' perceptions of AWE scores:

T1: " The scores are more or less the same with the students' levels. One of these students was especially less proficient than the other and I gave the lowest score to that student. As

the student got the lowest score by the system and it was roughly the same band. The system also gave the others an average score like mine. I guess it just gave one student a lower score than I thought, because of grammar mistakes.”

T2: “ The system gave scores in parallel with the scores I gave. I can say that it gave similar scores and provided similar explanations with the criteria we use, and even almost the same scores, so it was similar.”

T3: “For example, the 60-80 range for me corresponds to the range of 4. Although there is a wide range, we have given scores at the same intervals. When I look at the comments provided under the scores and when I evaluate the papers, I agree with the problems and deficiencies mentioned by the system for the essays, both in terms of grammar and content. Hence, I think it evaluates the essays similarly and I find it reliable.”

T4: “The scores given by the system were similar. Actually, they were very similar, so I am satisfied with the scoring, but the students may not be able to envision how much it coincides with our own rubric.”

As seen in the excerpts above, the majority of the instructors found the AWE scores similar to their own scores, and they expressed that the AWE system similarly justified the given score with the holistic rubric used by the instructors. Here are a sample score and explanations for the certain score provided by the system (see Figure 4.1)

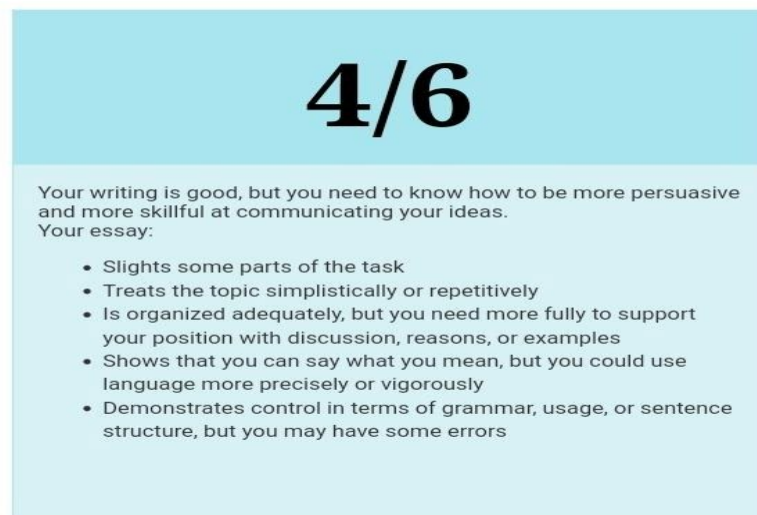


Figure 4.1. *A sample score and explanations provided by the AWE tool for the certain score*

To conclude the findings regarding the correlations among instructors’ scores, AWE scores and the external rater’s scores put forward that AWE scores were positively and significantly correlated with the scores provided by the instructors and the external rater. Moreover, it was found that the inter-rater agreement between two instructors and the inter-rater agreement between an instructor and an AWE tool can be similar.

Moreover, the teachers noted that there were similarities between their assessment and the AWE assessment in terms of the similarity of the scores, the score intervals and the rationales for the scores in each interval.

4.3. RQ2: The Change in EFL Students' Writing Accuracy After the Provision of AWE Formative Feedback and Teachers' Formative Feedback

In the first phase of the study, both the students in the experimental group (n=18) and the students in the control group (n=20) received writing instruction on writing opinion paragraphs for three weeks and both groups were provided with teachers' formative feedback. In addition, they were trained on using the correction symbols guide to correct their errors. During the fourth week, both groups received instruction on the structure of the opinion essay and the students in the experimental group were trained on using the AWE system, and to make sure students could make use of AWE formative feedback, the terms used by the system were clarified. The materials used to familiarize the students with the system (see Appendix J) were provided by the Educational Testing Service (ETS) after permission for the study was granted and the existing manuals and guidebooks on the system were used for further familiarization.

Following the introduction of the AWE tool and an introduction to writing opinion essays, both groups wrote first drafts and second drafts, which were also the final drafts, of an opinion essay during the fifth week, sixth week, and seventh week. The students wrote their essays on a different topic each week, and they were asked to write more than 250 words. In other words, both groups wrote two drafts for each of the three different topics and they received feedback for each draft, and both groups wrote each draft of the essays on the same days and under the same conditions. Hence, the only difference between the groups was the type of feedback they received to correct their errors as the control group received the teacher's formative feedback, whereas the experimental group received AWE formative feedback to correct their errors and write their second drafts.

In order to determine the error types to be included in comparative analysis, first drafts of all three writing assignments were examined and the errors on students' first drafts on each topic were counted by two raters. After the common errors made by the students were found as a result of frequency analysis, experienced teachers working in the study context were consulted on whether they agree with the types of errors found to be frequent at the intermediate level according to the study, and if these errors were

commonly made by the students in the study context based on their observations. Consequently, it was also expressed by the experienced teachers that the frequent error types found as a result of analysis were the common errors made by students in the study context in their opinions as well. Table 4.3. illustrates the findings related to common errors made by the participants.

Table 4.3. *Frequent error types in students' first drafts*

Error Types	The First Essay			The Second Essay			The Third Essay		
	N*	Mean	%	N*	Mean	%	N*	M	%
Punctuation	150	3,94	13,2	214	5,63	17,4	170	4,47	16,4
Article	209	5,5	18,4	112	2,94	9,1	105	2,76	10,1
Spelling	92	2,42	8,1	102	2,68	8,3	74	1,94	7,1
Ill-formed verbs	88	2,31	7,7	105	2,76	8,6	76	2	7,3
Preposition	77	2,02	6,8	85	2,23	6,9	60	1,57	5,8
Subject-verb agreement	69	1,81	6	61	1,60	5	65	1,71	6,3
Run-on sentences	63	1,65	5,5	57	1,50	4,6	71	1,86	6,9
Singular/Plural noun	55	1,44	4,8	104	2,73	8,5	78	2,05	7,5
Wrong word choice	52	1,36	4,6	68	1,78	5,5	84	2,21	8,1
Capitalization	48	1,26	4,2	64	1,68	5,2	27	0,71	2,6
Sentence structure	45	1,18	4	35	0,92	2,9	51	1,34	4,9
Fragment	43	1,13	3,8	46	1,21	3,7	57	1,5	5,5
Wrong Tense	22	,57	1,9	51	1,34	4,1	16	0,42	1,5
Missing noun or verb	19	0,50	1,7	26	0,68	2,1	18	0,47	1,7
Wrong word order	19	0,50	1,7	17	0,44	1,4	12	0,32	1,2
Unnecessary word	14	0,36	1,2	14	0,36	1,1	18	0,47	1,7
Wrong word form	14	0,36	1,2	10	0,26	0,8	9	0,23	0,9
Quantifier-noun agreement	11	0,28	1	10	0,26	0,8	10	0,26	1
Pronoun	10	0,26	0,9	10	0,26	0,8	9	0,23	0,9
Countable/Uncountable	10	0,26	0,9	18	0,47	1,5	8	0,21	0,8
Faulty comparison	9	0,23	0,8	10	0,27	0,8	9	0,23	0,9
Compound words	8	0,21	0,7	5	0,13	0,4	5	0,13	0,5
Possessive Errors	6	0,15	0,5	4	0,10	0,3	2	0,05	0,2
Negation	4	0,10	0,4	2	0,05	0,2	2	0,05	0,2
Total	1137		100	1230		100	1036		100

As analyzing the changes in all the error types would be beyond the scope of this study, the error types with a mean score of more than 1 were considered as the most frequent errors, and accordingly, 11 most frequent error types were included in the further comparative analyses between the groups. As a result, the common errors to be included in the study were determined as fragment, run-on sentence, sentence structure errors, subject-verb agreement errors (SVA), ill-formed verbs, wrong word choice, missing and extra articles, preposition, spelling, capitalization and punctuation errors. Although the results showed that singular/ plural noun errors and tense errors were common errors with mean scores of more than 1, these error types were not included in the comparative analysis as the AWE system does not provide corrective feedback on these errors. Hence,

comparative analyses were conducted to examine the change in these 11 categories.

Once the errors on the first and the second drafts were counted, the error counts were normalized based on the average word count of students' essays. The error counts on both drafts of the first, the second, and the third essays were normalized to the average of, respectively, 276, 282, and 295 words. Afterward, it was examined whether the data met the normality assumptions. Shapiro-Wilk, skewness-kurtosis, and histogram graphs were examined to determine if there was a normal distribution. Nonetheless, it was found that the data did not meet the normality assumptions. Therefore, Wilcoxon signed-rank analysis was used to determine the change in pre-feedback and post-feedback error counts, and Mann-Whitney U analysis was used to determine whether there was a significant difference between the groups according to the post-tests. Data were analyzed using the SPSS 20 statistical package program. The following parts summarize the weekly findings of Wilcoxon signed-rank analyses and Mann-Whitney U analyses in detail.

4.3.1. The change in students' writing accuracy from the first drafts to the second drafts of the first essay

The students in the experimental and control groups wrote their first essays in the fifth week. Although it was assumed that the students had a similar level of writing proficiency since they were all intermediate-level students, the similarities between the groups in terms of their writing accuracy were also analyzed according to the pre-feedback error counts of the students in the experimental and control groups. According to the comparative analysis results, it was determined that there was only one significant difference in the "spelling" error counts between groups before the provision of AWE feedback. The findings obtained as a result of the analysis are presented in Table 4.4.

Table 4.4. *The first essay- comparison of pre-feedback error counts between groups*

Errors types (pre-feedback)	Experimental (N=18)		Control (N=20)		U	P
	Mean Ranks	Sum of Ranks	Mean Ranks	Sum of Ranks		
Fragment	18,08	325,50	20,78	415,50	154,50	,455
Run-on	20,00	360,00	19,05	381,00	171,00	,792
RW	20,31	365,50	18,78	375,50	165,50	,670
SVA	20,06	361,00	19,00	380,00	170,00	,770
Ill-formed verb	20,28	365,00	18,80	376,00	166,00	,682
Wrong word	20,83	375,00	18,30	366,00	156,00	,481
Article	22,56	406,00	16,75	335,00	125,00	,108
Preposition	20,86	375,50	18,28	365,50	15,50	,474

Table 4.4. (Continued) *The first essay- comparison of pre-feedback error counts between groups*

Spelling	23,44	422,00	15,95	319,00	109,00	,038*
Capitalization	22,72	409,00	16,60	332,00	122,00	,087
Punctuation	18,56	334,00	20,35	407,00	163,00	,619

* Statistically significant ($p < .05$)

As seen in Table 4.4, according to the Mann-Whitney U analysis, spelling errors showed a significant difference in favor of the experimental group ($U = 109.50$, $p < .05$). In other words, there was no significant difference between the groups in all other variables (except for the spelling error) examined in the context of the first essays ($p > .05$). This situation reveals that before the students in the experimental group were exposed to AWE feedback, the students in the experimental and control groups were similar to each other in terms of writing accuracy except for spelling errors.

In order to determine the accuracy change depending on the feedback type, the Wilcoxon signed-ranks test was conducted to find out whether there was a significant difference in the pre-feedback and post-feedback error counts of the students in the experimental and control groups, and the results are presented in Table 4.5.

Table 4.5. *Comparison of pre-feedback and post-feedback error counts of students' first essays*

The First Essay		Experimental group					Control group				
		N	Mean Rank	Sum of Ranks	Z	p	N	Mean Rank	Sum of Ranks	z	P
Fragment (pre-post)	Negative Ranks	8	4,50	36,00	-2,521	,012*	12	7,50	90,00	-3,111	,002*
	Positive Ranks	0	,00	,00			1	1,00	1,00		
	Ties	10					7				
Run-on (pre-post)	Negative Ranks	15	8,00	120,00	-3,408	,001*	14	8,50	119,00	-3,352	,001*
	Positive Ranks	0	,00	,00			1	1,00	1,00		
	Ties	3					5				
Sentence structure (pre-post)	Negative Ranks	12	6,50	78,00	-3,059	,002*	11	6,00	66,00	-2,937	,003*
	Positive Ranks	0	,00	,00			0	,00	,00		
	Ties	6					9				
SVA (pre-post)	Negative Ranks	15	8,00	120,00	-3,408	,001*	18	9,50	171,00	-3,724	,000*
	Positive Ranks	0	,00	,00			0	,00	,00		
	Ties	3					2				
Ill-formed verb (pre-post)	Negative Ranks	17	9,00	153,00	-3,621	,000*	19	10,00	190,00	-3,824	,000*
	Positive Ranks	0	,00	,00			0	,00	,00		
	Ties	1					1				
Wrong word (pre-post)	Negative Ranks	9	5,00	45,00	-2,666	,008*	10	5,50	55,00	-2,803	,005*
	Positive Ranks	0	,00	,00			0	,00	,00		
	Ties	9					10				
Article (pre-post)	Negative Ranks	17	9,00	153,00	-3,621	,000*	8	4,50	36,00	-2,521	,012*
	Positive Ranks	0	,00	,00			0	,00	,00		
	Ties	1					12				

Table 4.5. (Continued) *Comparison of pre-feedback and post-feedback error counts of the students' first essays*

Preposition (pre-post)	Negative Ranks	14	7,50	105,00		14	7,50	105,00			
	Positive Ranks	0	,00	,00	-3,296	,001*	0	,00	,00	-3,297	,001*
	Ties	4					6				
Spelling (pre-post)	Negative Ranks	17	9,00	153,00		16	8,50	136,00			
	Positive Ranks	0	,00	,00	-3,621	,000*	0	,00	,00	-3,517	,000*
	Ties	1					4				
Capitaliza- tion (pre-post)	Negative Ranks	14	7,50	105,00		13	7,00	91,00			
	Positive Ranks	0	,00	,00	-3,296	,001*	0	,00	,00	-3,182	,001*
	Ties	4					7				
Punctua- tion (pre-post)	Negative Ranks	18	9,50	171,00		19	10,79	205,00			
	Positive Ranks	0	,00	,00	-3,724	,000*	1	5,00	5,00	-3,733	,000*
	Ties	0					0				

Based on negative ranks

* Statistically significant ($p < .05$)

As seen in Table 4.5, the revisions made depending on the type of feedback showed a statistically significant difference in the context of the error types examined in both groups ($p < .05$ for all variables). In both groups, the error counts were in favor of negative ranks which indicate a reduction in errors, and both feedback types provided to the groups significantly reduced the error counts in each category.

In addition to within groups comparison, with the Mann-Whitney U analysis, it was examined whether the type of feedback given to the first drafts of the initial essay showed a significant difference in the post-feedback scores of the experimental and control groups. The results are presented in Table 4.6.

Table 4.6. *Comparison of post-feedback error counts of students' first essays between groups*

The First Essay (post-feedback)	Experimental (N=18)		Control (N=20)			
	Mean Ranks	Sum of Ranks	Mean Ranks	Sum of Ranks	U	P
Fragment	20,22	364,00	18,85	377,00	167,00	,693
Run-on	18,00	324,00	20,85	417,00	153,00	,371
Sentence structure	18,94	341,00	20,00	400,00	170,00	,730
Subject-verb agreement	17,89	322,00	20,95	419,00	151,00	,210
Ill-formed verbs	15,89	286,00	22,75	455,00	115,00	,044*
Wrong word choice	19,61	353,00	19,40	388,00	178,00	,951
Article	9,78	176,00	28,25	565,00	5,00	,000*
Preposition	18,75	337,50	20,18	403,50	166,50	,686
Spelling	18,39	331,00	20,50	410,00	160,00	,465
Capitalization	18,39	331,00	20,50	410,00	160,00	,357
Punctuation	13,06	235,00	25,30	506,00	64,00	,000*

* Statistically significant ($p < .05$)

According to the results of the comparative analysis, the error types that reduced significantly were ill-formed verbs, ($U = 115.00$, $p < .05$), article ($U = 5.00$, $p < .05$) and

punctuation errors ($U = 64.00$, $p < .05$). In the post-feedback results, in terms of the ill-formed verbs, article, and punctuation error counts, the average of the error counts of the students in the experimental group decreased significantly, while the mean scores of the error counts of these error types did not decrease significantly in the control group.

These results reveal that the feedback given by the AWE system in the experimental group was more effective in correcting the ill-formed verbs, articles, and punctuation errors compared to the feedback given by the teacher in the control group. In other words, in the comparison made over the post-feedback error counts of the groups, there was a significant difference between the two groups in terms of the reduction in the ill-formed verbs, articles, and punctuation errors. Sample AWE feedback and teachers' feedback provided for ill-formed verbs, articles, and punctuation errors made by the students in this study are displayed below in Table 4.7.

Table 4.7. *Sample feedback provided by teachers and the AWE system*

Error Type	Sample errors from students' essays	Teacher's feedback	Sample AWE feedback
Ill-formed verbs	They always want to following the	WF	This verb may be incorrect. Proofread the sentence to make sure you have used the correct form.
Missing or Extra Article	First reason why internet is beneficial for people is ... On the contrary, you can meet new person , and ...	^ or X	You may need to use an article before this word. Consider using the article the You may need to use an article before this word. Consider using the article a
Punctuation	Money brings happiness but reach people ... There are three reasons to believe that , internet is the best invention.	P	You may need to place a comma after this word You may need to remove this comma

As shown in Table 4.7, while the teacher feedback is provided with error codes, the AWE system provided longer and detailed explanations compared to teachers' feedback.

In addition, Table 4.7 shows that the AWE tool also provided hints about how to correct certain types of errors.

4.3.2. The change in students' writing accuracy from the first drafts to the second drafts of the second essay

In order to determine if there was a significant difference between the groups in terms of writing accuracy, the pre-feedback error counts of the students in the experimental and control groups were compared. According to the comparative analysis results, it was determined that there was only one significant difference in the "spelling" factor between groups before the provision of AWE feedback.

The results showed that there was a significant difference between the groups in terms of spelling errors ($U = 105.50$, $p < .05$) and capitalization errors ($U = 90.50$, $p < .05$). However, there was no significant difference between the groups in all other variables (except for the spelling and capitalization errors) examined in the context of the second essay ($p > .05$). This situation reveals that before the students in the experimental group were exposed to AWE feedback, the students in the experimental and control groups were similar to each other in terms of writing accuracy. The findings obtained as a result of the analysis are presented in Table 4.8.

Table 4.8. *The second essay-comparison of pre-feedback error counts between groups*

Errors types (pre-feedback)	Experimental (N=18)		Control (N=20)			
	Mean Ranks	Sum of Ranks	Mean Ranks	Sum of Ranks	U	P
Fragment	16,81	302,50	21,93	438,00	131,50	,158
Run-on	17,19	309,50	21,58	431,50	138,50	,228
RW	17,56	316,00	21,25	425,00	145,00	,317
SVA	17,36	312,00	21,43	428,00	141,50	,264
Ill-formed verb	17,56	316,00	21,25	425,00	145,00	,317
Wrong word	19,56	352,00	19,45	389,00	179,00	,988
Article	17,11	308,00	21,65	433,00	137,00	,217
Preposition	16,78	302,00	21,95	439,00	131,00	,158
Spelling	15,36	276,50	23,23	464,00	105,50	,028*
Capitalization	14,53	261,00	23,98	479,00	90,50	,008*
Punctuation	16,14	290,00	22,53	450,00	119,50	,077

* Statistically significant ($p < .05$)

In order to examine the change in writing accuracy depending on the type of feedback, the Wilcoxon signed-ranks test was conducted to determine whether there was a significant reduction from the pre-feedback to post-feedback error counts of the students in the experimental and control groups, and the results are presented in Table 4.9.

Table 4.9. *Comparison of pre-feedback and post- feedback error counts of students' second essays*

The Second Essay		Experimental group					Control group				
		N	Mean Rank	Sum of Ranks	Z	p	N	Mean Rank	Sum of Ranks	Z	P
Fragment (pre-post)	Negative Ranks	12	7,13	85,50	-2,796	,005*	12	6,50	78,00	-3,064	,002*
	Positive Ranks	1	5,50	5,50			0	,00	,00		
	Ties	5					8				
Run-on (pre-post)	Negative Ranks	11	6,00	66,00	-2,936	,003*	15	8,00	120,00	-3,411	,001*
	Positive Ranks	0	,00	,00			0	,00	,00		
	Ties	7					5				
Sentence structure (pre-post)	Negative Ranks	6	3,50	21,00	-2,207	,027*	7	4,00	28,00	-2,366	,018*
	Positive Ranks	0	,00	,00			0	,00	,00		
	Ties	12					13				
SVA (pre-post)	Negative Ranks	13	7,00	91,00	-3,180	,001*	15	8,00	120,00	-3,408	,001*
	Positive Ranks	0	,00	,00			0	,00	,00		
	Ties	5					5				
Ill-formed verbs (pre-post)	Negative Ranks	13	7,00	91,00	-3,180	,001*	19	10,00	190,00	-3,823	,000*
	Positive Ranks	0	,00	,00			0	,00	,00		
	Ties	5					1				
Wrong word (pre-post)	Negative Ranks	14	7,50	105,00	-3,296	,001*	11	6,36	70,00	-2,432	,015*
	Positive Ranks	0	,00	,00			1	8,00	8,00		
	Ties	4					8				
Article (pre-post)	Negative Ranks	17	9,00	153,00	-3,621	,000*	17	9,00	153,00	-3,622	,000*
	Positive Ranks	0	,00	,00			0	,00	,00		
	Ties	1					3				
Preposi- tion (pre-post)	Negative Ranks	15	8,00	120,00	-3,409	,001*	19	10,00	190,00	-3,824	,000*
	Positive Ranks	0	,00	,00			0	,00	,00		
	Ties	3					1				
Spelling (pre-post)	Negative Ranks	13	7,00	91,00	-3,180	,001*	18	10,50	189,00	-3,783	,000*
	Positive Ranks	0	,00	,00			1	1,00	1,00		
	Ties	5					1				
Capitaliza- tion (pre-post)	Negative Ranks	11	7,00	77,00	-2,981	,003*	16	8,50	136,00	-3,516	,000*
	Positive Ranks	1	1,00	1,00			0	,00	,00		
	Ties	6					4				
Punctua- tion (pre-post)	Negative Ranks	18	9,50	171,00	-3,724	,000*	20	10,50	210,00	-3,921	,000*
	Positive Ranks	0	,00	,00			0	,00	,00		
	Ties	0					0				

Based on negative ranks

* Statistically significant ($p < .05$)

As shown in Table 4.9, both types of feedback led to a statistically significant reduction in the context of all variables (error types) examined in the experimental and control groups ($p < .05$ for all variables). The change in error counts in both groups were in favor of negative ranks that is the reduction in error counts.

In addition to the within-groups comparison, post-feedback error reduction rates were examined by the Mann-Whitney U analysis to determine whether the type of feedback given in the context of the typical errors resulted in a significant difference in the post-feedback error counts of the experimental and control groups. The results obtained were as presented in Table 4.10.

Table 4.10. *Comparison of post- feedback error counts of students' second essays between groups*

The Second Essay (post-feedback)	Experimental (N=18)		Control (N=20)		U	P
	Mean Ranks	Sum of Ranks	Mean Ranks	Sum of Ranks		
Fragment	19,17	345,00	19,80	396,00	174,00	,849
Run-on	18,75	337,50	20,18	403,50	166,50	,641
Sentence structure	18,50	333,00	20,40	408,00	162,00	,534
SVA	17,67	318,00	21,15	423,00	147,00	,196
Ill-formed verbs	19,06	343,00	19,90	398,00	172,00	,798
Wrong word	17,44	314,00	21,35	427,00	143,00	,261
Article	10,67	192,00	27,45	549,00	21,00	,000*
Preposition	20,11	362,00	18,95	379,00	169,00	,734
Spelling	18,22	328,00	20,65	413,00	157,00	,385
Capitalization	18,94	341,00	20,00	400,00	170,00	,619
Punctuation	10,72	193,00	27,40	548,00	22,00	,000*

* Statistically significant ($p < .05$)

According to the comparative analysis of the errors on the second drafts, it was determined that there was a significant difference only in the article ($U = 21.00$, $p < .05$) and the punctuation ($U = 22.00$, $p < .05$) errors between the groups. When the article and punctuation error counts on students' second drafts were analyzed, it was be seen that the mean error counts of the students in the experimental group decreased significantly, while the mean scores of the students in the control group did not decrease significantly. These results show that the feedback given by the system in the experimental group was more effective in correcting the article and punctuation errors compared to the feedback given by the teacher in the control group. In the comparison made on the post-feedback accuracy of the groups, apart from the article and punctuation errors, there was no significant difference in reduction of error types on students' second drafts of the second essay.

4.3.3. The change in students' writing accuracy from the first drafts to second drafts of the third essay

To be able to investigate if there was a significant difference between the groups in terms of writing accuracy, the pre-feedback error counts of the students in the experimental and control groups were compared. According to the comparative analysis results, it was determined that there was no significant difference ($p > .05$) between groups before the provision of AWE feedback. The findings obtained as a result of the analysis are presented in Table 4.11.

Table 4.11. *The third essay- comparison of pre-feedback error counts between groups*

Errors types (pre-feedback)	Experimental (N=18)		Control (N=20)		U	P
	Mean Ranks	Sum of Ranks	Mean Ranks	Sum of Ranks		
Fragment	20,25	364,50	18,83	376,50	166,50	,696
Run-on	22,94	413,00	16,40	328,00	118,00	,072
RW	21,44	386,00	17,75	355,00	145,00	,317
SVA	20,47	368,00	18,63	372,50	162,50	,613
Ill-formed verb	20,83	375,00	18,30	366,00	156,00	,496
Wrong word	20,47	368,00	18,63	372,50	162,50	,613
Article	20,39	367,00	18,70	374,00	164,00	,654
Preposition	18,31	329,00	20,58	411,50	158,50	,534
Spelling	20,19	363,50	18,88	377,50	167,50	,718
Capitalization	21,39	385,00	17,80	356,00	146,00	,331
Punctuation	19,56	352,00	19,45	389,00	179,00	,988

* Statistically significant ($p < .05$)

To determine the change in writing accuracy depending on the feedback type, the Wilcoxon signed-ranks test was conducted to determine whether there was a significant difference in the pre-feedback and post-feedback writing accuracy of the students in the experimental and control groups, and the results are summarized in Table 4.12.

Table 4.12. *Comparison of pre-feedback and post-feedback error counts of students' third essays*

The Third Essay (pre-post)		Experimental group					Control group				
		N	Mean Rank	Sum of Ranks	Z	p	N	Mean Rank	Sum of Ranks	Z	P
Fragment (pre-post)	Negative Ranks	13	7,00	91,00	-3,181	,001*	13	7,00	91,00	-3,181	,001*
	Positive Ranks	0	,00	,00			0	,00	,00		
	Ties	5					7				
Run-on (pre-post)	Negative Ranks	16	8,50	136,00	-3,516	,000*	15	8,00	120,00	-3,408	,001*
	Positive Ranks	0	,00	,00			0	,00	,00		
	Ties	2					5				
Sentence structure (pre-post)	Negative Ranks	11	6,00	66,00	-2,936	,003*	9	5,00	45,00	-2,668	,008*
	Positive Ranks	0	,00	,00			0	,00	,00		
	Ties	7					11				
SVA (pre-post)	Negative Ranks	12	6,50	78,00	-3,059	,002*	16	8,50	136,00	-3,516	,000*
	Positive Ranks	0	,00	,00			0	,00	,00		
	Ties	6					4				
Ill-formed verbs (pre-post)	Negative Ranks	14	7,50	105,00	-3,296	,001*	13	7,00	91,00	-3,180	,001*
	Positive Ranks	0	,00	,00			0	,00	,00		
	Ties	4					7				
Wrong word (pre-post)	Negative Ranks	14	7,50	105,00	-3,296	,001*	14	7,50	105,00	-3,296	,001*
	Positive Ranks	0	,00	,00			0	,00	,00		
	Ties	4					6				

Table 4.12. (Continued) Comparison of pre-feedback and post- feedback error counts of students' third essays

Article (pre-post)	Negative Ranks	16	8,50	136,00			10	6,40	64,00		
	Positive Ranks	0	,00	,00	-3,517	,000*	1	2,00	2,00	-2,758	,006*
	Ties	2					9				
Preposi- tion (pre-post)	Negative Ranks	11	6,00	66,00			12	6,50	78,00		
	Positive Ranks	0	,00	,00	-2,934	,003*	0	,00	,00	-3,059	,002*
	Ties	6					8				
Spelling (pre-post)	Negative Ranks	15	8,00	120,00			16	8,50	136,00		
	Positive Ranks	0	,00	,00	-3,408	,001*	0	,00	,00	-3,517	,000*
	Ties	3					4				
Capitali- zation (pre-post)	Negative Ranks	10	5,50	55,00			9	5,00	45,00		
	Positive Ranks	0	,00	,00	-2,805	,005*	0	,00	,00	-2,666	,008*
	Ties	8					11				
Punctua- tion (pre-post)	Negative Ranks	16	8,50	136,00			20	10,50	210,00		
	Positive Ranks	0	,00	,00	-3,516	,000*	0	,00	,00	-3,920	,000*
	Ties	2					0				

* Statistically significant ($p < .05$)

As illustrated in Table 4.12, both types of feedback led to a statistically significant difference on all variables (11 error types) examined in the experimental and control groups ($p < .05$ for all variables). The difference in both groups was in favor of negative ranks, which indicates that both feedback types resulted in a significant reduction in error counts, and improved writing accuracy. Furthermore, to examine if there was a significant difference between the feedback types considering the error reduction, the post-feedback writing accuracy was examined by the Mann-Whitney U analysis to find out whether the type of feedback given in the context of error correction shows a significant difference in the writing accuracy of the second drafts of the groups. The results obtained are presented in Table 4.13.

Table 4.13. Comparison of post- feedback error counts of students' third essays between groups

The Third Essay (post-test)	Experimental (N=18)		Control (N=20)		U	P
	Mean Ranks	Sum of Ranks	Mean Ranks	Sum of Ranks		
Fragment	20,42	367,50	18,68	373,50	163,50	,618
Run-on	17,86	321,50	20,98	419,50	150,50	,319
Sentence structure	19,22	346,00	19,75	395,00	175,00	,863
SVA	18,08	325,50	20,78	415,50	154,50	,352
Ill-formed verbs	17,06	307,00	21,70	434,00	136,00	,152
Wrong word	19,06	343,00	19,90	398,00	172,00	,809
Article	12,33	222,00	25,95	519,00	51,00	,000*
Preposition	17,97	305,50	19,88	397,50	152,50	,578
Spelling	18,25	328,50	20,63	412,50	157,50	,356
Capitalization	19,56	352,00	19,45	389,00	179,00	,956
Punctuation	15,42	277,50	23,18	463,50	106,50	,027*

* Statistically significant ($p < .05$)

According to the results given in Table 4.13, it was determined that there was only a significant difference in the article ($U = 51.00$, $p < .05$) and punctuation ($U = 106.50$, $p < .05$) errors between the groups after different types of feedback was received by the experimental and the control groups. In the second drafts, based on the article and punctuation error counts, the mean error counts of the students in the experimental group decreased significantly, while the mean scores of the students in the control group did not reduce significantly.

These results reveal that the feedback given by the AWE system in the experimental group was more effective in correcting the article and punctuation errors than the feedback given by the teacher in the control group. Apart from the article and punctuation errors, in the comparison made on the post-feedback writing accuracy of the groups, there was no significant difference due to the type of feedback.

In addition to quantitative data, in order to explore whether students' opinions align with the quantitative findings regarding the change in writing accuracy resulting from the use of AWE feedback, qualitative data were collected through weekly reflection reports and semi-structured interviews were carried out with the experimental group. Weekly reflection reports were collected every week after the students utilized AWE feedback to write their second drafts and reviewed the AWE feedback given for their second drafts, and the interviews were conducted at the end of the study. Both data collection instruments asked students to reflect on the errors they were able to correct, and they were asked to specify the types of errors that they were able to correct thanks to the use of AWE feedback.

Qualitative analysis of the data revealed the main error categories and the specific error types which students stated they were able to correct easily by using AWE feedback. While some students mentioned the names of the main error categories in the weekly reflection reports and during the interviews, some of them identified the specific error types they corrected the most by the use of AWE feedback. As a result, a total of 171 error types and categories were identified. Among these, 110 belonged to error types specifically named by the students, whereas 61 of them belonged to main error categories, respectively, grammar, mechanics, and usage categories. Table 4.14. summarizes the findings of the analysis.

Table 4.14. *The error categories and the error types students stated that they were able to correct easily*

Main Error Categories	Error types	N*	%
Grammar	Ill-formed verbs	11	6.4
	Subject-verb agreement errors	9	5.3
	Fragment errors	4	2.3
	Sentence structure errors	3	1.8
	Others*	31	18.1
Usage	Articles	26	15.2
	Wrong word choice	7	4.1
	Prepositions	3	1.8
	Others*	9	5.3
Mechanics	Punctuation	24	14
	Spelling	16	9.3
	Capitalization	7	4.1
	Others*	21	12.3
Total		171	100

N* = Number of the codes

Others* = The main error categories

As seen in Table 4.14, according to the students, some of the frequently mentioned error types that they were able to correct easily were article errors (n=26), punctuation errors (n=24), spelling errors (n=16), ill-formed verbs (n=11), subject-verb agreement errors (n=9), wrong word choice errors (n=7) and capitalization errors (n=7). Furthermore, it can be seen in Table 4.12 that according to students, they were able to correct the errors in the mechanics category more compared to the other main error categories. Thus, complementary qualitative data analysis supports the results of the quantitative data analyses since the result of the quantitative data analyses revealed that in addition to being effective in reducing errors in grammar, usage, and mechanics categories, the AWE feedback was more effective in correcting the article and punctuation errors than the feedback given by the teacher to the control group.

All in all, both quantitative and qualitative data gathered in order to explore the effect of AWE feedback on correcting errors in L2 writing put forward that both AWE feedback and teacher feedback were effective in reducing the 11 error types included in this study significantly. Nonetheless, the AWE feedback was more effective in reducing usage and mechanics errors compared to the teacher's formative feedback. According to the students, although the AWE feedback helped to identify a variety of error types, most of the errors they believed they were able to correct were in the usage and the mechanics categories. This result in general yielded that AWE scoring was found as effective as

teacher scoring in all categories of error correction and it further helped for identifying more errors in usage and mechanics while writing in L2.

4.4. RQ3: EFL Students' Views on the Use of AWE in Learning and Assessment of Writing in L2

In order to elicit the experimental group students' opinions about the use of AWE feedback, weekly reflection reports were collected each week from 18 students who participated in the study regularly, and semi-structured interviews were conducted with 13 volunteers at the end of the study. The analysis of the collected qualitative data were performed by using the Constant Comparison Method. As a result of the analysis, 922 codes were identified regarding students' opinions which were formed under five main categories. The main categories that emerged from the codes were *advantageous features of the AWE system* (224 codes), *improving writing performance by the use of AWE feedback* (324 codes), *positive effects of AWE tool on gaining autonomy* (115 codes), *comparison of AWE feedback and teachers' feedback* (157 codes), *drawbacks of using the AWE tool* (102 codes). The main categories related to students' views on the use of AWE in L2 writing can be seen in Table 4.15.

Table 4.15. *The main categories related to students' views on the use of AWE in L2 writing*

Main categories	N*
Advantageous features	224
Improvement in writing performance	324
Gains in autonomy	115
Comparison of AWE feedback and teachers' feedback	157
Drawbacks	102
Total	922

N*: Number of codes

4.4.1. Advantageous features of the AWE system

The first main category emerging from students' statements was advantageous features of the AWE system. The first main category comprised 224 codes in total and three subcategories. A large number of respondents indicated that AWE systems are advantageous due to *the technical advantages of the AWE system* (n=146), *quality of AWE feedback* (n=36), and *quantity of AWE feedback* (n=42). Table 4.16 displays the subcategories related to the advantageous features of the AWE system.

Table 4.16. *The subcategories related to the advantageous features of the AWE system*

Advantageous features of AWE system	Explanatory statements	N*
Advantageous features are...		
technical advantages	<i>You can use the system regardless of time and place</i>	146
quality of feedback	<i>It provides detailed and explanatory feedback</i>	36
quantity of feedback		42
Total	<i>It finds all the errors and evaluates every aspect of writing</i>	224

N*: Number of codes

As seen in table 4.16, one of the recurrent sub-categories emerging from participants' opinions was the technical advantages of the AWE system (n=146). With regard to the advantages of using an AWE tool, the majority of the students mentioned that the system can be used on different devices, such as smartphones and computers, and it can be used at any time. Hence, the availability of the system was a prominent reason why participants viewed using the AWE system is advantageous. Correspondingly, using the AWE system was described as a practical way to receive feedback and correct errors by most students. The reason why the AWE system was found practical by the students was the fact that even though they wrote their drafts on paper at school when they practice writing at home the system copies the first draft and allows users to make changes to their first draft as opposed to writing everything once again. The way the system allows users to revise and resubmit their essays is shown in Figure 4.2.

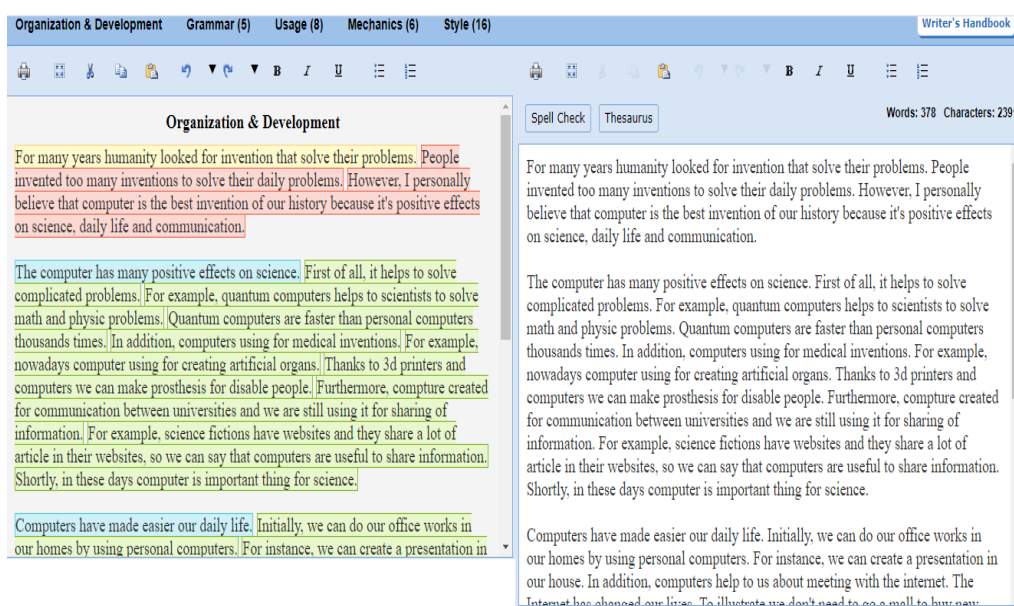


Figure 4.2. *A sample view of the revision box of the AWE tool*

As seen in Figure 4.2, the system allowed users to revise their drafts and make necessary changes on a copy of the draft on the same page. Thus, writing on the AWE system could be viewed as practical and time-saving by the participants. Moreover, participants pointed out that receiving immediate feedback saved their time and allowed them to revise and resubmit their essays a couple of times. Therefore, receiving immediate feedback saved students' time and encouraged them to write more than two drafts. In accordance with being able to write more, participants pointed out that as the system allowed them to write up to ten drafts and receive immediate feedback upon writing each draft, they were able to practice more than they did in a classroom setting. Moreover, the AWE system's user-friendly interface and its ability to store students' all essays were the other technical advantages of the AWE system according to the participants. In addition, some participants noted that as the AWE system uses algorithms to detect errors and provide feedback, they trust the system. Concerning their trust in the system, participants pointed out that as an automated system can't get tired, busy, and can't overlook errors, they believe that the system can find a large number of errors. The excerpts below show students' opinions regarding the technical advantages of using AWE feedback:

S1: "I would like to use AWE feedback to correct my errors because, as I said, it gives immediate feedback and the tool is easy to use and very accessible."

S4: "The strengths of the system are that it keeps records of errors and it keeps track of everything."

S5: "I think the system can't overlook errors because it has an established system. On the other hand, teachers might be fatigued while assessing writing. Additionally, most of the time, when we write in handwriting, we might write illegibly. Teachers may not have noticed this, or they may not understand. That's why we may not get a chance to correct our errors if the errors are not identified. However, a machine would identify them thanks to its system."

S8: "Since writing frequently is important for the development of my writing skills, we can get feedback more easily thanks to this system, we do not have to wait, so we can write more. We don't have to wait for the teacher."

As can be interpreted from the excerpts above, the students pointed out many technical advantages that have been brought about by the use of the AWE tool. Thus, many participants stated that using the AWE system is a convenient way to write and receive feedback thanks to the easy-to-use features of the system.

Another sub-category related to advantageous features of AWE feedback was *the quality of AWE feedback* (n=36). The students found AWE feedback clear, and they pointed out AWE system provides detailed explanations and gives direct or indirect hints

which may enable students to correct their errors easily. The excerpts below show the students' opinions about the quality of AWE feedback:

S9: "When I click on the errors, the feedback was more explanatory for me as it shows the categories on the top and hints appear in the feedback, so I corrected more errors."

S10: "I liked the system, and it helped me to reduce my errors because it explained what the mistakes are in a simple way. It did not only showed the name of the error type, but also included clues about how to fix it, so it was useful for us to correct our errors."

As can be understood from the excerpts, the students pointed out that they found AWE feedback easy to understand and assisting since it had a different system compared to correction symbols used by teachers to give corrective feedback. AWE feedback explains why errors occur in students' essays and most AWE feedback includes hints to help students correct their errors. Thus, it can be deduced that in terms of corrective feedback, AWE feedback might be considered as a more direct type of feedback compared to correction symbols.

In addition to the quality of feedback, the *quantity of feedback* (n=42) provided by the AWE system was a recurrent subcategory. The participants noted they were pleased with the amount of feedback given by the system and they expressed that the abundant quantity of feedback provided by the system helped them to improve writing accuracy on their second drafts. The following excerpts exemplify students' views:

S1: "It shows all the errors, and it tells you exactly what category the error falls into. Since it finds all the errors even if it is missing comma, you can see all of the errors on the system"

S6: "It shows most of the errors. There is nothing it does not show. For instance, I used to forget all the articles. It shows them a lot."

S8: "There are many categories and error types shown on the feedback box."

The excerpts above show that the students were pleased with the number of errors identified on the system and the fact that they were able to see the number of errors made under each category. Furthermore, it can be understood that students were in favor of being notified of all their errors. Hence, it might be inferred that students may have preferred receiving feedback on a variety of writing aspects.

4.4.2. The effects of AWE feedback on improving writing performance

The second main category emerging from students' statements was the effect of AWE feedback on improving writing performance. In addition, the three sub-categories in relation to the positive effect of AWE feedback on improving L2 writing were identified as the *allowance of AWE feedback to reduce errors after the revision of*

feedback (n=249), *to increase students' attention to avoid the common errors while writing* (n=21) and *to promote learning through feedback* (n=54). The subcategories related to the effect of AWE on improving writing performance can be seen in Table 4.17.

Table 4.17. *The subcategories related to the effect of AWE on improving writing performance*

AWE feedback improves writing performance	Explanatory statements	N*
AWE feedback improves writing performance thanks to its allowance to ...		
reduce errors after revision of feedback	<i>I corrected most of my errors</i>	249
increase attention to avoid common errors	<i>I realized the frequent errors and tried not to make those errors</i>	21
promote learning through feedback	<i>I learned that nouns are followed by articles</i>	54
Total		324

N*: Number of codes

As can be seen in Table 4.17, one of the sub-categories was the allowance of AWE feedback to *reduce errors after the revision of feedback* (n=249). The majority of the students pointed out that they were able to correct their errors thanks to the feedback provided by the system. Students were able to notice a reduction in their errors with the help of feedback provided by the AWE system. The excerpts below show students' opinions about reducing their errors by using AWE feedback:

S5: "When I wrote the second draft in the initial weeks, it was more difficult to correct all the errors and I still had a few errors. However, on the second draft of the final essay, I was able to correct all the errors on my first attempt. Currently, no errors appear in any way. That's why I feel like I improved myself in writing."

S9: "Since all the feedbacks are very explanatory, I was able to understand my errors and correct them very easily. Therefore, it was easier for me to correct my mistakes compared to the past. For example, I was able to fix all the grammar errors. Apart from that, I was able to correct all the mechanical errors as well."

As can be understood from the excerpts above, the students found AWE feedback helpful to reduce their errors and write their second drafts with the help of feedback provided. In addition, a large number of students noted in their weekly reflection reports that reducing their errors has gotten easier over time, and they were able to correct more errors, especially in the final week of the AWE use in writing. As shown in the excerpts above, students attributed the reason for error reduction to the detailed and qualified feedback provided by the AWE system. That is, students were able to understand their

errors and tried to overcome them in time. Hence, it can be inferred that exposure to AWE feedback led to improved writing accuracy.

Another recurrent subcategory was that AWE feedback increased students' attention to avoid the common errors (n=21). In addition to the reduction in their errors in general, students were able to notice the recurring errors in their writing which led to an awareness of the nature of common errors they made. In other words, students were able to realize the frequent errors they made while writing, and they paid attention to avoid making the same errors while writing their second drafts and other essays. The excerpts below show how AWE feedback led students to avoid repeating errors:

S2: "I pay much more attention to the things I used to forget while writing, or the errors I used to make frequently."

S4: "Thanks to the feedback I received, now I know what kind of errors will occur in my writing, and I try not to make those errors."

S10: "I think it helped me to improve my writing skills because my errors decreased over time, and I paid more attention to them as I knew what my common errors were. I wanted to make fewer mistakes while writing. Thanks to the system, I don't make the same errors while writing anymore."

As seen in excerpts, being informed about the common errors they make while writing in L2 resulted in reducing the number of errors that occurred frequently in their essays, and they tended to avoid these kinds of errors to improve their L2 writing proficiency. Thus, it can be concluded that AWE feedback not only helped students to reduce their errors from the first draft to the second draft but also allowed them to reduce the errors that might appear on other essays.

Finally, it was found that according to the students, AWE feedback *promotes learning* (n=54). The students' statements revealed that they might have reached beyond completing the given task since the students expressed that in addition to correcting their errors, they learned how to improve writing accuracy through the feedback provided by the AWE tool. That is, they only corrected their errors but the feedback they received also resulted in learning gains in terms of language use. The excerpts below exemplify students' beliefs regarding this issue:

S1: "It makes me see even the minor errors. Although they are minor, they become a major issue when I repeat them. For example, in this way, I learned what comes in front of nouns or words, and I use what I learned in every sentence since the system showed me how to use them."

S2: “I have probably revised the feedback given by the system about a thousand times, I used to have a problem in constructing sentences, and I was structuring sentences much differently than I do now. I have learned how to construct my sentences properly by revising the feedback there many times.”

S4:” Even though I have used the system for a short period of time, I have learned a lot. In other words, the system gives feedback in a certain order that makes me understand better what is wrong.”

As seen from the excerpts above based on the students’ statements, it can be inferred that students believed that they enhanced their knowledge of various aspects of writing, such as grammar, mechanics, sentence structuring while revising their papers and attempting to correct their errors.

4.4.3. Positive effects of the AWE tool on gaining autonomy

Another main category emerging from the students’ statements was the positive effect of the AWE tool on gaining autonomy (n=115). Accordingly, four subcategories were formed regarding the relationship between the AWE tool and gaining autonomy. The sub-categories were that the *AWE tool raises awareness of strength and weaknesses of students with regard to writing in L2* (n=17), *it fosters independence from the teacher* (n=68), *it increases self-correction opportunities* (n=21), and *it increases motivation* (n=9). The subcategories of gaining autonomy thanks to the AWE tool can be seen in Table 4.18.

Table 4.18. *The subcategories related to the positive effects of the AWE tool on gaining autonomy*

Positive effects of AWE tool on gaining autonomy are	Explanatory statements	N*
AWE tool increases autonomy by resulting in...		
self-awareness of strengths and weaknesses in L2 writing	<i>I realized that need to improve my grammar</i>	17
independence from teacher	<i>We don't need to wait for a teacher to give feedback</i>	68
increased self-correction opportunities	<i>I noticed the areas of improvement and studied accordingly</i>	21
increased motivation	<i>Receiving feedback quickly and easily makes you want to practice</i>	9
Total		115

N*: Number of codes

The first sub-category related to gaining autonomy thanks to the AWE tool was that *the system raises learners’ self-awareness of their strengths and weaknesses* (n=17). The students pointed out that being able to see the errors grouped under certain categories

S1: “Based on the feedback I received, I realized that I mostly make grammar errors because I have received more feedback on the grammar category. Other categories seem almost immaculate. It means that I am better in some aspects of writing, and I need to focus more on some areas, such as grammar. Hence, the feedback is useful in that sense.”

All Feedback

Trait Feedback Analysis

Trait	Count
Common: 1 error	1
Usage: 3 errors	3
Mechanics: 7 errors	7
Style: 4 errors	4

Average Number of Words per Sentence: 15.2
 Total Number of Sentences: 17
 Unique Words Count: 63

All Feedback

Mechanics

Mechanics	Count
Spelling	0
Capital Punctuation	0
Missing Punctuation Mark	0
Missing Quotation Mark	0
Missing First Letter	0
Missing Initial Capital Letter	0
Missing Apostrophe	0
Missing Comma	6
Hyphen Error	0
Fixed Word	0
Compound Word	0
Quotation	0
Extra Comma	1

Average Number of Words per Sentence: 15.2
 Total Number of Sentences: 17
 Unique Words Count: 63
 Total Number of Words: 258

Furthermore, it was revealed that according to the students, using the AWE tool *fosters independence from the teacher* (n=68). As most of the writing practice took place at school guided by the teachers, students were mostly dependent on the feedback

provided by the teachers to improve their writing ability. However, students pointed out that thanks to the AWE tool, they were able to practice outside the school, and they started to feel less dependent on their teacher to practice improving their writing in L2. The excerpts below explain how the students benefited from the AWE tool to become more independent:

S1: "In addition to writing essays, I would also use the system for my hobbies. For instance, I like to write stories in English, but since it is something I do outside the school, I can't ask my teachers to correct my errors in my stories. However, I will be able to use the system whenever I need to write in English, so I can also use it for other purposes."

S7: "When the program gives feedback, you can easily understand why you made an error thanks to the highlighted errors and clear explanations, so I do not feel like I need to consult my teacher to correct the errors."

S9: "As I said, it became more practical as we could rewrite at home on the system because normally we feel like we have to ask the teachers and talk about how to correct errors, but now it is more comfortable to correct errors on my own because the explanations appear on the system."

S11: "Thanks to easy access to feedback, we were not depending on the time we spend in the classroom. We were also able to work at home, so we were able to do everything on our own."

As can be understood from the excerpts above, although students used to feel as if receiving feedback from the teacher was the only way to correct their errors, they were able to self-regulate their own learning process. In addition, the students expressed that they used to need further verbal explanations to correct their errors. However, they started to feel more self-sufficient and less dependent on the time spent in the classroom to correct their errors. Moreover, students were able to receive feedback regardless of time and place boundaries; hence, they had a chance to receive feedback on their other writings apart from the ones they were supposed to write as part of the academic assessment. Teachers need to give feedback to many students in a limited time, and according to students' views AWE system created a space for individual work in which students can correct their errors on their own at their own pace. Thus, it can be inferred that using supplementary tools, such as the AWE tool outside the classroom can allow learners to be actively involved in their own learning process.

Another subcategory related to gaining autonomy was that the feedback *increased their self-correction opportunities* (n=21). The participants stated that the AWE feedback urged them to find out how to correct their error by spending time and effort as the AWE

tool was not a direct provider of correction. The excerpts below show students' opinions about being led to develop strategies by the AWE system to improve writing accuracy:

S9: "Because the system shows the errors but does not correct them automatically, I try to find information on how to correct those errors. I think it is memorable when I search for information, or it becomes easier for me to correct them."

S13: "It was easy to see how many mistakes I had, and which categories the errors fall into, so it helped me search for information to correct the errors I have made frequently. Additionally, it was easier to focus on certain issues and study accordingly."

As can be seen from the excerpts above, since the students were guided by the system to search for and gather information to increase their writing accuracy, it can be concluded that AWE feedback might boost students' learning efforts. Therefore, rather than expecting everything from the teacher, students took the initiation to be in charge of their own learning and become more responsible for their own learning. Self-correction opportunities were welcomed by the students since when they searched for information on their own to improve their L2 writing, learning became more memorable for them which strengthened the links for new learnings.

The fourth and final subcategory was that the AWE tool *increases students' self-motivation* to write more and receive more feedback. The reason why motivation is a subcategory under autonomy is that while some believe that motivation can be an outcome of autonomy (Deci & Ryan, 1985; Dickinson, 1995), many also believe that motivation is essential to initiate self-regulated learning and it promotes learning effort (Spratt, Humphreys & Chan, 2002; Vandergrift, 2005). Hence, it can be said that motivation and autonomy are interconnected factors that influence learning. Regarding the link between motivation and autonomy in this study, the majority of the students were pleased with receiving immediate feedback and being able to write up to ten drafts. Therefore, some students expressed that they were motivated to use the system to receive formative feedback. In fact, it was observed that even though students were asked to utilize the AWE feedback while revising their first and the second drafts of each essay, many students kept revising and submitting revised drafts to correct all the errors appearing on their essays. The following excerpt is an example of students' statements regarding this matter:

S12: "I think such websites motivate people to write more because it is both easy to access, and we can get receive feedback quickly."

As visible from the quote above, AWE feedback fostered an intrinsic type of motivation to write more effectively in the target language. That is, rather than writing to take a grade or to be promoted by the teacher as extrinsic sources of motivation, students used feedback provided by the AWE system to enhance their own learning in L2 writing.

4.4.4. Comparison of the AWE feedback and teachers' feedback

When students were asked to reflect on their experiences of using teacher's feedback and AWE feedback to improve their writing in L2, 157 codes related to the comparison of AWE feedback and teachers' feedback were identified. The sub-categories regarding the comparison of AWE feedback and teachers' feedback involved *the similarities between the two types of feedback* (n=38), *the differences between the feedback types* (n= 88), and *preferred use of feedback* (n=31). The total number of codes included in this main category and sub-categories are presented in Table 4.19.

Table 4.19. *The subcategories related to the comparison of AWE feedback and teachers' feedback*

Comparison of AWE feedback and teachers' feedback	Explanatory statements	N*
The comparison of feedback types showed that the..		
similarities of AWE feedback and teachers' feedback	<i>Both types of feedback are indirect and focused on similar error types</i>	38
differences of AWE feedback and teachers' feedback	<i>AWE feedback is instant, teacher feedback is delayed</i>	88
the preferred use of feedback in L2 writing	<i>We can use AWE feedback in general, but teachers can comment on our ideas</i>	31
Total		157

N*: Number of codes

As a result of the qualitative data analysis, it was revealed that there were some *similarities between teachers' feedback and AWE feedback* (n=38). One of the most prominent similarities between these feedback types was that both the AWE feedback and the feedback given by the teachers focused on similar error types, especially the error types under the grammar, usage, and mechanics categories. In addition, the participants stated that they were familiar with the error types as the error types were similar to the error types included in the correction symbols guide used by the teachers. Moreover, the students pointed out that both types of feedback guided students to correct their errors on their own instead of correcting the errors directly. Thus, it can be inferred that both teachers' feedback and AWE feedback can be considered as indirect feedback types. Additionally, some of the students stated that both types of feedback helped them to reduce their errors. In other words, according to the students, both AWE feedback and

teachers' feedback were helpful in terms of improving writing accuracy. The following excerpt can exemplify students' opinions concerning the similarities between the two feedback types:

S10: "I think everything is the same because the teacher shows the errors without correcting the errors, and the system shows what the errors are and where they are without correcting them. Both of them help us to reduce our errors by showing our errors and checking how much we have improved when we correct the errors, so both of them are useful."

On the other hand, students highlighted that there are some differences between these two feedback types (n= 88) even though they share some similarities. Some of the differences frequently mentioned by students were that they can only receive teacher feedback in the school environment. For this reason, it was stated that the process of writing essays and receiving feedback was limited to the lesson hours. In addition, the students stated that their teachers might have difficulty in providing feedback for the essays they write outside of the classroom due to their workload, and it would not be possible to ask to receive further guidance from the teachers during the time allocated to the writing lessons in the classroom due to the number of students in the class and the duration of the class. In addition, students stated that writing the first and second drafts in the classroom was quite time-consuming compared to writing on the AWE tool. Thus, the AWE feedback and teacher's feedback are different due to some challenges and constraints both students and teachers encounter in the school environment. The excerpts below show the differences between these two feedback types:

S4: "You can get instant feedback on the system, but since there are many people in the classroom, teachers cannot always give us instant feedback because we make a lot of errors, and they have to mark even the slightest errors."

S11: "When we use teachers' feedback we need to wait for a while because we only see the teachers at school, but we can use the system whenever we want."

In addition to differences, such as the availability and practicality of both types of feedback, some other differences were also pointed out regarding the explanations provided to help students correct their errors. While the majority of the students stated that the feedback given by the AWE system under most categories were more explanatory than correction symbols used by the teachers, the feedback given under the organization and development category was found insufficient and rather generic compared to teacher's feedback. The reason why the AWE feedback on the organization and development category was found less specific than the teachers' feedback is that the AWE

tool shows reminders and directs students to the writers' handbook to review the writing guide instead of showing the errors made under this category. Furthermore, the students expressed that teachers can guide them better than AWE feedback in terms of content, mainly due to the lack of content-related feedback on the AWE system. Thus, according to the students, teachers can provide more personalized feedback in areas, such as organization and content. Therefore, it might be inferred that although the students found the AWE feedback more explanatory in general, it was revealed that they would appreciate being given feedback indicating the problems regarding coherence as well. The excerpts below exemplify students' views:

S2: "I think the only difference is that teachers can inform us if the examples we write are relevant, or they can give feedback on meaning and ideas."

S4: "The teacher can assess the writing in terms of meaning, organization, and content, but the system doesn't show problems with content."

Students also commented on the type of feedback they would like to use to improve their writing skills and how they would like to benefit from AWE feedback or teachers' feedback. It was revealed that *the preferred way to use feedback* (n=31) was a mixed use of AWE feedback and teachers' feedback. Although some students preferred to receive teacher's feedback to use in the future, a majority of the students stated that they would like to either continue using AWE feedback, or they would prefer to receive both teachers' feedback and AWE feedback. Furthermore, when students were asked about how they would like to benefit from these two types of feedback, the majority of the students expressed their willingness to continue using the AWE tool in the absence of a writing instructor, especially outside the school for self-regulated learning purposes, and to be able to practice writing more. Apart from the use of the AWE tool for writing essays, all the students highlighted that throughout their university education they will need to use a tool similar to the AWE tool, especially for making presentations, writing articles, and submitting assignments. Hence, the AWE system is preferred by the students in general as complementary to teacher feedback rather than a replacement for it.

4.4.5. Drawbacks of using the AWE tool

As a result of qualitative data analysis, it was revealed that the drawbacks of using the AWE tool (n= 102) was another main category emerging from students' statements. In relation to the drawbacks of using the AWE tool, two sub-categories were found within the main category. These sub-categories were *difficulties in correcting certain errors*

(n=74), and *system-related issues* (n=28). These sub-categories are displayed below in Table 4.20.

Table 4.20. *The subcategories related to the drawbacks of using the AWE tool*

Drawbacks of using the AWE tool	Explanatory statements	N*
Some drawbacks of the system are ...		
difficulties in correcting certain errors	<i>I couldn't change some of the repeated words</i>	74
system-related issues	<i>The system detected a website name as an error</i>	28
Total		102

N*: Number of codes

Students expressed their concerns about the error types they had *difficulty in correcting* while writing their second draft. It was unveiled that there were certain errors that were rather challenging to correct compared to others. According to the students, they had difficulty in correcting the error types which fall within the style category. Primarily, the error types the student had difficulty in correcting were the repetition of words and short sentences. With regard to the challenges students faced while attempting to correct repetitive words, some students pointed out that as the AWE tool is a machine, it might have disregarded the topic they wrote essays about, and while certain words seemed essential for them to use in an essay, those words might have been detected as repetitive words by the system. The following excerpts are instances of that:

S2: "I did not encounter any problems while using the system, but as I said it shows too many repetition errors. I don't know if it could understand, or how it could understand, but it was a problem."

S6: "I couldn't correct some style errors. For example, let's say the topic is about money and happiness. I have to use the words money and happiness a lot while explaining it. Then a warning appears on the system and it says you have repeated the same word too many times. I couldn't correct all of them."

In addition to the repetition errors, many students pointed out that they had difficulty in correcting the errors that fall into short sentence error types. It was also stated that although students attempted to connect short sentences with connectors, and they tried to write complex sentences upon receiving feedback, it was rather burdensome to correct the particular error type, and they had difficulty in extending the length of sentences. The excerpts below should exemplify students' statements concerning this issue:

S5: "In general, I can't prevent myself from writing short sentences. I connected most of them with connectors, but still, it was harder to correct them."

S7: "The system sometimes says that a sentence is short, but I couldn't find what else to write for some of those sentences."

As can be seen from the excerpts above, compared to the errors in other categories, such as grammar, usage, and mechanics, students found the errors appearing within the style category and short sentence error type relatively harder to correct. The main reason for this might be students' level since students might be writing shorter sentences in lower levels and may not be able to comprehend the significance of writing complex and compound sentences or they might be in need of further guidance and input regarding this issue. Thus, it can be inferred that receiving AWE feedback in isolation may not be sufficient for correcting certain errors, or students may need further writing instruction to improve certain aspects of writing.

In addition to difficulties in correcting particular errors, it was pointed out by the students that they experienced some system-related issues (n=28). To illustrate, although the analysis of the weekly reflection reports and the interviews revealed that students haven't experienced any major technical issues and they found the interface user-friendly, some students noted that the interface seemed complicated the very first week and they have gotten used to receiving AWE feedback by time. Hence, it can be concluded that although students went through a training week before they started to use AWE feedback to correct their errors, a lengthy training period might help students to familiarize themselves with technological tools. In addition, some students expressed that the system wasn't able to identify a few proper names and detected them as errors. This might indicate that certain proper names, such as personal names and city names may not exist in the corpus used by the system. The following excerpts exemplify students' statements:

S4: "I wrote the name of a pharaoh, and the system thought that it was a spelling error."

S11: "I think the system doesn't know some names. For example, I wrote PUBG (an online game) when I was giving an example, and it appeared as an error."

In conclusion, the students' views on the use of AWE feedback were mainly positive, and the students were content with the use of AWE feedback to improve writing accuracy in general. The results revealed that according to the students, there are advantageous features of the AWE tool that made it attractive and convenient to be used in stages of writing in L2. To illustrate, most students perceived the AWE tool as useful, practical, easily accessible, and they were pleased with the features included in the

system. In addition to the advantageous features, the students expressed that AWE was also beneficial since it helped them to improve their writing ability. In other words, the students were convinced that the AWE feedback led to improving writing accuracy and reducing the errors they made in L2 writing. Furthermore, the students' statements revealed that they found AWE helpful in terms of gaining autonomy. To put it differently, it was put forward by the students that the AWE diagnostic feedback and trait feedback analysis accompanying it raised their awareness of their strengths and weaknesses. Thanks to the easy access to the system and the numerous opportunities provided by the system for students to practice writing and receive feedback outside the school without a teacher's supervision allowed them to gain independence and increased their learning effort in addition to increasing their motivation to improve their writing accuracy. In addition, comparing the implementation of AWE and their previous experiences with teacher feedback, the students' highlighted that there were some similarities and differences between the AWE assessment and teachers' assessment. It was pointed out that the AWE feedback was similar to teacher feedback as both feedback types included formative corrective feedback, and they both contributed to writing accuracy. It was also revealed that although AWE feedback might be more explanatory in terms of the formative, corrective feedback, the feedback given on content and organization needs to be more guiding and personalized. Hence, some students expressed their willingness to receive a mix of both feedback types or to be assessed concerning the content and organization aspects of their writing. Despite the advantages of using AWE in L2 writing and the similarities between the two types of assessments, some drawbacks and concerns were pointed out. Thus, when implementing AWE in EFL writing, all these points must be taken into consideration.

4.5. RQ4: EFL Instructors' Views on the Use of AWE in Teaching and Assessment of Writing in L2

At the beginning of the data collection period, a three-week training period was included to make sure that the teachers had enough information about the system. The focus of each week of the training process was different, and the instructors were asked to take reflective notes each week while making observations and examinations on the system regarding the focal points of the research. In the first week, the instructors were given a tutorial introducing the system, and they were provided with both students' and

teachers' accounts. In the second week, teachers were provided with a brief tutorial session including the main features of the system, and they were provided with informative documents. Throughout the second week, the instructors observed the features of the system with the help of the training they received and the documents provided (see Appendix K). In the third week, teachers were asked to view how the same essays they gave scores and feedback for were assessed by the system, and they were asked to view their feedback and scores as well as the AWE feedback and automated scores.

When the three-week training and observation process was over, semi-structured interviews were conducted with the teachers. The teachers were asked to answer the questions and express their feelings, thoughts, and opinions about the system. The interviews were analyzed qualitatively using the Constant Comparison Method. As a result of qualitative data analysis, a total of 503 codes were identified and these codes were formed under five main categories. The main categories emerging from the codes were the **challenges faced by teachers in teaching writing** (85 codes), **advantages of using the AWE tool** (177 codes), **disadvantages of the AWE tool** (63 codes), **comparison of AWE assessment and teachers' assessment** (95 codes) and **implementing AWE in EFL writing** (83 codes). The main categories are displayed in Table 4.21 below.

Table 4.21. *The main categories emerging from the instructors' statements and viewpoints on the use of AWE in teaching and learning writing*

Main categories	N*
The challenges faced by teachers in teaching writing	85
Advantages of using the AWE tool	177
Disadvantages of the AWE tool	63
Comparison of AWE assessment and teachers' assessment	95
Implementing AWE in EFL writing	83
Total	503

N*= Number of the codes

4.5.1. The challenges faced by teachers in teaching writing

During the semi-structured interviews, when the instructors were asked to reflect on the AWE tool and the AWE feedback, they also touched upon their own experiences of teaching writing. The first main category, the challenges faced by teachers in teaching writing (n=85), consisted of two subcategories which are the challenges faced by teachers

during *the practice stage* (n=27) and *the feedback stage* (n=58). A common view among the instructors was that teaching writing is challenging and the main challenges occur during practice and feedback stages. The subcategories in relation to the challenges faced by the teachers can be seen below in Table 4.22.

Table 4.22. *The subcategories related to the challenges faced by teachers in teaching writing*

The challenges faced by teachers in teaching writing	Explanatory statements	N*
There were challenges faced by teachers during the...		
practice stage	<i>Due to time constraints, we can't make students practice writing sufficiently in class</i>	27
feedback stage	<i>Because of our other responsibilities and workload, we can't give feedback often</i>	58
Total		85

N*: number of codes

As seen in table 4.22, one of the main challenges reflected in instructors' statements was the challenges faced during the *practice stage* (n=27). Concerning the practice stage, many teachers stated that due to the time constraints, the time that can be allocated to writing practice is limited. Correspondingly, the teachers noted that although they would want to be able to make students write several drafts for each writing task, the time spent in the classroom doesn't allow teachers to assign more than one or two drafts. Hence, due to time constraints, teachers might fail to achieve the main goals of the process writing approach. In addition to the time constraints, the teachers stated that they have difficulty encouraging students to complete writing whole essays during a lesson hour. Thus, although all the instructors acknowledged the importance of writing practice, they were having difficulties in allocating enough time for practice and encouraging students to complete writing essays in the classroom. The following excerpt represents the kind of challenges faced during the practice stage:

T2: "We need to give writing instruction, make them practice writing, give feedback and make the students write again and we have only four lesson hours to do that. It's hard to make students write even two drafts."

As seen above in Table 4.22, the reported challenges surfaced mostly in relation to feedback stages (n=58) of writing in L2. Given that providing unfocused feedback requires instructors to find every single error and provide feedback on each aspect of writing, the teachers are expected to place all the feedback on students' drafts. With regard to the widespread challenges faced during the feedback stage, the instructors also pointed

out that finding errors and showing all of them on paper is difficult as it is hard to fit all the corrective and constructive feedback on the paper. Furthermore, the instructors stated that the most prominent challenge faced during the feedback stage is that providing feedback to lengthy essays of a large number of students is time-consuming and tiresome. The following excerpt exemplifies this issue:

T6: "Initially, I read everything to understand the content. Then, I reread to find the errors. Then, for example, if the student didn't use transitional things, I read them again to mark or highlight them, so it takes a lot of time for me to assess a paper. Depending on the essay and how tired I feel, assessing a single essay can go up to 30 minutes."

In addition to the challenges they face while providing feedback for writing, the teachers also mentioned how the challenges they face while providing feedback can affect the students. According to the instructors, as providing feedback to students might take days, a long time passes between the first draft and the second draft. In addition, the instructors noted that, due to the challenges and constraints, the quality of feedback decreases, and they might ignore or overlook some issues in writing. Thus, the challenges that the teachers face while giving feedback might affect students' learning experiences. The following excerpt shows how the instructors feel about this matter:

T2: "There are at least 15 to 20 people in a class that when we can give feedback to students a really long time passes. For example, when students submit an essay, at least a week can intervene, or at least the weekend intervenes, or sometimes it can be forgotten. Sometimes in the classroom, our efforts may not be enough because of pacing or some other issues. Hence, the students finally get feedback, they can't remember why they made an error or why they wrote in a certain way because they can't get instant feedback."

Although the aim of this study was not to find the difficulties teachers usually have while teaching L2 writing, when they were asked to share their views on the AWE system, teachers mentioned the difficulties they face assessing writing. Therefore, in line with the purpose of the study, it was important to report the challenges which the instructors mentioned while reflecting on the system as this might be linked to the potential of AWE in helping instructors overcoming some of these challenges.

4.5.2. Advantages of using the AWE tool

When the instructors' statements were examined, one of the most prominent main categories emerging from instructors' statements and viewpoints was that there were certain advantages of using the AWE tool (n=177). Data analysis revealed that there were

advantages under six sub- categories. This main category includes six sub-categories which are *the quality of feedback* (n=32), *the quantity of feedback* (n=44), *technical advantages* (n=26), *being appealing to learner's preferences* (n= 12), *to lessen teachers' burden* (n=28), and *helping students to gain autonomy* (n=35). Table 4.23 below shows the sub-categories found in relation to the advantages of using the AWE tool in detail.

Table 4.23. *The subcategories related to the advantages of using the AWE tool*

Advantages of using the AWE tool	Explanatory statements	N*
The advantages of using the AWE tool are...		
quality of feedback	<i>The feedback is guiding and explanatory</i>	32
quantity of feedback	<i>The system found the errors I couldn't notice</i>	44
technical advantages	<i>Since the feedback is automated is fast and practical</i>	26
being appealing to learners' preferences	<i>Writing on a computer, or receiving feedback on phones can be preferred by the new generations</i>	12
lessening teachers' burden	<i>It would save a lot of time for us to focus on other matters</i>	28
helping students to gain autonomy	<i>It allows students to practice and receive feedback outside the classroom</i>	35
Total		177

N*: number of codes

One of the emerging sub-categories in relation to the advantages of the AWE tool was the quality of feedback (n=32). With regard to the quality of feedback, the teachers pointed out that the errors are clearly pointed out on the system and the explanations provided for the errors were clear. Additionally, it was stated that as the system provided indirect feedback instead of automatically correcting the errors, the feedback is favourable and useful.

T2: "I especially liked the feedbacks because it gave the category of each error one by one. Students click on the categories to find errors, and they see them individually, so it is better for students to focus on the errors. As the errors are highlighted and explained in detail, the feedback is clear."

T4: "It shows errors such as grammar, punctuation, spelling one by one. I think it is very nice in this respect because the student can directly see what the error is. I think it's nice that the feedback doesn't correct the error, so it just says it's an error and what the error type is."

Another sub-category revealed as a result of qualitative data analysis was the satisfaction with the quantity of the AWE feedback (n=44). Although the majority of the

instructors stated that the quantity of AWE feedback provided on errors, especially the errors in grammar, usage, and mechanics, was exceedingly sufficient, two instructors expressed concerns as to whether the quantity of feedback was too loaded. Nonetheless, as stated earlier, the majority pointed out that they also prefer to provide unfocused feedback and prefer students to see all their errors in writing. Additionally, almost all the teachers highlighted that while viewing the AWE tool and comparing the AWE feedback and their own feedback given to the same essays, they noticed that the AWE tool detected some errors which were overlooked by them. For instance, some instructors noted that they overlooked some errors in mechanics category. Regarding the unseen errors, the teachers stated that a number of factors affect their performance while giving feedback and some errors may go unnoticed. Furthermore, the teachers added that AWE feedback is advantageous as it can identify the overlooked errors in particular categories and allow learners to correct them as students are expected to improve writing accuracy based on the feedback they receive and being able to detect and providing corrective feedback on errors is of great importance for the revision stage. The excerpts below exemplify how the teachers feel about this issue:

T1: “The system can give detailed feedback on so many things, especially it showed all the mechanical ones, and even some things I missed myself. It looks like the system did not miss what I missed, mechanically, grammatically and, in terms of usage.”

T4: “Due to fatigue, we can overlook some mechanical errors or some errors such as the use of a / an / the or spelling mistakes. I think it's helpful to see them, so it's nice that the feedback provided by the system is abundant.”

T6: “The system can detect an error you missed. Sometimes we don't notice when reading the first draft. We notice it while reading the second draft. We need to mark that we forgot this or there was a typo here. In this case, I feel bad on my behalf. Rightly, I do not lower the grade of the student because when something like this happens, I find the student right, so I think it's an advantage to detect all the mechanical errors.”

As seen in the excerpts above, the instructors noticed that the system found the error that was overlooked by them. When the teachers were asked about how they feel about the fact that the system found some errors that weren't identified by them, they added positive remarks stating that it is advantageous to detect as many errors as possible since providing corrective feedback affects students' improvement positively. The following excerpt is an example of the instructor's views on this issue.

T2: “It actually makes a difference because for example the 'am / is / are' use or 'a / an / the' use are actually important issues because students have difficulties using these at any level,

and we always complain about it. Maybe, we overlook functional words because we do not pay much attention to them, ignore them, and we do not correct it, but it can cause confusion in some sentences. Hence, it is important to prevent this problem by providing feedback.”

In addition to their views on the quality and the quantity of AWE feedback, the instructors also expressed that the AWE system has some technical advantages (n=26) that can facilitate the teaching and learning to write in L2. Some of the technical advantages highlighted by the instructors were the user-friendly interface, the practicality of using such a tool, and the availability of the system. Therefore, it can be inferred from the instructors’ statements that using such tools as AWE can facilitate the writing-to-learn process, especially for the learners. The excerpt below exemplifies the views regarding the technological advantages of using the AWE tool in L2 writing:

T5: “As the tool we viewed is a technological system, and the new generation can adapt better to technological systems, it becomes a very practical method to practice writing. They can use it in every sphere of life. For example, it can even be used while on the bus. It can actually be used everywhere. Then they won’t have issues they experience when they write on paper, so in my opinion, it’s a practical tool.”

Another subcategory emerging from the instructors’ statements was that the AWE tool appeals to learners’ preferences (n= 12) in certain ways. Stating the generational gap between the teachers and the learners, it was noted that learners’ preferences might differ from those of the teachers or the previous generations. It was also noted that the learners’ preferences might vary within a group of students, and technological tools, such as AWE might help to acknowledge individual differences assist students in improving their writing ability by allowing them to come up with individual learning strategies. The following excerpts are examples of the instructors’ views on this issue:

T3: “I think, especially some students will benefit from this. For example, I have students with ailments such as migraine, and others who have a lot of trouble in writing in a classroom environment, or writing a second draft seems like torture to some. However, I had students who were much more comfortable in the computer environment. I think those students would rather write on such a platform.”

T4: “Students would want to use this type of tool individually and instead of handwriting because they are not as used to handwriting as we do because they are a different generation. Such a platform might be more beneficial for some of them.”

Furthermore, the instructors expressed that using some tools similar to the AWE tool can lessen teachers’ burden (n=28). As mentioned earlier, the challenges that the instructors face during the writing stages were a recurrent issue brought to researchers’ attention throughout the semi-structured interviews. With regard to the challenges they

frequently face while teaching writing, the instructors suggested that using such tools or the AWE tool can lessen their burden. The following excerpts are some examples:

T1: "I think using such systems can reduce teachers' burden."

T2: "I think it could be a very good system to ease the burden on teacher."

T3: "Using the system can actually speed up the feedback process."

The final sub-category related to the advantages of using the AWE tool was to help students gain autonomy (n=35). Throughout the semi-structured interviews, it was highlighted that students are not really autonomous as they mainly depend on teachers' feedback to improve their writing ability. Accordingly, it was also pointed out by the instructors that encouraging students to use the AWE tool can help them to gain autonomy and be more involved in their learning process.

The instructors noted that the system can be useful for students to gain autonomy as it allows students to practice more often and with less limitation. Additionally, the instructors stated that the AWE tool can be used by the students who lack guidance and feedback provided by teachers. Moreover, some teachers stated that the AWE tool might motivate students to practice more often, and it might increase students' learning efforts to improve their writing abilities as the type of feedback provided by the system may be conducive for students to increase their learning efforts. The following excerpts exemplify the instructors' statements:

T2: "As students are provided with 10 rights to correct their errors, they can be actually more responsible for their own progress, and since the system does not give direct feedback, students need to think, study and find their errors in every chance they receive feedback. That is why the students are required to improve both their English knowledge and writing ability. It can also positively affect the willingness to work because it is much more effective to search for information as they can learn different things."

T5: "I think students can use the system. As a result, I think it will be beneficial because as far as I see the system is very suitable for individual work. Moreover,, I think it would be a great resource for a student who does not take a course in any school, and the students who do not have the opportunity to work with a teacher."

4.5.3. Disadvantages of the AWE tool

The qualitative data analysis revealed that in addition to the advantages of using the AWE tool, there were certain disadvantages of the AWE tool (n=63) which was also a main category emerging from the instructors' statements. Furthermore, two sub-categories were found in relation to the disadvantages of the AWE tool. The sub-

categories were *the feedback-related issues* (n=42) and *the system-related issues* (n=21). The subcategories related to the disadvantages of the AWE tool can be seen below in Table 4.24.

Table 4.24. *The sub-categories related to disadvantages of the AWE tool*

Disadvantages of the AWE tool	Explanatory statements	N*
Disadvantages of the AWE feedback are		
feedback-related issues	<i>It would be better if it provided more personalized feedback on organization</i>	42
system-related issues	<i>Students may not understand some words on the system</i>	21
Total		63

N*: number of codes

Related to the first sub-category, *feedback-related issues* (n=42), the instructors brought to the researcher's attention that the AWE system did not provide feedback on tense errors and singular/ plural noun errors. They also pointed out that the feedback provided for certain aspects of writing might be insufficient. As stated earlier in the method and the results chapters, the AWE tool doesn't provide feedback on singular and plural noun errors and tense errors.

The lack of corrective feedback provided for these particular error types was also noticed by the instructors and it was viewed as a disadvantage and the necessity of incorporating feedback on these error types was a common view among the interviewees. Furthermore, the AWE feedback provided for the organization and development of writing was found insufficient and too general.

The instructors stated that for students to benefit from formative feedback provided for the organization of writing should be more specific and personalized. Finally, the lack of feedback provided on content was another drawback of the AWE tool according to the instructors. Correspondingly, the teachers noted that the importance of providing feedback on content should not be overlooked and it should be regarded as important as corrective feedback provided to improve writing accuracy. The following excerpts exemplify the instructors' views concerning this issue:

T1: "I think the feedback given for other categories is enough, it has explained the problems very well but the system is also weak in terms of providing feedback on content, but I don't think it can do that since it is a machine. Human intelligence is needed to do that."

T5: “It would be better to see some feedback on meaning and ideas, I think it should be added to the system if possible.”

T7: “It gives the same feedback for all of them in the organization and development part. The student cannot understand much from there. In other words, when directed there, it gives general feedback in terms of essay structure, not the errors.”

As can be seen in the excerpts, the feedback provided by the system on content and organization needs to be elaborated, and human intervention and guidance might be needed concerning these issues.

In addition to instructors' concerns regarding the lack of feedback on certain areas, it was also stated that there are some *system-related issues* (n=21) that should be improved or adapted to be more suitable for EFL learners. Concerning the system-related issues, the instructors stated that the language used by the system may be difficult for all students to understand as there might be unfamiliar words if students' level of proficiency is low. Moreover, two of the instructors added that a rubric with a wider range might be more useful to be used in writing evaluation, and it may help students to better understand their level of writing proficiency. The excerpt below is an example of the instructors' statements related to the system-related issues:

T6: “There are some terminological words. Thus, the student using the system must be at the minimum intermediate level. If possible, it should be upper because I think the students may either get lazy to read them, or they will not be able to fully understand them. For example, words such as neglect, misinterpret are some difficult words in the system, so I think it would be better to simplify the language in feedback for our students”

As can be understood from the excerpt above and the instructors' views, it can be inferred some adaptation to the system might be needed for EFL learners to benefit from the AWE tool even more. Hence, students' level of proficiency and the context that an AWE tool can determine how an AWE tool can be used.

4.5.4. Comparison of AWE assessment and teachers' assessment of writing in L2

Another main category found as a result of qualitative analysis was the comparison of AWE feedback and teachers' feedback (n=95). With regard to the comparison of these two feedback types, *the similarities between the two types of feedbacks* (n=51) and *the difference between the two types of feedbacks* (n=44) were identified as two sub-categories. The sub-categories and the number of codes belonging to those two sub-categories can be seen in Table 4.25.

Table 4.25. *The sub-categories related to comparison of AWE assessment and teachers' assessment*

Comparison of AWE and teachers' assessment	Explanatory statements	N*
When the AWE and teachers' feedback are compared there are... similarities	<i>It doesn't correct the errors but identifies them just like we do</i>	51
differences	<i>There are more subcategories on the system for the error types</i>	44
Total		95

N*: number of codes

According to the instructors, *the similarities* (n=51) between their assessment of L2 writing and AWE assessment of writing were mainly related to the feedback and the scores obtained by the AWE system and the instructors. As reported earlier in the results of the first question, the qualitative analysis of the instructors' opinions about the scores given for the essays which were also evaluated by them showed that the instructors found the scores given by the system similar to the scores they gave to the same essays. In addition, some instructors noted that the explanations given by the system regarding the scores resemble the explanations attached to intervals on the holistic rubric used by them. In terms of feedback, although it was stated that the feedback found to be similar to the feedback given in categories other than content and organization, the instructors pointed out that both AWE feedback and the corrective feedback provided by the instructors were found to be similar in terms of error types which feedback was given for, and the appearance of feedback. The excerpt below exemplifies the instructors' views:

T2: "I think we found the same errors and gave similar feedback. For example, I also show students' word level or grammatical errors in the text and write comments to their ideas or other things separately above. I saw similar comments on the system apart from the corrective feedback."

With regard to *the differences* (n=44) between the AWE assessment and instructors' assessment of writing, some of the differences mentioned by the teachers were that on the AWE system, there are more error types specified within the categories. Therefore, more specific feedback was given in some categories. To illustrate, teachers were able to identify the error with the symbols used by identifying the general category of some errors, while AWE was also able to indicate additional error types within the error categories identified by the teachers. In addition, it was stated by the teachers that there was a difference in feedback types. It was stated by some teachers that AWE feedback appears to be more like corrective feedback, but teachers can add explanations in terms

of idea development in addition to corrective feedback, and they use motivational symbols or expressions. Regarding this point, it was expressed that the teachers build rapport with the students and the feedback given was more subjective as the instructors were affected by this situation. The following excerpts are examples of the instructors' views on differences:

T1: "The system gave more feedback and found more grammar and mechanics errors. This was nice for me to see. As I said in terms of content, there are differences. There are differences in terms of punctuation because I did not pay too much attention to punctuation and I did not give specific feedback. On the other hand, the system gave so much feedback on punctuation."

T3: "I can make more detailed or directive touches about the content. In a simpler way, I can direct students regarding what to do, or if the student does not understand, they have the opportunity to speak with me. In short, the feedback differs in terms of the ability to give ideas"

T7: "There is a difference with the feedback we give because we give a lot of motivational feedback. We write motivational comments, sentences to encourage the student for the better."

4.5.5. Implementing the AWE tool in EFL writing

The final main category emerging from the instructors' viewpoints was the ways of implementing AWE in EFL writing (n=83). In relation to the ways of implementing the AWE tool in writing, two sub-categories were found. The first subcategory was *the use of the AWE tool in the classroom* (n=54), and the second sub-category was *the use of the AWE tool outside the classroom* (n=29). Table 4.26 displays the subcategories and the number of codes found in relation to implementing AWE in EFL writing.

Table 4.26. *The subcategories related to implementing AWE in EFL writing*

Implementing AWE in EFL writing	Explanatory statements	N*
The AWE feedback can be used...		
in the classroom	<i>AWE can focus on errors and we can focus on content</i>	54
outside the classroom	<i>We can assign homework to be submitted on AWE</i>	29
Total		83

N*: number of codes

In relation to implementing the AWE tool in teaching and learning writing in L2, teachers pointed out that the AWE tool can be used *in the classroom* (n=54) during various stages of process writing, such as pre-writing, while writing, and revision stages.

Nonetheless, the instructors agreed that using AWE alone in the classroom setting would not be sufficient and that a mix of both AWE and teacher's presence was the ideal way to include AWE in teaching writing. Regarding utilizing the AWE tool to feedback and revision processes, the instructors noted that using AWE to give corrective feedback and providing teacher's feedback by adding comments in terms of organization or content will both increase the effectiveness of the feedback received by students, and it can speed up and facilitate the provision of the feedback process. Furthermore, two teachers stated that in addition to the provision of individual feedback, AWE feedback can be used while giving whole class feedback as the interface of the system allows teachers to access the essays written by the students and detect the common errors to address the needs of the students with ease. Thus, it was a common view amongst the instructors that a mixed-use of two types of feedback, that is, to combine AWE corrective feedback with teacher's feedback provided on certain aspects of writing would be beneficial for the students. The following excerpts from the interviews emphasize this:

T1: "If I used it, I would use it like this; I would like the students to correct all their mechanical errors on the tool, but I also wish they could forward them to me as well. Maybe, first, they would upload it here, they would see the mechanical errors and change them, then they would send it to me. I would review the remaining errors and review the essay in terms of content."

T4: "The teacher can only be expected to give feedback on the organization and ideas after the system finds errors. For example, I could have a general look at the first draft, and if there is not a major problem with ideas, I would wait for the students to correct errors using the system, and then I could give feedback about the organization if there is a situation that I really deem necessary."

With regard to the use of the AWE tool *outside the classroom* (n=29), the instructors stated that AWE can be used individually by students who want to gain autonomy outside of the classroom. In addition, teachers stated that the AWE system can be used to assign students more homework and increase the opportunity for students to practice outside the classroom. The excerpts below show how the instructors would use the AWE feedback outside the classroom:

T2: "For writing, we write essays or paragraphs, and we give them homework or write it in class. Instead of printing it on paper, we can ask students to write them directly on the system and correct their mistakes on it, or I can first show the students how the system works and tell them how they can improve themselves and make them write more paragraphs or essays."

T3: "If I could implement the process writing process, for example, if I consider our own process, I would definitely like to read a paper from the students, but if we had such a system, students would have more chance to write, more writing could be assigned."

As can be understood from the excerpts above, the instructors stated that besides supporting the use of AWE in the classroom, the use of AWE outside the classroom will also support students' effort of developing writing skills and it will enable teachers to make students practice outside the classroom. Hence, the instructors would want to use the AWE system to make students submit and revise assigned homework outside the classroom.

In conclusion, the results of the research revealed that teachers face many problems and limitations while teaching writing. Considering the process writing process, teachers stated what are the advantages and disadvantages that AWE will provide to EFL students when writing in L2. In addition, based on their observations, teachers compared AWE and teachers' assessment in terms of assessment and revealed the similarities and differences between them. Finally, as a result of their observations, teachers expressed their willingness to use AWE and stated that they would like to use AWE in the classroom and outside the classroom in a supportive way, although not alone, in process writing stages.

5. DISCUSSION

5.1. The Correlation between the Holistic Scores Given for EFL Students' Essays by the Instructors, and the Correlation between the Instructors Scores and the AWE Scores

The first question asked if there was a correlation among the scores given for EFL students' essays by the writing instructors, the AWE system, and the external rater. The findings revealed that the correlation coefficient between AWE scores and instructors' scores was significantly and positively high ($r = .862, p < .05$). That is, when the AWE scores increase, the instructors' scores increase as well, or when the scores of one variable decreases, the scores of other variable decreases too. It was also found that the correlation coefficient between instructors' scores and the external rater was significantly and positively high ($r = .804, p < .05$) which means that the instructors' scores and the external rater's scores increase and decrease together. Moreover, the correlation between the AWE scores and the external rater score was significantly and positively high ($r = .876, p < .05$). In other words, the AWE scores and the external rater's scores tended to increase and decrease together as well. Taking all together, it can be suggested that the agreement between AWE scores and the instructors' scores can reach a similar rate of agreement between an instructor and another external rater. Additionally, the fact that AWE scores may agree well with instructors supports the idea that AWE scores and human scores show similarities when both are given on rubrics with criteria (Xi, 2010). Hence, the similar agreement rates among AWE scores, instructors scores, and the external rater's scores can be explained by the fact that both the instructors working in the study context and the AWE system evaluate essays using a holistic rubric. As holistic rubrics are mainly concerned with overall writing performance rather than assessing the components of writing separately and adding up points in different sections to score essays, the AWE assessment and instructors' assessment might have shown similarities.

In addition, when the correlation coefficient between each instructors' scores and AWE scores were compared with the correlation coefficient between instructors' scores and the external rater's scores, the results indicate that in some cases, the correlation between the AWE scores and the instructors' scores were higher compared to the correlation between instructors' scores and external rater's scores. In other words, it can be assumed that AWE scores might not only be were positively and significantly

correlated with instructors' scores but also the correlation between AWE scores and instructors' scores can be higher than the correlation between two instructors. This substantiates previous findings in the literature showing that both human rater scores and AWE scores can be positively and significantly correlated with other raters, and in some cases, the correlation between AWE scores and other raters can be higher (Attali & Burstein, 2006, Burstein & Chodorow, 1999, Ramineni et al., 2012, Shermis et al., 2002). Nonetheless, it should be noted that despite the numeric differences in the correlations, both the AWE scores and the external rater's scores were positively and significantly correlated with the instructors' scores in the same intervals except for two comparisons. That is, when the correlation between an instructor and the AWE scores were significantly and positively high, the correlation between the instructor's and the external rater's scores were significantly and positively high as well in most cases. Additionally, when the correlation values were in different intervals, the correlation between the instructors' scores and the AWE scores was higher. Thus, it can be assumed that the AWE tool used in this study was consistent in correlating well with the instructors' scores. The current study, therefore, contributes to a growing body of evidence indicating that AWE scores may show consistency in correlating well with instructors' scores in classroom-based L2 writing assessment environments.

Despite the fact that the results concur well with the previous studies, the findings differ from some previous results in the literature which reported weak or moderate correlations found between human rater scores and AWE scores (Li et al., 2014; Wilson & Roscoe, 2020) since this study revealed a high correlation between instructors' mean scores and AWE scores. This may have occurred due to contextual differences as the majority of the previous studies were conducted in high-stakes environments in which students are evaluated by the use of standardized tests, such as TOEFL and GRE tests rather than classroom assessment environments. Another explanation why higher correlation values were found in this study compared to the findings in some previous studies can be that, as also reported by the instructors taking part in this study, the rubric used in the study context and the rubric used by the AWE system share some surface similarities, and the scores given by the system and the instructors were reported to be in the same or similar intervals. Moreover, the instructors pointed out that the explanations provided by the system resembled the explanations attached to the intervals. In other words, because the rubric used by teachers in the study context was similar to the rubric

used by AWE, teachers were familiar with using a rubric as AWE uses and that might be the reason why there was not much difference between AWE's and instructors' evaluation of the essays. Another reason why there was a significant and positively high correlation between AWE scores and instructors' scores might be that although AWE scores can be given on many different scales such as four-point or six-point scoring scales, in this study AWE holistic scores were given on a six-point scale while the instructors used a holistic rubric had seven intervals with the first interval being zero as in "no answer". Therefore, there were six main intervals in the instructors' scoring rubric as well. On the other hand, in some of the studies finding a weak or moderate correlation between the AWE scores and the human scores, there were notable differences between the AWE holistic rubric and the rubric they used those studies. To illustrate, the instructors participating in Wilson and Roscoe's study (2020) used an analytical rubric with scores ranging from 0 to 11 and an AWE tool scoring essays out of 6 to investigate the correlation between instructors' scores and the AWE scores. As a result, the type and format of the rubric/scale used and the instructors' familiarity may affect the correlation in scoring.

In addition to the studies conducted in NES and EFL contexts, the findings regarding the correlation between the instructors' scores and the AWE scores are in contradiction with a previous study conducted in the Turkish EFL context as much higher values for correlation coefficient were found between the instructors in the present study compared to those reported by Doğan (2013). To the researcher's best knowledge, this is also the only study that investigated the correlation between instructors' scores and AWE scores in a Turkish EFL. The study conducted by Doğan (2013) revealed that correlations among the three instructors were moderate ranging from .58 to .65, whereas the correlation between the mean scores of instructors and the e-rater was reported as high ($r = .89$). As indicated by Doğan (2013), the low agreement among instructors' scores may point to a need for a training period taking place prior to assessment. In this respect, in the current study, it was ensured that the instructors were experienced in assessing essays with the rubric used in this study context and a standardization session was held prior to the assessment. Hence, training sessions provided by the testing unit and a standardization session held before the assessment may have contributed to the agreement values. These sessions might have helped the instructors become familiar with the assessment tool, clarify what was expected in each band of the rubric, and achieve a standard evaluation by adjusting themselves to the statements in the rubric. Taking all together,

standardization sessions, training, and experience are important to achieve inter-rater reliability between human raters. As the findings in the literature suggest, the correlation between human scores might be lower than the correlation between human scores and AWE scores. As the AWE tools have been used as a second-rater in high-stake testing environments (Hegelheimer & Lee, 2012), it can be assumed that AWE tools can be used to resolve discrepancies between instructors. Furthermore, they can be used as a second-rater in EFL contexts as the findings conducted in EFL contexts suggest that a higher agreement can be achieved when AWE scores complement a human score rather than another human score (Doğan, 2013; Ebyary & Windeatt, 2010).

To put it in a nutshell, it can be indicated that the findings regarding the correlations among instructors' scores, AWE scores, and the external rater's scores compare well with most studies in the literature and a previous study conducted in the Turkish EFL context. Although the findings are based on limited data collected in a single context, the results point to the likelihood that AWE scores can be consistent with the scores provided by instructors, and an AWE tool might be used to score essays in the EFL classroom assessment context. It is worth mentioning that the studies which examined the agreement between AWE tools and human raters in the Turkish EFL context are rather scarce. Therefore, the results of this study provide considerable insight into the use of AWE tools for assessing students' writing performance in a higher education context. Using such automated tools can help teachers deal with students' writing performances in crowded classes, save time and energy for improving writing performance in a busy and tight schedule, and create opportunities to work on issues such as error correction, feedback, and material preparation in the time spared (Li et al., 2015; Zupanc & Bosnic, 2016). That said, although utilizing an AWE tool to score essays would most likely ease teachers' burden, using an AWE tool in isolation to evaluate writing performance requires much consideration as a summative assessment is a crucial component of the decision-making process which has a direct impact on students, teachers' insight, and the curriculum. Therefore, in addition to accurate and consistent scoring performances, teachers' judgment and insights should be involved in summative assessment leading up to a decision-making process. Nonetheless, using AWE as a second evaluator is becoming more common, especially in a high-stakes testing environment as the studies show that AWE scores are in good agreement with human raters (Attali & Burstein, 2006; Burstein & Chodorow, 1999; Ramineni et al., 2012; Shermis et al., 2002). Considering the fact

that usually two raters are involved in the evaluation of writing performance and mean scores of two raters are taken into consideration in ESL and EFL contexts, AWE can be used as a second-rater, or it can be utilized to resolve discrepancies between instructors. In this respect, AWE appears as a practical and feasible tool to be used in EFL/ESL contexts to enhance scoring of students' writing performances, save time and energy, and support teachers in writing classes. Especially in EFL contexts, such as the one in this study, in which instructors have various writing classes with too many students, AWE tools may lessen the burden of the instructors and help them spare more time for enhancing teaching. To conclude, AWE scores might be utilized in summative assessment in order to ensure consistency in scoring, and it might also be used to inform students about their performances and to encourage them to increase their efforts in L2 writing.

5.2. The Change in EFL Students' Writing Accuracy after the Provision of AWE Formative Feedback and Teachers' Formative Feedback

The findings of the study point to the effectiveness of both AWE feedback and teacher feedback in improving writing accuracy as the results emerging from the data gathered from the experimental and the control groups revealed that the difference between pre-feedback and post-feedback error counts were significant in both groups. In other words, both teacher's feedback and AWE feedback led to a significant decrease in writing errors. This concurs well with the previous studies in the literature suggesting that using AWE formative feedback result in positive changes in writing performance by reducing errors (Attali, 2004; Li et al., 2015; Long, 2013; Polermo & Thomson, 2018; Ranalli et al., 2017; Warschauer & Grimes, 2008). In addition, the findings are consistent with two earlier studies conducted in the Turkish EFL context. In one of those studies, Bulut (2019) found that although both AWE feedback provided to the experimental group and the teacher feedback provided to the control group were effective in improving writing accuracy significantly, AWE feedback led the students in the experimental group to receive better scores in the post-feedback assessment of essays. Likewise, Gençer (2019) found that AWE feedback has positive feedback on students' L2 writing performances both in the short-term and long-term. Similarly, the results of this study contribute to a growing body of findings suggesting that AWE feedback can be used in formative assessment of L2 writing and help learners to improve their writing

performances. Therefore, the findings of the current study regarding the effectiveness of AWE feedback on improving writing accuracy support the notion that AWE has great potential in not only guiding students toward improving their writing performances but also reducing teachers' burden of correcting errors on each student's essay on a regular basis and facilitating the assessment process.

Besides the general improvement in writing accuracy, the error reduction in 11 categories was examined. Accordingly, it was ascertained that although both teacher feedback and AWE feedback led to a significant decrease in error counts, the comparison of the experimental and the control groups revealed that there was a significant difference between the two groups in three categories which are ill-formed verbs, articles, and punctuation categories. In other words, it was revealed that the AWE feedback was more effective in correcting ill-formed verbs, articles, and punctuation error types compared to teacher feedback. Furthermore, it was seen that there was no significant difference between the groups regarding the error decrease in the ill-formed verbs category when the changes in the second and the third essay drafts were examined. Hence, the AWE feedback was consistently more effective than teacher feedback in correcting article, punctuation errors and correcting ill-formed verb errors, albeit the AWE feedback was more effective in correcting the errors made in those three categories, especially in articles and punctuation categories, there was no significant difference between the effectiveness of teacher feedback and AWE feedback on correcting the errors in the other eight categories (fragment, run-on sentence, sentence structure errors, subject-verb agreement errors (SVA), wrong word choice, preposition, spelling, and capitalization). Therefore, the findings regarding the difference between the experimental and the control groups in terms of the error reduction in certain error categories substantiate other previous findings in the literature. To exemplify, both studies conducted by Khoii and Doroudian (2013), and Wang et al. (2013) reported significant findings in favor of AWE feedback as in both studies the experimental groups significantly outperformed the control groups in writing accuracy. Additionally, a study conducted by Long (2013) found that AWE feedback was more effective in correcting errors in mechanics category compared to other main categories. In a similar vein, the results of the present study put forward that although the control group was not outperformed by the experimental group in all the error categories, the significant difference found in the three categories was in favor of the effectiveness of AWE feedback, especially in mechanics and usage aspects of writing.

A possible explanation for these results might be that as illustrated in Table 4.7, a striking difference can be noticed between the AWE feedback and teacher feedback provided for these errors. To illustrate, while teachers use codes, such as ‘^’ for ‘article’ and ‘P’ for ‘punctuation’ to address errors, the AWE tool used in this study specified these errors in more detail. That is, the system provided students with explanations and clues especially for the errors under mechanics category, such as article and punctuation errors. Hence, it can be assumed that although error codes are frequently used to give indirect corrective feedback and preferred by teachers to save time, students may not understand how to correct these errors and they may need further explanations (Lee, 2003). To conclude, although both teacher feedback and the AWE feedback used in this study were indirect feedback types, the AWE feedback might have assisted the students better to correct the ill-formed verb errors, article errors, and punctuation errors with the help of the clarifications the system provides. This finding has important implications for L2 writing assessment as it shows that AWE feedback is time-saving and it may also provide guidance and effective feedback.

With regard to the error types that AWE was more successful in correcting, the finding of this study correlates favorably with Ranalli et al.’s study (2017) as the current study revealed that AWE feedback led to a significant reduction in certain error categories such as wrong word, punctuation, subject-verb agreement, and articles. Hence, the findings of the current study further support the idea suggested by Ranalli et al. (2017) that AWE feedback can be more useful to improve form-based aspects of writing. Likewise, a comparative study conducted by Long (2013) showed that AWE feedback was more effective in correcting errors, such as punctuation, spelling, short sentence, and confused words. What is more, AWE feedback was found to be more focused on linguistic and stylistic features of writing, whereas teacher feedback was more focused on meaning. Furthermore, Warshauer and Grimes (2008) and Li et al. (2015) reported that the AWE feedback was more effective on word-level errors and most of the improvement was mainly on grammar and mechanics aspects of writing.

As indicated by the previous studies in the literature, the findings of the current study provide further evidence for the effectiveness of AWE feedback on word-level errors. It was found that the AWE feedback was more effective in improving the mechanics aspect of writing as it was more effective in reducing errors in punctuation and article categories. Hence, similar to the findings of other studies (Long, 2013; Ranalli et

al., 2017; Warschauer & Grimes, 2008), the evidence found in this study points to the effectiveness of AWE feedback on form-based aspects of writing. That said, the present findings are significant in at least three major respects. To begin with, the results indicate that AWE feedback can be as effective as teacher feedback in improving writing accuracy and, therefore integrating AWE into the formative assessment of classroom-based writing might result in positive results as EFL and ESL students attach great importance to accuracy in writing and receiving regular formative corrective feedback (Lee, 2003). Moreover, considering the fact that punctuation, article, and the ill-formed verbs were among the most frequent errors the students made in the study context, the results may indicate that using AWE in formative assessment can save a great amount of time for the teachers as the more frequent the errors are the more time teachers need to spend indicating those errors. Furthermore, as the results suggest that AWE feedback might be more effective than teacher feedback in correcting these error types, AWE feedback can lead to reducing the occurrence of these frequent errors significantly in the long term. Finally, although this study did not focus on the effectiveness of AWE on discourse-level writing performance, the findings might point to the usefulness of AWE feedback on word-level and sentence-level improvement. Thus, the findings may suggest that AWE feedback can be used to provide formative corrective feedback on errors, and it can be combined with teacher feedback to ease teachers' burden by allowing them to focus only on the aspects of writing which AWE might disregard or does not provide feedback on.

5.3. EFL Students' Views on the Use of AWE in Learning and Assessment of Writing in L2

The results of the present study point to overall satisfaction with the AWE tool and the AWE feedback. When the students participating in this study were asked to reflect on their experiences utilizing AWE feedback to improve writing accuracy in L2, the students expressed that they were satisfied with the quantity of the feedback provided by the system as the feedback given on errors was abundant. They were also pleased with the quality of the feedback they received as the feedback was reported to be clear and explanatory. In addition to the AWE feedback, the features of the system and advantages of using an online automated writing assessment tool, such as its availability and practicality were pointed out. Taken together, the findings highlight that the students were satisfied with the AWE tool and the AWE feedback. Hence, the results lend support to

previous findings in the literature regarding the students' positive attitudes towards AWE. To illustrate, when El Ebyary and Windeatt (2010) conducted a study examining ELT trainees' attitudes towards AWE with pre-treatment and post-treatment questionnaires, it was revealed that over the course of the study, the participants' attitudes changed positively. Even though the participants expected to be dissatisfied with the AWE feedback, according to the post-treatment questionnaire analysis, the ELT trainees were pleased with the feedback. It shows that when learners are given an opportunity to receive feedback from AWE tools, they can get used to such feedback in time and may receive long-term benefits. In the current study, students' general satisfaction with the feedback they receive in a short treatment period may suggest the use of such tools for language learning on learners' behalf.

On the other hand, AWE tools have not always received positive reactions but are perceived negatively by some students in the early studies as well. To illustrate, Chen and Cheng (2008) conducted a study aiming to explore EFL university students' attitudes towards the use of AWE in L2 writing, and some students and the course instructors held negative reactions against the system. Nevertheless, when Cheng (2017) conducted another similar study on perceptions of EFL university students towards AWE, it was found that the AWE feedback wasn't only effective in improving students' writing performance, but it was also perceived positively by the students. Likewise, many other recent studies (Li et al., 2015; Parra, 2019; Zhang, 2020) revealed that the students were satisfied with their experiences with AWE and they showed positive attitudes. Taken together, these differences between early studies and the recent studies can be accounted for in part by the time the studies were conducted and the features of the AWE tool utilized in these studies. In other words, as suggested by Wang, Chen and Fang (2011) students' perception towards to use of technology in education is parallel with their familiarity with it. As stated by one of the instructors participating in the present study, the current generation of students may not feel as comfortable as the previous generation in the traditional pen and paper approach to writing assessment and they may prefer to utilize technological tools for both instruction and assessment purposes. Hence, the results point to the likelihood that depending on their level of familiarity with computer technologies and the AWE tools, in particular, students can be willing to use AWE tools in writing instruction and assessment, or their attitudes may change positively once they get familiarized with such tools. In this respect, the findings of the previous studies and

the current study support looking for new alternatives in writing assessment like the implementation of AWE tools and familiarize students with these digital tools. Without the presence of the teacher, learners can benefit from such tools outside the physical borders of the classrooms.

It is also worth noting that students found the AWE tool trustworthy by addressing the fact that teacher's fatigue and illegibility of handwriting may affect teacher's assessment; and thus, some errors may be overlooked. Therefore, the automated feedback was perceived as trustworthy by the students. This aligns with the previous findings suggesting that AWE tools are found trustworthy by students (Grimes & Warschauer, 2010; Li et al., 2014). Trust in automated tools stems from the information gained about user's performance through the system, how a system process, and the purpose of the system used (Roscoe, Allen, Johnson & McNamara, 2018). Hence, a reasonable explanation of why the students who participated in the current study found the AWE tool trustworthy may be that students were able to understand how the automated tool works. Moreover, the students were already familiar with the formative assessment process and receiving corrective feedback, and therefore they were informed about the purpose and expected outcomes of using this tool. Additionally, the students were provided with tutorials and materials explaining how the AWE tool works. Thus, they were informed about how the AWE tool used in this study process and provide information. Furthermore, the feedback AWE provides was found clear, understandable, quite objective and it was outlined neatly and this may contribute to the information the students gained about their performance through the system. That is to say, once the students are familiar with the format of the feedback the system generates, the purpose of the feedback given, and the information given on their performance, they can track their errors and benefit from the feedback more. As a result, these factors may contribute to the students' positive attitude towards such automated writing systems.

In addition, many advantageous features of the system were reported by the students, including its being easily accessible, practicality, providing instant feedback, and its allowance to receive explanatory and abundant feedback. These factors may have also contributed to students' writing performance as it was also revealed that students believe the AWE feedback helped them to reduce their writing errors, increased their attention to avoid common errors, and led to learning through formative feedback. Hence, it can be concluded that students found AWE feedback helpful in terms of improving

writing accuracy, increasing their awareness, and promoting learning from errors. This concurs well with previous findings in the literature (Li et al., 2015; Palermo & Wilson, 2020; Parra, 2019; Warschauer & Grimes, 2008) including two studies conducted in the Turkish EFL context (Bulut, 2019; Gençer, 2019) which found similar findings in favor of AWE feedback for improving L2 writing. To exemplify, Gençer (2019) found that according to the EFL students at a university, the AWE tool was helpful and the majority of the students were satisfied with their experiences of automated feedback especially in terms of grammar and mechanics. In the current study, learner satisfaction may have stemmed from their positive attitudes towards the system.

It is also noteworthy that the AWE feedback can lead to a mediated learning experience (MLE). According to Feuerstein, Klein, and Tannenbaum (1991), for an assessment type to be considered as MLE, the assessment type should serve a formative purpose. In other words, the feedback should be given intentionally, the information should be transferred to future situations, and the experience should be perceived as meaningful and appreciated by the participant. As suggested by Li et al., (2014) the AWE tools can meet three fundamental criteria of MLE, which are intentionality, transcendence, and meaningfulness. Therefore, the main reason why the students of the current study reported that their interaction with AWE led to learning might be that the information gathered through revision was transferred to similar situations in the following revisions. Hence, it can be assumed that the knowledge, resulted from AWE feedback, might possess the quality of being transcendent, and the feedback may lead to an MLE. Concurrently, AWE feedback can help students to realize what they need to do to improve their writing ability. Moreover, since AWE feedback can be received both outside and inside of a classroom, it can also be assumed that AWE feedback could be a tool resulting in mediated learning experience both in and out of the classroom.

Another noteworthy finding resulting from the students' statements was that the AWE feedback had positive effects on gaining autonomy. It was expressed by the students that using AWE feedback raised their awareness of their strengths and weaknesses which ultimately increased their learning efforts. As the AWE tool utilized in this study provided trait feedback analysis which shows the common errors made under main categories and the error types within each category, it helped to diagnose weaknesses and strengths of students in L2 writing. Thus, the students noted that being able to diagnose their weaknesses easily helped them to detect the areas of improvement

and it led them to put effort to focus on improving those areas in particular. Therefore, it can be interpreted that the AWE feedback may have raised students' metalinguistic and metastrategic awareness. That is, students were able to detect the writing problems occurring in their essays and they were able to come up with tactics or build up personal strategies to tackle these problems. Both metalinguistic awareness and metastrategic awareness are significant types of metacognitive awareness that are crucial for a revision stage to be meaningful since detecting errors or flaws in writing and developing strategies improve writing ability (Cotos, 2014). According to Horning and Robertson (2006), attaining different types of metacognitive awareness are often acquired by expert writers compared to inexperienced writers. Hence, thanks to the opportunity to receive feedback on multiple drafts, the students may have raised metacognitive awareness which led them to gain autonomy and put more effort into overcoming their weaknesses. In addition, the relationship between writing experience and raising metacognitive awareness brings the question of whether receiving AWE feedback led to metalinguistic and metastrategic awareness since students were in intermediate level, and whether the same effect would be observed in lower-levels. Therefore, future studies examining the effect of AWE feedback on gaining metalinguistic and metastrategic awareness in different levels of English proficiency can shed light on this question.

Moreover, with regard to gaining autonomy through AWE, it was put forward by the students that using the AWE feedback fostered independence from the teacher as they were able to use the AWE feedback outside the school and they were also able to use AWE feedback for many different purposes in addition to the assignments. To illustrate, the students expressed their willingness to use AWE feedback for online and offline writing activities, such as story or blog writing. Furthermore, almost all the students expressed their willingness to use the AWE feedback when they continue their studies in their faculties for many different purposes, such as getting ready for presentations, checking language accuracy before handing in their assignments, and writing research papers in English. That is to say, the findings emerging from students' statements highlighted the importance of being able to access an automated writing evaluation tool. Although students receive formative feedback on a regular basis when they are in preparatory programs, they won't be able to receive traditional feedback from teachers regarding their writing errors in the following years.

Similar to many other EFL contexts, in Turkey, English-medium higher education programs are established in most universities, and English is prioritized as a foreign language (Kırkgöz, 2019). Nonetheless, writing in L2 is one of the most challenging areas for students studying in English-medium programs in the Turkish EFL context (Kamaşak, Sahan & Rose, 2021). A study conducted by Gurel (2010) examining the challenges Turkish students face in academic writing in English revealed that university students face many challenges, such as linguistic difficulties, not being able to express ideas in English, interlanguage errors, and lack of writing. Moreover, some advisors who participated in the study stated that they correct errors as a last resort and urge students to improve their writing ability through their individual efforts since providing detailed feedback concerning all the errors can be too demanding (Gurel, 2010). Therefore, it can clearly be seen that, especially in higher education, although students reach a certain level of proficiency before they continue their studies in their departments, they need guidance and formative feedback regarding their ability to write in L2. In this respect, the assistance they can receive from their lecturers and advisors concerning writing accuracy and fluency can be quite limited. Hence, it is abundantly clear that throughout their university education, Turkish students may need to receive help, guidance, and formative feedback regarding their writing accuracy and fluency. As pointed out by the students in this study, an implication of the findings might be that the AWE tools can be utilized to promote autonomy, self-regulated learning, and it can be used for various purposes stated above.

Another notable finding in relation to gaining autonomy was that the AWE tool increased students' motivation to write more and receive more feedback. In other words, the findings indicate that AWE may have changed students' revising behavior positively by motivating them. The findings regarding the relationship between using AWE and motivation are consistent with previous results suggesting that AWE tools increase students' motivation to write in L2 and help them revise their writing (Bulut, 2019; Li et al., 2015; Warschauer & Grimes, 2008; Wilson & Czik, 2016; Wilson & Roscoe, 2020). Given that intrinsic motivation is believed to be essential to initiate self-regulated learning and promoting learning effort (Pintrich, 1999; Stone, 2017), without sufficient intrinsic motivation, students wouldn't use the feedback outside the classroom for the sake of self-improvement. As the students in the current study expressed their willingness to use AWE feedback outside the classroom for self-improvement, the students might have become more motivated as their revisions were more learning-oriented, and they were motivated

to improve their L2 writing abilities rather than for the solemn purpose of submitting assignments. All in all, the students used feedback provided by the AWE system to enhance their own learning in L2 writing and become motivated to progress in L2 writing.

According to Cotos (2014), the type of AWE feedback has a direct impact on students' experiences and perceptions, and the results concerning students' preferences were in favor of color-coded feedback. Likewise, in the current study, the students noted that as the errors were highlighted and colored, they attracted attention, and since the explanations appeared on the highlighted errors, they found the AWE feedback clear. Thus, the AWE feedback utilized by the students in the present study may have led to positive learning experiences owing to the quality and quantity of feedback given by the system. In the study conducted by Cotos (2014) it was found that feeling accomplishment, proof of improvement, and previous disappointment with traditional feedback were the main reasons why the students felt excited and motivated by the feedback received by the system. Therefore, it can be presumed that the students who participated in the current study may have felt motivated to practice writing as the system allows them to revise an essay 10 times and provides feedback for each revision. Thus, students are able to see the proof of improvement gradually and feel accomplished seeing the reduction in each error. It can also be presumed that viewing the number of errors in categories clearly and realizing that the number of errors decreases after each revision may have boosted the feeling of accomplishment.

Emerging from the students' reflection on their experiences with the traditional feedback and the AWE feedback, it was revealed that there are some similarities and differences between the traditional feedback and AWE feedback. Some of the similarities reported by the students were the types of error categories and error types for which feedback was provided, the usefulness of both feedback types, and the guidance received by the feedback received from the system and the teacher. Likewise, in a study conducted by Calvo and Ellis (2010), the students stated that both feedback types were similar in terms of writing fulfillment and feedback gains. The reason why the AWE feedback was found similar to teachers' feedback might be that both feedback types can be considered as indirect feedback types since both do not entirely correct the errors, but they clarify the type of error made. In this respect, both feedback types drive learners' attention to their errors and create space and opportunity for them to gain awareness of their own errors for self-correction. Moreover, both feedback types can be considered as formative corrective

feedback as the students are expected to improve writing accuracy by utilizing the feedback they receive. As a result, AWE feedback was compatible with traditional feedback in leading learners to work on their own errors.

In addition to the similarities, some differences between the two feedback types were also reported by the students. Among the differences favoring the use of AWE, it was mentioned by the students that AWE feedback was more available compared to teacher feedback. The reason might be that prior to the introduction of AWE feedback, the students in the study were heavily dependent on teachers' feedback which can be only received at school. While receiving only teachers' feedback, students may face some challenges, such as time constraints and limited feedback depending on teachers' burden. Instead, they can access an AWE tool whenever they need to. Additionally, the nature of feedback was reported to be different from teachers' feedback since AWE feedback was more explanatory compared to teachers' feedback. This may have occurred due to the fact that the corrective formative feedback is provided through correction symbols whereas the AWE feedback specifies the full name of the error made by the students in addition to explanations appearing on errors. Hence the AWE feedback may have perceived as more straightforward compared to traditional correction symbols.

In terms of the scope of feedback, although AWE feedback was found abundant regarding certain categories, such as grammar, usage, and mechanics, the lack of feedback in some aspects of writing such as organization and content was also pointed out by the participants. This is in a similar vein with Zhang and Hyland (2018) who found out that the focus of the two feedback types was different as AWE feedback was rather form-focused whereas the teacher's feedback focused more on content and organization. This contradicts with Warschauer and Grimes' study (2008) which revealed that the AWE feedback had a positive influence on students' performance in terms of content. Another drawback pointed out by the students of this study was that they experienced difficulty in correcting certain errors such as short sentences and repetition of words. Therefore, when asked what type of feedback they would prefer to receive, the majority of the students expressed their willingness to receive a combination of both feedback types. Although the students were satisfied with AWE formative corrective feedback given on errors, some expressed willingness to receive teacher feedback concerning meaning and organization. It can therefore be assumed that AWE feedback may allow learners to receive formative corrective feedback through the system at all times and may ease the

burden of teachers by allowing them to focus only on the organization and meaning of the written text.

Taking all these points into consideration, it can be indicated that both teacher feedback and AWE feedback have advantages and disadvantages. Hence, no single feedback type meets all students' needs. However, as mentioned by the students, a combination of both feedback types may cater to students' needs and provide them with more opportunities to receive feedback in addition to increasing the quality of feedback. In fact, the study by Wilson and Cziki (2016) which investigated the effects of combining teacher feedback with AWE feedback on students' writing motivation revealed that the students who received a combination of AWE feedback and teacher feedback were more motivated to revise and improve their writing compared to the teacher-feedback-only group. Additionally, another study conducted by Palermo and Thomson (2018) revealed that when an AWE tool is combined with conventional writing instruction, students were able to compose essays in higher quality compared to the group receiving solely conventional feedback. Likewise, when Mohsen and Abdulaziz (2019) combined two feedback types in comparison to an AWE feedback-only condition, it was found that a hybrid mode led to a significant improvement in L2 writing

Since the number of studies investigating the effectiveness of the combination of AWE feedback and teacher feedback is rather small, more studies would be required to further explore the effectiveness of combining feedback types rather than using only teacher feedback or AWE feedback. Nonetheless, both the students' statements in the current study and the previous findings in the literature put forward that combining AWE feedback with a different type of feedback (i.e. traditional teacher feedback) that could focus on different aspects of writing such as the organization and content could increase the quality of feedback and boost students' L2 writing performance. Hence, future studies are needed to better determine whether a hybrid mode would be ideal for both students and teachers since such claims are beyond the scope of the current study.

5.4. EFL Instructors' Views on the Use of AWE in Teaching and Assessment of Writing in L2

Even though revealing the challenges in L2 writing was not one of the purposes of the study, when teachers were asked to reflect on AWE tool, one of the themes emerging from the interviews was the challenges faced by the teachers in teaching L2

writing. It was highlighted many times during the interviews that providing feedback to a classroom full of students in a short time is a quite demanding task, and the teachers admitted having difficulties during the practice stage and revision stages of writing due to time constraints and teachers' burden. Besides, providing quality feedback is a time-consuming task and teachers often feel tiresome when they give feedback on multiple drafts on many students. Furthermore, the challenges faced in teaching and revising written texts can also affect students' achievement negatively as the quality of feedback can decrease and some errors might be overlooked due to the factors mentioned earlier.

On the other hand, reflecting on their experiences of the AWE tool and AWE feedback and their impression of them, the teachers stated that they were satisfied with the quality and quantity of AWE feedback as they were impressed with the AWE tool's ability to detect errors, especially in grammar, usage, and mechanics categories. Moreover, the type of indirect, explanatory and unfocused feedback provided by the system was found favorable. Furthermore, when the teachers compared the feedback they gave to essays and the feedback provided by the AWE system, most of them realized that some errors went unnoticed within their evaluation which provides further evidence that some problems pertaining to the adequacy and quality of teacher feedback may occur due to time constraints in classroom settings (Gorgogeta & Vlachos, 2019). Regarding the importance of detecting errors and the scope of corrective feedback, the instructors stated that it is advantageous to detect as many errors as possible since providing corrective feedback affects students' improvement positively. Likewise, in many studies EFL learners prefer teachers to correct all of their errors and inform them about different aspects of their writing performance (Amrhein & Nassaji, 2010; Üstünbaş & Çimen, 2016).

Instructors reported some major advantages of utilizing AWE feedback as the availability of the system, its providing instant feedback, matching some students' preferences, and helping learners gain autonomy. Taking all these points into consideration, it can be presumed that AWE can be used as a form of assessment that holds the potential for decreasing students' dependence on classroom assessment and promoting self-regulated learning while allowing teachers to improve writing instruction. That is, the active involvement of both students and teachers is crucial for the revision stage of writing, and an increase in students' involvement can facilitate teachers' assessment practices. In fact, to reduce teachers' burden, some AWE programs which are

designed for classroom use are currently offered to students free of charge in some countries (Lee, 2008; Matsumoto & Akahori, 2008; Palermo & Thomson, 2019). Li et al. (2015) reported that the implementation of AWE changed the way the teachers responded to papers positively. Since they saved a considerable amount of time by not providing feedback on some areas, such as grammar and mechanics, they had more time to focus on content and organization aspects of writing. To conclude, AWE feedback may hold the potential to help learners become more autonomous and decrease teachers' burden.

Although many advantages of using AWE were highlighted by the instructors, it was also noted that using AWE feedback has some disadvantages, such as lack of feedback on certain errors and lack of assessment of certain aspects of writing. Likewise, the teachers in the study conducted by Li et al. (2015) pointed out that AWE feedback needs to be improved in terms of its response to content and organization. The reason might be that even though AWE tools can provide diagnostic feedback on content and organization (Schultz, 2013), the feedback provided in these categories may have been found insufficient and general. It can be assumed that individualized feedback on these categories could be preferred by both students and the teachers.

When the instructors were asked to compare the scores and the feedback they provided for essays with the ones provided by the system, they noted the similarity between their own scores and AWE scores. In addition, they highlighted the advantage of the explanations provided along with the scores given by the system. Conversely, the findings of the study conducted by Li et al. (2014) showed that the instructors agreed with the low grades given by AWE; however, they disagreed with the high grades stating that AWE scored some papers higher than they would. This may have occurred since the rubric used in the study conducted by Li et al. (2014) was an analytical rubric. On the other hand, in the current study both the system and the teachers in the study context used holistic rubrics to score the essays. Moreover, the resemblance of the explanations included on the rubric used in the study context and the rubric on the system were found compatible by the teachers. As both quantitative and qualitative findings regarding the correlation between the scores given by the instructors and the AWE system indicate similarity, AWE scores can be utilized by the teachers or the students if needed. However, given that our findings are based on limited data, the results concerning the results may not be generalizable to other settings.

In addition to the similarities observed by the instructors, some differences between the two feedback types were also identified. These were the additional error types detected by the system and the nature of the feedback type provided by the system. To put it differently, the instructors expressed that while they use one correction symbol to identify a general category, the AWE tool identifies many other error types under a broad category. To illustrate, while the teachers mark the punctuation errors with the letter "p", the system identifies sub-categories, such as missing comma, extra comma, and missing apostrophes. Although this difference was perceived positively by the instructors, it was also stated that teachers' feedback can be more personalized as teachers can add motivational comments, suggestions, and other remarks. Likewise, Zhang and Hyland (2018) pointed out that AWE feedback lacks some characteristics of teacher feedback such as motivational comments and recommendations for further practice.

Taking all these points into consideration, when asked about their opinion of the use of AWE in process writing stages and their preferences regarding the implementation of such tools, the instructors expressed their willingness to use AWE. They stated that they would like students to utilize AWE feedback outside the classroom in a supportive way. Additionally, they noted that they might consider implementing AWE feedback in different stages of process writing or they would utilize the AWE feedback to lessen their workload. In other words, although the teachers expressed that the AWE feedback has many advantages and offers a similar type of feedback, there were some drawbacks of the system. Hence, their preference of the implementation of AWE was more of a hybrid mode of feedback which combines both AWE feedback and teacher feedback. Although the teachers in this study did not use AWE feedback in their classrooms for long periods of time, the findings in the literature (Li et al., 2015; Link et al., 2014; Warschauer & Grimes, 2008) point to the likelihood of the instructors' presumptions to be correct. To illustrate, Link et al. (2014) found that five teachers who implemented AWE in their writing lessons for a year and an AWE in writing instruction and various stages of writing, such as pre-writing, peer review, and revision expressed that the system was competent in finding and dealing with grammatical and mechanical issues of writings. Moreover, the instructors mentioned that they found AWE programs useful for boosting learners' metalinguistic ability, reducing teachers' workload, and providing feedback on sentence-level. The findings of the current study and the previous studies regarding the teachers' perceptions towards AWE are crucial since teachers are the ones who can implement

innovative tools in a classroom setting, and how teachers perceive the use of innovative technologies will impact their classroom practices. Thus, for students to benefit from innovative tools, such as AWE tools, teachers need to acknowledge the potential benefits of implementing such tools and their attitude towards these tools should be positive as well as the students' attitudes.

6. CONCLUSION

6.1. Summary of the Findings

The aim of this research was to develop an understanding of the role AWE might play in the Turkish EFL context in terms of assessment of L2 writing. Therefore, the current study was designed to investigate 1) the correlation between AWE scores and instructors' holistic scores given to students' essays, 2) the impact of AWE formative feedback on writing accuracy, 3) the instructors' views on AWE scores, and the use AWE feedback in L2 writing and 4) the students' views on the use of AWE for formative assessment of L2 writing. In an effort to examine the effectiveness of using AWE in L2 writing, an AWE tool was implemented in the classroom assessment to provide feedback to the 18 students in the experimental group, whereas 20 students in the control group were provided with teacher's feedback during the revision stage of process writing. In order to determine the effectiveness of both feedback types in reducing errors and whether there was a significant difference between the feedback types regarding their effect on improving writing accuracy, the change in accuracy from first drafts to second drafts of each essay were compared within groups with the Wilcoxon signed-rank test and between groups with Mann- Whitney U test. Furthermore, in an effort to examine the students' views on using AWE feedback in L2 writing assessment and utilizing AWE feedback to improve writing accuracy, written reflection reports were collected every week and semi-structured interviews were conducted with the students. Both written reflection reports and the semi-structured interviews were analyzed by using the Constant Comparison Method.

In addition to the use of AWE by students, seven instructors were trained in the use of AWE in writing assessment. Following the training, essays that were assessed by teachers were uploaded to the AWE system to determine the correlation coefficient between the scores given by the teachers and the AWE tool for the same papers. For this purpose, the Spearman correlation coefficient was used for statistical analysis. Moreover, for teachers to reveal their views on the use of AWE in L2 writing and to reflect on the AWE scores and AWE feedback and compare them to the scores and feedback they gave students on their writing semi-structured interviews were conducted. The qualitative data gathered through semi-structured interviews with teachers were analyzed by using the Constant Comparison Method.

The results showed that the correlation coefficient between the scores of AWE and instructors' scores was significant and positively high ($r = .862$, $p < .05$), meaning the scores of two variables increased and decreased together. It was also found that the correlation coefficient between the instructors' scores and the external rater, who was the eighth instructor participating only in the scoring process, was significantly and positively high ($r = .804$, $p < .05$) which means that two variables increased and decreased simultaneously. Furthermore, when the correlation coefficients between the scores given by each instructor and AWE were compared to the correlation coefficients between the scores given by the instructors and the external rater, the results indicate that in some cases the correlation between the scores given by AWE and the scores given by the instructors was higher than the correlation between the instructors' scores and the external rater's scores given to students essays. In addition to the quantitative data analysis, the qualitative data analysis of the semi-structured interviews revealed that the instructors' perceptions regarding the AWE scores were positive as the AWE scores, intervals and the accompanying explanations on the holistic criteria were found similar to the scores they gave for the essays and the holistic rubric used in the study context. Taken together, the results suggest that in addition to the agreement between the external rater and the instructors, there was an agreement between the AWE scores and the scores given by the instructors. In addition, there are similarities observed by the teachers in terms of the similarity of the scores, the score intervals, and the rationales for the scores in each interval as well as the points taken into consideration when scoring.

With regard to the efficiency of teacher's feedback and the AWE feedback on reducing errors in 11 error types and improving writing accuracy, the change in writing accuracy from the first draft to second drafts of three different essays written by the experimental group and the control group were compared both between and within groups. To begin with, the results of Wilcoxon signed-rank tests show that both the feedback that was given by the AWE system and the traditional feedback given by the teachers were effective in reducing all of the 11 error types included in the analysis. Moreover, the results of the Mann-Whitney U tests revealed that although the AWE feedback was more effective in correcting three error types which are ill-formed verbs, punctuation, and article errors according to the first within-group analysis, the next two withing group analyses revealed that AWE was more effective in correcting only punctuation and article errors. Therefore, it can be concluded that AWE feedback was

more effective in correcting errors in the mechanics and usage categories than teacher feedback.

In addition to the statistical analyses of quantitative data, the students were asked to reflect on their views on using AWE to revise their essays and improve writing accuracy by using AWE. The qualitative data were collected through weekly reflection reports and semi-structured interviews and it was found that the supplementary qualitative data analyses support the results of the quantitative data analysis as the results of the qualitative data analyses showed that the AWE feedback was not only effective in reducing errors in the categories of grammar, usage, and mechanics, but it was also more effective in correcting certain the errors in the mechanics and usage categories, such as article and punctuation errors than other types of errors according to the students. All in all, both the quantitative and qualitative data that were collected to explore the effectiveness of AWE feedback in correcting errors in L2 writing show that both AWE feedback and teacher feedback were effective in significantly reducing the 11 error types included in this study. Nonetheless, the AWE feedback was more effective in reducing mechanics errors compared to the teachers' formative corrective feedback and according to the students, although the AWE feedback addressed a variety of error types, most of the errors they believed they were able to correct were in the mechanics and usage categories.

Along with the use of AWE to reduce errors, the students were asked to reflect on their experience with the AWE tool and the AWE feedback as well as the teacher feedback. The student's views on the use of AWE feedback were overwhelmingly positive, and the students were generally satisfied with the use of AWE feedback to improve writing accuracy. The results showed that according to the students, there are beneficial features of the AWE tool that make this tool attractive and convenient to be used in the stages of process writing in L2. To illustrate, most students found the AWE tool useful, convenient, and easy to access, and they were satisfied with the features included in the system. In addition to the beneficial features, students expressed that AWE was also beneficial in helping them improve their writing skills. In other words, students believed that the feedback from AWE leads to an improvement in their writing accuracy and reduces the mistakes they make in L2 writing.

Moreover, the students' statements indicated that they found AWE helpful in gaining autonomy. In other words, it was put forward by the students that the diagnostic

feedback from AWE and the accompanying analysis of the feature feedback has increased their awareness of their strengths and weaknesses. The ease of access to the system and the numerous opportunities it provides for students to practice writing and receive feedback outside the school without the supervision of a teacher allowed them to gain independence and increased their learning efforts, in addition to increasing their motivation to improve their writing accuracy. When comparing the implementation of AWE and their previous experiences with teacher feedback, students also emphasized that there were some similarities and differences between the AWE assessment and teacher assessment. They pointed out that AWE feedback was similar to teacher feedback in that both types of feedback include formative, corrective feedback and both contribute to writing accuracy. It also became clear that although AWE feedback could be more explanatory in terms of formative corrective feedback, the feedback given on content and organization needs to be more directive and personalized. Therefore, some students expressed their willingness to receive a mixture of both types of feedback or to be assessed on the content and organizational aspects of their writing. Therefore, despite the advantages of using AWE in L2 writing and the similarities between the two types of assessment, some disadvantages and concerns were also mentioned.

The final research question asked how the instructors view the use of AWE in L2 writing and the teachers were asked to reflect on their own impressions of the AWE system based on their experiences of using the system for three weeks. The main purpose of the three weeks the teachers spent familiarizing themselves with the system was to help them reach a deeper understanding of the potential of using such tools in summative and formative assessment of L2 writing. The study showed that teachers face many challenges and shortcomings while assessing students' writing. Regarding the writing process, teachers discussed the benefits and drawbacks that AWE would bring to L2 writing classes. Furthermore, based on their findings, teachers compared AWE and teachers' assessments in terms of assessment, revealing similarities and distinctions. Finally, as a result of this comparison, teachers demonstrated a desire to use AWE and indicated that they would like to use AWE in the classroom and outside of the classroom in a complementary manner, if not exclusively, in writing assessment.

6.2. Conclusion

The value of feedback in helping L2 learners become proficient writers and improve their writing accuracy cannot be overstated (Hattie & Timperley, 2007; Parr & Timperley, 2010). Given that L2 learners make more word-level and sentence-level errors than L1 learners (Ranalli et al., 2017; Silva, 1993), it is widely agreed that corrective feedback can be particularly useful in improving the writing performance of L2 learners. It is also argued that the immediacy of feedback is another factor contributing to the effectiveness of the feedback provided to the learners (Dihoff et al., 2004). Moreover, in EFL writing context, students are largely dependent on teachers to receive formative corrective feedback. However, for teachers to provide formative corrective feedback to a large number of students is a time-consuming and labor-intensive part of classroom-based assessment. Hence, integrating innovative forms of assessment into L2 writing classes which can facilitate self-assessment and provide instant feedback is of utmost importance. In this regard, the potential of AWE tools to promote learning and teaching L2 writing by providing immediate formative feedback has gained considerable attention.

It has been suggested by many researchers (Lee, 2017; Page, 2003; Shermis & Burstein, 2013; Warschauer & Ware, 2006) that the AWE tools, which can score essays and provide immediate feedback on written compositions, may reduce teachers' burden while helping learners become more autonomous, allow them to improve writing accuracy and generate scores compatible to teachers' scores. The previous literature on the use of AWE in writing assessment has yielded favorable results in terms of correlation between AWE scores and human-rater scores (Attali & Burstein, 2006; El Ebyary & Windeatt, 2010; Ramineni, 2013; Shermis et al., 2002), the effectiveness of AWE feedback in improving writing accuracy (Bulut, 2019; Gençer, 2019; Li et al., 2015; Mohsen & Abdulaziz, 2019; Palermo & Thomson, 2018; Waschauer & Grimes, 2008), students views on the use of AWE (Li et al., 2014; Wilson & Czik, 2016) and teachers' views on the use of AWE (Li et al., 2015; Link et al., 2014). In a similar vein, the findings of the present study were consistent with the previous literature.

While most of the previous literature has focused on the use of AWE for testing and scoring purposes in both L1 and L2 contexts, studies exploring the use of AWE for formative assessment purposes, especially in EFL context still remains scarce. Therefore, this study contributes to the existing literature on the use of AWE in EFL context. Since the findings of the previous literature and the current study showed that AWE feedback

was effective in reducing errors in grammar, usage and mechanics, it can be expected that the use of AWE tools might become more common in classroom-based writing assessment and the interest in using AWE for providing feedback on errors may increase.

Furthermore, as the findings of this study put forward that the AWE scores were in agreement with the instructors' scores participating in the study, and both AWE feedback and teacher feedback helped learners reduce errors, it can be concluded that AWE appears to be a tool which can be utilized in L2 writing assessment. Thus, the results indicate that integrating AWE tools into the summative and formative assessment of L2 writing can reduce teachers' workload while allowing learners to assess their writing performance and monitor their L2 writing development by promoting learner autonomy. Considering the fact that L2 learners would benefit from immediate formative feedback and plentiful opportunities for practicing writing and revising their written compositions, using AWE tools can promote learning.

In addition to the quantitative results, the qualitative results of the study put forward that some of the advantages of using the AWE tool were motivating the students, facilitating self-study and self-assessment, promoting learning, and gaining autonomy. Therefore, the results of this study support that tools such as AWE can be used in EFL environments in a detailed and comprehensive manner. In other words, such tools can be used not only in the classroom environment but also outside the classroom. Although some issues with AWE feedback were pointed out by the students and teachers, the participants' views on the AWE use were mainly positive. Hence, the results indicated that even though AWE feedback may not replace the teacher feedback altogether, it can complement teacher feedback in a way that enables teachers to focus on global issues of writing, such as content and organization. Such use of AWE can allow teachers to save time while giving feedback and provide more effective feedback on other aspects of writing while AWE feedback is utilized to provide feedback on errors.

In conclusion, the importance of finding useful and effective tools that provide formative and summative assessment cannot be denied in the new era of digitalization that has entered our lives with the recent developments and gained momentum due to the current pandemic process. Considering the importance of students getting feedback from reliable sources when they are on their own, it is important to use effective tools. In this respect, students can work autonomously, and they can support this process by using tools such as AWE. Therefore, in order to acknowledge the potential of innovative technologies

and use the necessary tools in writing, introducing tools such as AWE is of great importance, especially in EFL writing. Apart from that, encouraging students to use AWE-like tools outside the classroom can allow learners to become autonomous and have more opportunities to assess themselves and receive feedback on their writing. The Table 6.1 below summarized the conclusion of the current study.

Table 6.1. *The conclusion of the study*

Research questions	Conclusion
1) Is there a relationship between AWE holistic scores and instructors' holistic scores given to students' essays?	• The correlation between the AWE tool and instructors was similar to the correlation between an instructor and the external rater • In some cases, the correlation between AWE scores and instructors scores were higher than the correlation between two instructors • The AWE tools, which can score essays, may reduce teachers' burden by generating scores compatible with teachers' scores
2) Is there a significant difference in terms of writing accuracy between a group of students receiving AWE formative feedback with a group of students receiving teacher feedback?	• The most common errors were mainly in usage and mechanics categories • Both AWE feedback and teacher feedback helped learners reduce errors in grammar, usage and mechanics • The AWE feedback was more effective in using certain errors in usage and mechanics categories which were among the most common errors • AWE appears to be a tool that can be utilized in formative assessment of L2 writing and it can reduce teachers' workload by providing effective feedback on common errors
3) How do EFL students view the use of AWE in L2 writing?	• There were some of the advantageous features of the AWE tool making it appealing to the students • According to the students the AWE tool facilitated self-study and self-assessment, promoted learning, and allowed them to gain autonomy. • The tool helped learners improve writing accuracy and monitor their L2 writing development by promoting learner autonomy. • Such tools can be used not only in the classroom environment but also outside the classroom
4) How do EFL instructors view the use of AWE in L2 writing?	• Even though AWE feedback may not replace the teacher feedback altogether, it can complement teacher feedback in a way that enables teachers to focus on global issues of writing, such as content and organization • AWE can allow teachers to save time while giving feedback and they provide more effective feedback on other aspects of writing while AWE feedback is utilized to provide feedback on errors • Teachers can utilize the AWE feedback not only for classroom-based assessment but also for allowing learners to assess themselves outside the classroom

6.3. Implications

Drawn from the aforementioned findings of the current study, some pedagogical implications for students, teachers, school administrators, teacher educators, curriculum designers, and tool developers might be suggested. To begin with, in Turkey, where English is taught as a foreign language, although students start learning English starting from primary school, little attention is paid to productive skills as speaking and writing skills are mostly excluded in high-stake exams and students' main motive to learn English is to pass exams (Eryılmaz & Yeşilyurt, 2020). Therefore, Turkish students have little experience in writing in English let alone academic writing, and writing in L2 might be significantly more difficult for EFL students compared to ESL and NES students. That said, EFL students practice writing in L2 mostly when they attend an English preparatory program of universities. Even though students practice writing and receive formative feedback regularly throughout the year spent in such programs, the assistance and guidance they receive regarding their writing ability is restricted to the time they spend at school and the main or the only source of formative feedback is the instructors. Hence, a one-year program may not be sufficient to improve EFL students' writing proficiency significantly without additional support or guidance. Moreover, the challenges EFL students face while writing in L2 might continue throughout their academic education. In this respect, since the findings show that AWE is compatible with teacher feedback in correcting errors and can be effective in reducing the types of errors made quite often by students, students can use AWE's feedback to learn about the errors they make frequently. Moreover, AWE feedback can be utilized to realize weaknesses and strengths thanks to trait feedback analysis and diagnostic feedback. In addition, because AWE is easily accessible and suitable to be used outside the classroom, students can use this system for additional practice and multiple revisions. Additionally, AWE can be used by students while continuing their studies at the faculty to check and correct errors and improve linguistic accuracy in written texts, such as assignments and research papers. With all these possibilities, AWE can help students gain autonomy and engage in self-improvement activities without depending on teachers.

Moreover, it was revealed that before this study took place, none of the teachers were fully aware of AWE tools and their capabilities. Nevertheless, shortly after this study was conducted, the COVID-19 pandemic has required most instructors to move writing instruction, practice, and assessment online and adapt the processes of teaching writing

to online education by implementing technology to enhance the efficacy of writing lessons at promoting learning. Although new challenges have emerged due to this rapid and radical transformation, this experience provided us with an opportunity to explore the potential of integrating technology in EFL writing. Thus, it is more clear now than ever that educational technology needs to be used more pervasively and effectively in writing lessons. One of the implications of this study regarding teachers is that teachers may receive great benefits when assistive technology such as AWE tools are integrated into the summative and formative assessment of L2 writing. By doing so, teachers can reduce their workload significantly. Moreover, teachers can use the trait feedback analysis provided by the AWE tools to identify students' weaknesses and strengths to adapt their approaches to students' needs accordingly. In addition to students' needs, teachers can acknowledge students' preferences regarding feedback by employing AWE tools as an automated and easily accessible tool that might improve students' self-regulated learning effort by motivating them. Regarding the use of AWE formative feedback, teachers can use AWE to provide corrective feedback during in-class writing stages, and thus they can save a lot of time by giving individual or corrective feedback to students on other aspects of writing, such as meaning and organization. In addition, teachers can encourage students to use AWE outside the classroom by assigning writing tasks, as the opportunities for practice and revision in the classroom are limited due to the crowd of the classes and the intensity of the curriculum. Additionally, the possibilities that such systems can provide for writing instruction and peer-feedback can be explored and integrated into different writing stages. In addition to formative assessment, AWE scores on summative assessment can be used to ensure inter-rater agreement and resolve discrepancies between teachers if needed.

The study also suggests implications for school administrators. To begin with, although almost all AWE systems are paid today, such systems can become free of charge if the use of these systems becomes widespread in the future. Taking this into account, teachers should be familiarized with the use of such tools with the help of subsequent training programs. Considering the potential benefits of these systems for teachers and students and the likelihood of popularization of such tools, the gradual integration of such tools into writing instruction, practice, and assessment administrators to integrate such systems into formative assessment will provide great convenience to all users in the future. In addition, although these systems are paid, they can be used for a more affordable

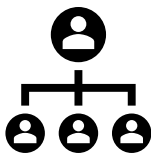
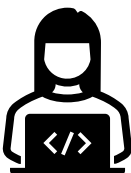
fee within the school-wide usage plans. Moreover, informative resources and technical support should be provided. In this regard, school administrators have a great responsibility for the widespread use of these systems both inside and outside of the classroom, considering the benefits of these systems for teachers and students. In order to facilitate the integration of AWE tools in classroom-based L2 writing, school administrators can equip schools with computer laboratories to allow students to access such tools at schools.

As mentioned earlier, the teachers who participated in this study were unaware of AWE tools and what they can offer. Although most of the teachers graduated from English language teaching departments and can be considered as experienced instructors, their lack of knowledge considering the existence of such assistive tools puts forward that teacher educators might consider implementing pre-service and in-service training programs for teacher candidates and teachers to equip them with knowledge and skills to integrate such tools in L2 writing lessons. Moreover, as a part of professional development, teacher educators need to mentor and support in-service teachers in applying technology standards and integrate technology-mediated instruction and assessment into writing classes.

Moreover, to support in-service teachers to effectively integrate AWE tools into writing classes, curriculum designers should acknowledge the potential of using such tools while designing syllabi and curriculum suitable for transforming educational practices. In other words, given that flexible educational settings and online learning environments have become an essential component of higher education programs, curriculum developers should design online writing tasks by integrating AWE tools in L2 writing stages, such as pre-writing, planning, practice, and revision stages. By doing so, curricula can be modified in light of the emerging needs of the students.

Implications drawn from this study suggest that tool developers can add features that better address the needs of EFL students. As the AWE tools are commonly used by NES and ESL context, potential difficulties that EFL students with lower-level language proficiency might face should be acknowledged and the terms and explanations on such tools should be simplified so that these tools can be utilized by lower-level EFL students as well. Furthermore, multiple scoring rubrics can be offered in order to allow users to choose the most suitable rubric concerning their settings and expectations. The Table 6.2 below summarized the implications of the current study.

Table 6.2. *The implications of the study*

Implications	
 Students	• Increasing awareness of their strengths and weaknesses through AWE feedback • Utilizing AWE feedback to reduce errors in L2 writing • Using AWE tools outside the classroom for self-assessment and self-improvement purposes • Receiving assistance from AWE tools throughout their academic education
 Teachers	• Combining AWE scores with instructors scores to ensure inter-rater agreement or resolve discrepancies • Trying out using such tools in writing instruction, practice and assessment • Acknowledging students' preferences regarding the type of formative feedback • Encouraging students to use AWE feedback outside the school • Combining AWE feedback with teacher feedback inside the school
 School Administrators	• Designing subsequent training programs • Integrating such tools into education gradually • Purchasing school-wide usage plans • Providing informative resources and technical support • Equipping schools with computer laboratories
 Teacher Educators	• Equipping pre-service and in-service teachers with knowledge and skills to integrate such tools in L2 writing lessons • Designing pre-service and in-service training programs for teacher candidates and teachers • Mentoring and supporting in-service teachers in applying technology standards
 Curriculum Designers	• Acknowledging the potential of using such tools while designing syllabi and curriculum • Designing online writing tasks by integrating AWE tools in L2 writing stages, such as pre-writing, planning, practice and revision stages
 Tool developers	• Simplifying the terms and explanations on such tools for lower-level EFL students • Integrating multiple scoring rubrics to provide more options

6.4. Suggestions for Further Research

The findings of the current study provide an insight into the use of AWE tools in the EFL context. Nevertheless, as the research investigating the use of AWE tools in the Turkish EFL context remains limited, future studies could explore this issue further. In future studies, exploring the differences between receiving a combination of AWE and teacher feedback, and utilizing either one of these feedback types in isolation might prove important. Investigating the effectiveness of a hybrid feedback mode and views of both students and teachers on combining teacher feedback with AWE feedback might be explored. Correspondingly, further research may reveal whether there is a significant difference between the effectiveness of using these two types of feedback in L2 writing assessment, and the effect of integrating both feedback types on writing improvement can fill the gap in the Turkish EFL context.

In addition to assessment in L2 writing lessons of AWE tools, the instructional use of these systems and their effectiveness on teaching and learning writing in L2 should be investigated. That said, teachers can be provided with the opportunity to integrate these tools into classroom use in different ways, and teachers' opinions about their preferences and different ways of using these systems can be investigated.

Further studies may also explore the use of AWE by students outside the classroom, as well as the effect of using it in a physical classroom setting on improving. By doing so, the relationship between AWE tools and the factors that have great effects on learning such as autonomy and motivation can be examined.

REFERENCES

- Ahangari, S., Hejazi, M., & Razmjou, L. (2014). The impact of scaffolding on content retention of Iranian post-elementary EFL learners' summary writing. *Procedia-Social and Behavioral Sciences*, 98 (1), 83-89.
- Allen, L. K., Crossley, S. A., Snow, E. L., & McNamara, D. S. (2014). L2 writing practice: Game enjoyment as a key to engagement. *Language Learning & Technology*, 18(2), 124-150.
- Allen, L. K., Jacovina, M. E., & McNamara, D. S. (2016). Computer-based writing instruction. In C. A. MacArthur, S. Graham, & J. Fitzgerald (Eds.), *Handbook of Writing Research* (pp. 316–329). New York, NY: The Guilford Press
- Alshakhi, A. (2019). Revisiting the writing assessment process at a Saudi English Language Institute: Problems and solutions. *English Language Teaching*, 12(1), 176-185.
- Amrhein, H. R., & Nassaji, H. (2010). Written corrective feedback: What do students and teachers think is right and why?. *Canadian Journal of Applied Linguistics*, 13(2), 95-127.
- Assessment Reform Group. (2002). *Assessment for learning: 10 Principles*. Cambridge: University of Cambridge.
- Ata, M. & Tıkız, G. (2019). The voice of college writing instructors on the academic challenges of the students in preparatory classes: A case study in a Turkish EFL context. *Journal of Foreign Language Education and Technology*, 5(1), 67-85.
- Atkinson, D. (2003). L2 writing in the post-process era: Introduction. *Journal of Second Language Writing*, 12(1), 3-15.
- Attali, Y. (2004). Exploring the feedback and revision features of Criterion. *Journal of Second Language Writing*, 14 (1), 191-205.
- Attali, Y., & Burstein, J. (2005). *Automated essay scoring with e-rater v. 2.0 (ETS RR-04-45)*. Princeton, NJ: Educational Testing Service.
- Attali, Y., & Burstein, J. (2006). Automated essay scoring with e-rater® V. 2. *The Journal of Technology, Learning and Assessment*, 4(3), 1-31.

- Attali, Y., & Powers, D. (2008). *Effect of immediate feedback and revision on psychometric properties of open-ended GRE Subject Test items: GRE Board Research Report*. Princeton, NJ: ETS.
- Bacha, N. N., & Bahous, R. (2008). Contrasting views of business students' writing needs in an EFL environment. *English for specific purposes*, 27(1), 74-93.
- Bachman, L. F., & Palmer, A. S. (1996). *Language testing in practice: Designing and developing useful language tests*. Oxford: Oxford University Press.
- Badger, R., & White, G. (2000). A process genre approach to teaching writing. *ELT Journal*, 54(2), 153-160.
- Bae, J. (2011). *Teaching process writing for intermediate/advanced learners in South Korea*. Unpublished Master's Thesis. Wisconsin: University of Wisconsin, The Graduate Faculty.
- Baehr, M. (2005). Distinctions between assessment and evaluation. *Program Assessment Handbook*, 7(1), 231-234.
- Bai, B., Wang, J., & Nie, Y. (2020). Self-efficacy, task values and growth mindset: what has the most predictive power for primary school students' self-regulated learning in English writing and writing competence in an Asian Confucian cultural context? *Cambridge Journal of Education*, 51 (1), 65-84.
- Baker, E. L., & O'Neil Jr, H. F. (1994). Performance assessment and equity: A view from the USA. *Assessment in Education: Principles, policy & practice*, 1(1), 11-26.
- Baker, S., Gersten, R., & Graham, S. (2003). Teaching expressive writing to students with learning disabilities: Research-based applications and examples. *Journal of learning disabilities*, 36(2), 109-123.
- Balfour, S. P. (2013). Assessing writing in MOOCs: Automated essay scoring and calibrated peer review™. *Research & Practice in Assessment*, 8 (1), 40-48.
- Baron, N. S. (2020). Talking, reading, and writing on smartphones. In Ling, R., Fortunati, L., Goggin, G., Lim, S. S., & Li, Y. (Eds.), *The Oxford Handbook of Mobile Communication and Society* (pp. 403-422). New Yorks: Oxford University Press.

- Bazerman, C. (1988). *Shaping written knowledge: The genre and activity of the experimental article in science*. Madison: University of Wisconsin Press.
- Biber, D., Douglas, B., Conrad, S., & Reppen, R. (1998). *Corpus linguistics: Investigating language structure and use*. Cambridge: Cambridge University Press.
- Biber, D., Nekrasova, T., & Horn, B. (2011). *The effectiveness of feedback for L1-English and L2 writing development: A meta-analysis. TOEFL iBT Research Report*: Princeton, NJ: Educational Testing Service.
- Bijami, M., & Raftari, S. (2013). Product, process, and genre approaches to writing: A comparison. *The Iranian EFL Journal*, 9(15), 9-18.
- Black, L., Daiker, D. A., Sommers, J., & Stygall, G. (1994). *Handbook of writing portfolio assessment: A program for college placement*. Oxford, OH: Miami University.
- Black, P., Harrison, C., Lee, C., Marshall, B., & Wiliam, D. (2004). Working inside the black box: Assessment for learning in the classroom. *Phi delta kappan*, 86(1), 8-21.
- Bloom, B. S. (1956). Taxonomy of educational objectives. Vol. 1: Cognitive domain. *New York: McKay*, 20, 24.
- Boud, D. (1995). Assessment and learning: contradictory or complementary? In Knight, P. (Ed). *Assessment for Learning in Higher Education* (pp. 35-48). London: Kogan Page.
- Brown, G. A., Bull, J., & Pendlebury, M. (2013). *Assessing student learning in higher education*. New York: Routledge.
- Brown, G. T. (2004). Teachers' conceptions of assessment: Implications for policy and professional development. *Assessment in Education: Principles, Policy & Practice*, 11(3), 301-318.
- Brown, H. D. (2007). *Teaching by principles: an interactive approach to language pedagogy*. New York: Pearson Education.
- Buitrago, C. R., & Díaz, J. (2018). Flipping your writing lessons: Optimizing time in your efl writing classroom. In J. Mehring & A. Leis (Eds.), *Innovations in Flipping the Language Classroom* (pp. 69-91). Singapore: Springer.
- Bulut, S. E. (2019). *The effects of automated writing evaluation on EFL students' writing achievement and motivation towards writing*. Unpublished Master's Thesis. İstanbul: İstanbul University, Department of English Language Education.

- Burstein, J., & Chodorow, M. (1999). Automated essay scoring for nonnative English speakers. Retrieved from https://origin-www.ets.org/Media/Research/pdf/erater_acl99rev.pdf
- Burstein, J., Chodorow, M., & Leacock, C. (2004). Automated essay evaluation: The Criterion online writing service. *Ai magazine*, 25(3), 27-27.
- Buttram, J. L. & Covert, R. W. (1977). *The discrepancy evaluation model: A systematic approach for the evaluation of career planning and placement programs*. Virginia: Evaluation Research Center.
- Cai, L. J. (2013). Students' perceptions of academic writing: A needs analysis of EAP in China. *Language Education in Asia*, 4 (1), 5-22.
- Calvo, R. A., & Ellis, R. A. (2010). Students' conceptions of tutor and automated feedback in professional writing. *Journal of Engineering Education*, 99(4), 427-438.
- Cen, Y., & Ross, H. (2009). Chinese higher education and evaluation in context: Guest editors' introduction. *Chinese Education & Society*, 42(1), 3-7.
- Chapelle, C. A., Cotos, E., & Lee, J. (2015). Validity arguments for diagnostic assessment using automated writing evaluation. *Language testing*, 32(3), 385-405.
- Charmaz, K. (2006). *Constructing grounded theory: A practical guide through qualitative analysis*. London: SAGE.
- Chaudron, C. (1988). *Second language classrooms: Research on teaching and learning*. Cambridge: Cambridge University Press.
- Chen, C. F. E. (2006). The development of e-mail literacy: From writing to peers to writing to authority figures. *Language Learning & Technology*, 10(2), 35-55.
- Chen, C. F. E., & Cheng, W. Y. E. C. (2008). Beyond the design of automated writing evaluation: Pedagogical practices and perceived learning effectiveness in EFL writing classes. *Language Learning & Technology*, 12(2), 94-112.
- Cheng, G. (2017). The impact of online automated feedback on students' reflective journal writing in an EFL course. *The Internet and Higher Education*, 34 (1), 18-27.
- Cizek, G. J. (1997). Learning, achievement, and assessment: Constructs at a crossroads. In G. D. Phye (Ed.) *Handbook of classroom assessment: Learning, adjustment, and achievement* (pp. 2-29). NY: Academic Press.

- Cizek, G. J., Andrade, H. L., & Bennett, R. E. (2019). Handbook of formative assessment in the disciplines. In H. L. Andrade, R. E. Bennett, and G. J. Cizek (Eds.), *Formative Assessment : History, Definition, and Progress* (pp. 1-17). New York, NY: Routledge.
- Cohen, Y., Ben-Simon, A., & Hovav, M.(2003). The effect of specific language features on the complexity of systems for automated essay scoring. *Paper presented at the international association of educational assessment annual conference*. Manchester, UK.
- Coniam, D. (2009). A comparison of onscreen and paper-based marking in the Hong Kong public examination system. *Educational Research and Evaluation*, 15(3), 243-263.
- Cotos, E. (2010). *Automated writing evaluation for non-native speaker English academic writing: The case of IADE and its formative feedback*. Unpublished Doctoral Dissertation. Iowa: Iowa State University, Faculty of Applied Linguistics and Technology.
- Cotos, E. (2014). *Genre-based automated writing evaluation for L2 research writing: From design to evaluation and enhancement*. New York: Springer.
- Cotos, E. (2015). Automated writing analysis for writing pedagogy: From healthy tension to tangible prospects. *Writing & Pedagogy*, 7(2), 197–231.
- Crooks, T. (2004). Tensions between assessment for learning and assessment for qualifications. *Paper presented at the Third Conference of the Association of Commonwealth Examinations and Accreditation Bodies*. Fiji: University of Otago.
- Crossley, S. A., Allen, L. K., Snow, E. L., & McNamara, D. S. (2015). Pssst...Textual features...There is more to Automatic Essay Scoring than just you! In *Proceedings of the 5th International Learning Analytics & Knowledge Conference*. Poughkeepsie, NY: ACM, (pp. 203-207).
- Cumming, A. (1989). Writing expertise and second-language proficiency. *Language learning*, 39(1), 81-135.
- Cumming, A. (1990). Metalinguistic and ideational thinking in second language composing. *Written communication*, 7(4), 482-511.
- Çakır, İ. (2010). Yazma Becerisinin Kazanılması Yabancı Dil Öğretiminde Neden Zordur? *Sosyal Bilimler Enstitüsü Dergisi*, 28, 165-176.
- Deci, E. L., & Ryan, R. M. (1985). The general causality orientations scale: Self-determination in personality. *Journal of research in personality*, 19(2), 109-134.
- Demirel, E. (2011). Take it step by step: Following a process approach to academic writing to overcome student anxiety. *Journal of Academic Writing*, 1(1), 222-227.

- Deqi, Z. (2005). The process-oriented approach to ESL/EFL writing instruction and research. *Celea Journal*, 28(5), 66-70.
- Derrida, J. (1979). Scribble (writing-power). *Yale French Studies*, (58), 117-147.
- Dickinson, L. (1995). Autonomy and motivation a literature review. *System*, 23(2), 165-174.
- Dihoff, R. E., Brosvic, G. M., & Epstein, M. L. (2003). The role of feedback during academic testing: The delay retention effect revisited. *The Psychological Record*, 53(4), 533-548.
- Dihoff, R. E., Brosvic, G. M., Epstein, M. L., & Cook, M. J. (2004). Provision of feedback during preparation for academic testing: Learning is enhanced by immediate but not delayed feedback. *The Psychological Record*, 54(2), 207-231.
- Dikli, S. (2006). An overview of automated scoring of essays. *Journal of Technology, Learning, and Assessment*, 5(1), 1-35.
- Dirgeyasa, I. W. (2016). Genre-based approach: What and how to teach and to learn writing. *English Language Teaching*, 9(9), 45-51.
- Doğan, A. (2013). *Automated essay scoring system: A reliability study*. Unpublished MA thesis. Mersin: Çağ University, Institute of Social Sciences.
- Drid, T. (2018). The fundamentals of assessing EFL writing. *Psychological & Educational Studies Review*, 20(20), 14-14.
- Earl, L., & Katz, S. (2006). Rethinking classroom assessment with purpose in mind. *Winnipeg, Manitoba: Western Northern Canadian Protocol*.
- El Ebyary, K., & Windeatt, S. (2010). The impact of computer-based feedback on students' written work. *International Journal of English Studies*, 10(2), 121-142.
- Elliott, S. (2003). Intellimetric™: From here to validity. In M. D. Shermis & J. C. Burstein (Eds.), *Automated essay scoring: A cross-disciplinary perspective* (pp. 71-86). Mahwah, NJ: Lawrence Erlbaum Associates.
- 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.
- Emig, J. (1971). *The composing processes of twelfth graders*. Urbana, IL: National Council of Teachers of English
- Enright, M. K., & Quinlan, T. (2010). Complementing human judgment of essays written by English language learners with e-rater® scoring. *Language Testing*, 27(3), 317-334.

- Ericsson, P. F. (2006). The meaning of meaning: Is a paragraph more than an equation? In P. F. Ericsson & R. H. Haswell (Eds.), *Machine scoring of student essays: Truth and consequences* (pp. 28–37). Logan: Utah State University Press.
- Eryılmaz, A., & Yeşilyurt, Y. E. (2020). Foreign language writing as a developmental process (foundation, expansion, development, and completion): The FEDCom Model. *International Journal of Contemporary Educational Research*, 7(2), 307-334.
- Evans, C. (2013). Making sense of assessment feedback in higher education. *Review of educational research*, 83(1), 70-120.
- Ezza, E. S. (2010). Arab EFL learners' writing dilemma at tertiary level. *English Language Teaching*, 3(4), 33-39.
- Faerch, C., & Kasper, G. (1986). One learner--two languages: Investigating types of interlanguage knowledge. In J. House & S. Blum-Kulka (Eds.), *Interlingual and intercultural communication* (pp. 211-227). Tuebingen: Gunter Narr.
- Farris, C. R. (1987). Current composition: beyond process vs. product. *The English Journal*, 76(6), 28-34.
- Ferris, D. R. (2002). *Treatment of error in second language writing classes*. Ann Arbor, MI: University of Michigan Press, 123.
- Ferris, D. R. (2003). Responding to writing. In B. Kroll (Ed.), *Exploring the dynamics of second language writing* (pp. 119-140). Cambridge: Cambridge University Press.
- Ferris, D. R. (2004). The “grammar correction” debate in L2 writing: Where are we, and where do we go from here?(and what do we do in the meantime...?). *Journal of second language writing*, 13(1), 49-62.
- Ferris, D. R., (2006). Does error feedback help student writers/new evidence on the short- and long-term effects of written error correction. In: Hyland, K., Hyland, F. (Eds.), *Feedback in Second Language Writing: Contexts and Issues* (pp. 81–104). Cambridge: Cambridge University Press.

- Ferris, D. R. (2010). Second language writing research and written corrective feedback in SLA: Intersections and practical applications. *Studies in Second Language Acquisition*, 181-201.
- Ferris, D. R. (2014). Responding to student writing: Teachers' philosophies and practices. *Assessing Writing*, 19(1), 6–23.
- Ferris, D. R., & Hedgcock J. S. (2005). *Teaching ESL composition: Purpose, process, and practice*. Mahwah, NJ: Lawrence Erlbaum.
- Ferris, D. R., & Roberts, B. (2001). Error feedback in L2 writing classes: How explicit does it need to be?. *Journal of second language writing*, 10(3), 161-184.
- Feuerstein, R., & Klein, P. S., & Tannenbaum, A. J. (1991). Mediated learning experience: A theoretical review. In R. Feuerstein, P. S. Klein, & A. J. Tannenbaum (Eds.), *Mediated learning experience (MLE): Theoretical, psychosocial and learning implication* (pp. 3–51). London: Freund.
- Fleissner, A. M. (2006). *Writing class: How perceived social background influences high school English teachers teaching writing*. Unpublished Doctoral Dissertation. New York: New York University, Department of Teaching and Learning.
- Flick, U. (2013). *The SAGE handbook of qualitative data analysis*. London: SAGE.
- Flower, L., & Hayes, J. R. (1981). A cognitive process theory of writing. *College composition and communication*, 32(4), 365-387.
- Foltz, P. W., Streeter, L. A., Lochbaum, K. E., & Landauer, T. K. (2013). Implementation and applications of the Intelligent Essay Assessor. In M. D. Shermis, & J. Burstein (Eds.), *Handbook of automated essay scoring: Current applications and future directions* (pp. 68–88). New York: Routledge.
- French, A. (2020). Academic writing as identity-work in higher education: forming a 'professional writing in higher education habitus'. *Studies in Higher Education*, 45(8), 1605-1617.
- Frodesen, J., & Holten, C. (2003). Grammar and the ESL writing class. In B. Kroll (Ed.), *Exploring the dynamics of second language writing* (pp. 141–161). Cambridge: Cambridge University Press.

- Fukuta, J., Tamura, Y., & Kawaguchi, Y. (2019). Written languaging with indirect feedback in writing revision: is feedback always effective?. *Language awareness*, 28(1), 1-14.
- Gabrielatos, C. (2002). *EFL writing: Product and process*. Retrieved from https://www.academia.edu/13907872/Gabrielatos_C_2002_EFL_writing_Product_and_process_ERIC_ED476839
- Galletta, A. (2013). *Mastering the semi-structured interview and beyond: From research design to analysis and publication*. New York. NYU press.
- Gardner, J. (2012). *Assessment and learning*. London: Sage.
- Gençer, Ö. (2019) *Automated writing evaluation use in an EFL context: From paragraph writing to essay writing*. Unpublished Master's Thesis. Ankara: Ufuk University, Institute of Social Sciences.
- Gilakjani, A. P. (2017). A review of the literature on the integration of technology into the learning and teaching of English language skills. *International Journal of English Linguistics*, 7(5), 95-106.
- Glaser, B., & Strauss, A., (1967). *The discovery of grounded theory*. Hawthorne, NY: Aldine Publishing Company.
- Glasson, T. (2009). *Improving student achievement: A practical guide to assessment for learning*. Carlton: Curriculum Corporation.
- Gorgogeta, M., & Vlachos, K. (2019). A case study to compare the effects of different practices applied when assessing language learners. *Research Papers in Language Teaching and Learning*, 10(1), 241-264.
- Graham, S., & Harris, K. R. (2016). A path to better writing. *The Reading Teacher*, 69(4), 359-365.
- Graham, S., & Perin, D. (2007). What we know, what we still need to know: Teaching adolescents to write. *Scientific Studies of Reading*, 11(4), 313-335.
- Graham, S., & Sandmel, K. (2011). The process writing approach: A meta-analysis. *The Journal of Educational Research*, 104(6), 396-407.

- Graham, S., MacArthur, C. A., & Fitzgerald, J. (2013). *Best practices in writing instruction*. New York: The Guilford Press.
- Grainger, P., Purnell, K., & Zipf, R. (2008). Judging quality through substantive conversations between markers. *Assessment & Evaluation in Higher Education*, 33(2), 133-142.
- Greenfield, P. M. (1972). Oral or written language: The consequences for cognitive development in Africa, the United States and England. *Language and speech*, 15(2), 169-178.
- Grimes, D. C. (2008). *Middle school use of automated writing evaluation: A multi-site case study*. Unpublished Doctoral dissertation. Irvine: University of California, Department of Information and Computer Science.
- Grimes, D. C., & Warschauer, M. (2010). Utility in a fallible tool: A multi-site case study of automated writing evaluation. *The Journal of Technology, Learning and Assessment*, 8(6), 4-43.
- Gurel, N. (2010). *An examination of linguistic and sociocultural variables in writing a dissertation among Turkish doctoral students*. Unpublished Doctoral Dissertation. (Publication. New York: State University of New York, Department of Learning and Instruction.
- Halliday, M. A. K. & Martin, J. R. (2003). *Writing science: Literacy and discursive power*. Pittsburgh : University of Pittsburgh Press.
- Hamp-Lyons, L. (1990). Second language writing: Assessment issues. In B. Kroll, (Ed.) *Second language writing: Research insights for the classroom* (pp. 69-87). Cambridge: Cambridge University Press.
- Hamp-Lyons, L. (2002). The scope of writing assessment. *Assessing writing*, 8(1), 5-16.
- Hargreaves, E. (2005). Assessment for learning? Thinking outside the (black) box. *Cambridge Journal of Education*, 35(2), 213-224.
- Harklau, L. (2002). The role of writing in classroom second language acquisition. *Journal of second language writing*, 11(4), 329-350.
- Harlen, W. (2005). Teachers' summative practices and assessment for learning—tensions and synergies. *Curriculum Journal*, 16(2), 207-223.

- Harlen, W. (2007). *Assessment of learning*. London: Sage.
- Harmer, J. (2007). *How to teach English*. England: Pearson Education.
- Haswell, R. (2005). Researching teacher evaluation of second language writing. In T. Silva & P. K. Matsuda (Eds.), *Second Language Writing Research: Perspectives on the Process of Knowledge Construction* (pp.105–120). London: Lawrence Erlbaum Associates.
- Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of educational research*, 77(1), 81-112.
- Hayes, J. R., & Flower, L. S. (1980). Identifying the organization of writing processes. In L. Gregg & E. R. Steinberg (Eds.), *Cognitive processes in writing* (pp. 3–30). Hillsdale, NJ: Erlbaum.
- Hedge T. (2005). *Writing*. New York: Oxford University Press.
- Hegelheimer, V., & Lee, J. (2013). The role of technology in teaching and researching writing. In M. Thomas, H. Reinders, & M. Warschauer (Eds.), *Contemporary computer-assisted language learning* (pp. 287-302). London: Bloomsbury.
- Hepplestone, S., Holden, G., Irwin, B., Parkin, H. J., & Thorpe, L. (2011). Using technology to encourage student engagement with feedback: a literature review. *Research in Learning Technology*, 19(2), 117-127.
- Hibbert, K. M., Van Deven, T., & Ros, S. (2012). Fundamentals of assessment and evaluation: clarifying terminology. In Hibbert K. M., Chhem R.K. , Van Deven T., Wang S.C., (Eds). *Radiology education: the evaluation and assessment of clinical competence* (pp. 9-11). Heidelberg: Springer.
- Hockly, N. (2019). Automated writing evaluation. *ELT Journal*, 73(1), 82–88.
- Horning, A., & Robertson, J. (2006). Revision: History, theory, and practice. In A. Horning, & A. Becker (Eds), *Basic Writers and Revision* (pp. 50–62). West Lafayette, IN: Parlor Press.
- Hsu, P. S. (2016). Examining current beliefs, practices and barriers about technology integration: A case study. *TechTrends*, 60(1), 30-40.

- Hu, Z., & Wang, Z. (2019). Chinese vocational college English teachers and majors' perceptions of the product approach and process approach and attitudes toward the application of the product process approach. *Creative Education*, 10(11), 2400-2427.
- Huang, S. C. (2012). Like a bell responding to a striker: Instruction contingent on assessment. *English Teaching: Practice and Critique*, 11(4), 99-119.
- Hyland, K. (2003). Genre-based pedagogies: A social response to process. *Journal of second language writing*, 12(1), 17-29.
- Hyland, K. (2007). Genre pedagogy: Language, literacy and L2 writing instruction. *Journal of second language writing*, 16(3), 148-164.
- Hyland, K. (2009). *Academic discourse: English in a global context*. London: Continuum.
- Hyland, K. (2015). *Teaching and researching writing*. New York: Routledge.
- Hyland, K. (2019). *Second language writing*. New York: Cambridge university press.
- Hyland, K., & Hyland, F. (2006). Feedback on second language students' writing. *Language teaching*, 39(2), 83-101.
- Ioannou-Georgiou, S., & Pavlou, P. (2003). *Assessing young learners*. Oxford: Oxford University Press.
- Izumi, S. (2002). Output, input enhancement, and the noticing hypothesis: An experimental study on ESL relativization. *Studies in second language acquisition*, 24(1), 541-577.
- Jackel, B., Pearce, J., Radloff, A. & Edwards, D. (2017) *Assessment and feedback in higher education. A review of literature for the Higher Education Academy*. York: The Higher Education Academy.
- Jefferson, K., & Radhakrishnan, V. (2020). An Exploration on issues and challenges in teaching writing skills to tertiary-level learners. In S. Kumar and R. Mohanasundram, *Innovations and Technologies for Soft Skill Development and Learning* (pp. 98-107). Hershey, PA: IGI Global.
- Johnson, E., & Jenkins, J. (2009). *Formative and summative assessment. Psychology of classroom learning: An encyclopedia*. Detroit: Macmillan Reference.

- Jones, S., & Tetroe, J. (1987). Composing in a second language. In A. Matsuhashi (Ed.), *Writing in real time. Modelling production processes* (pp. 34-57). Norwood, NJ: Ablex Publishing Corporation.
- Jordan, R. R. (1997). *English for academic purposes: A guide and resource book for teachers*. Cambridge University Press.
- Joughin, G. (2009). Assessment, learning and judgement in higher education: A critical review. In G. Joughin (Ed.), *Assessment, learning and judgement in higher education* (pp. 1-15). New York, NY: Springer.
- Jubran, S. (2017). A program based on the integration of product-process approach in teaching writing in Al-Balqa Applied University. *Dirasat, Educational Sciences*, 44(4), 361-367.
- Kamaşak, R., Sahan, K., & Rose, H. (2021). Academic language-related challenges at an English-medium university. *Journal of English for Academic Purposes*, 49(1), 1-40.
- Kamimura, T. (2000). Integration of process and product orientations in EFL writing instruction. *RELC Journal*, 31(2), 1-28.
- Kapoor, M. N., & Natarajan, V. (2012). Confidence level rating in multiple choice test leading to increase in reliability and decrease in standard error of measurement. *International Journal of Information Technology and Business Management*, 1(1), 35-44.
- Kaweera, C. (2013). Writing error: A review of interlingual and intralingual interference in EFL context. *English language teaching*, 6(7), 9-18.
- Kellaghan, T., & Stufflebeam, D. L. (2003). *International handbook of educational evaluation*. Boston: Kluwer Academic Publishers.
- Kellogg, R. T. (2008). Training writing skills: A cognitive developmental perspective. *Journal of writing research*, 1(1), 1-26.
- Kellogg, R. T., & Whiteford, A. P. (2012). The development of writing expertise. In E.L. Grigorenko, E. Mambrino, & D.D. Preiss (Eds.), *Writing. A Mosaic of New Perspectives*, (pp. 109-124). New York, NY: Psychology Press.
- Khoii, R., & Doroudian, A. (2013). Automated scoring of EFL learners' written performance: a torture or a blessing. In *Proceedings of the Edulearn14: 6th International Conference*

On Education And New Learning Technologies. Italy: International Education and Training Institution.

- Kırkgöz, Y. (2019). Investigating the growth of English-medium higher education in Turkey and the Middle East Region. In D. Staub (Ed.), *Quality assurance and accreditation in foreign language education global issues, models, and best practices from the Middle East and Turkey* (pp. 9–19). Cham: Springer International Publishing.
- Kroll, B. (1990). What does time buy? ESL student performance on home versus class compositions. In B. Kroll (Ed.), *Second language writing: Research insights for the classroom* (pp. 140–154). Cambridge: Cambridge University Press
- Kroll, B. (1998). Assessing writing abilities. *Annual Review of Applied Linguistics*, 18(1), 219-240.
- Lally, C. G. (2000). First language influences in second language composition: The effect of pre-writing. *Foreign Language Annals*, 33(4), 428-432.
- Landauer, T. K., Foltz, P. W., & Laham, D. (1998). An introduction to latent semantic analysis. *Discourse processes*, 25(2-3), 259-284.
- Landauer T. K., Laham D., & Foltz P. (2003). Automated scoring and annotation of essays with Intelligent essay assessor. In Shermis M.D. and Burstein J. (Eds.), *Automated Essay Scoring: A Cross-Disciplinary Perspective* (pp. 87-112). Mahwah, NJ: Lawrence Erlbaum.
- Laurillard, D. (2013). *Rethinking university teaching: A conversational framework for the effective use of learning technologies*. New York: Routledge.
- Lebler, D. (2008). Popular music pedagogy: Peer learning in practice. *Music education research*, 10(2), 193-213.
- Lee, I. (1997). ESL learners' performance in error correction in writing: Some implications for teaching. *System*, 25(4), 465-477.
- Lee, I. (2003). L2 writing teachers' perspectives, practices and problems regarding error feedback. *Assessing Writing*, 8(3), 216-237.
- Lee, I. (2010). Writing teacher education and teacher learning: Testimonies of four EFL teachers. *Journal of Second Language Writing*, 19(3), 143-157.
- Lee, I. (2014). Feedback in writing: Issues and challenges. *Assessing Writing*, (19), 1-5.

- Lee, I. (2017). *Classroom writing assessment and feedback in L2 school contexts*. Singapore: Springer.
- Lee, I., & Coniam, D. (2013). Introducing assessment for learning for EFL writing in an assessment of learning examination-driven system in Hong Kong. *Journal of Second Language Writing*, 22(1), 34-50.
- Lee, I., Mak, P., & Burns, A. (2015). Bringing innovation to conventional feedback approaches in EFL secondary writing classrooms. *English Teaching: Practice & Critique*, 14(2), 140-163.
- Lee, I., Mak, P., & Yuan, R. E. (2019). Assessment as learning in primary writing classrooms: An exploratory study. *Studies in Educational Evaluation*, 62, 72-81.
- Lee, S. (2008). Exploring the potential of a web-based writing instruction program and AES: An empirical study using My Access. *Multimedia Assisted Language Learning*, 11(2), 103-125.
- Leki, I. (1990). Coaching from the margins: Issues in written response. In B. Kroll (Ed.), *Second language writing: Research insights for the classroom* (pp. 57-68). Cambridge: Cambridge University Press.
- Leki, I. (2001). Material, educational, and ideological challenges of teaching EFL writing at the turn of the century. *International Journal of English Studies*, 1(2), 197-209.
- Leki, I. (2003). A challenge to second language writing professionals: Is writing overrated? In B. Kroll (Ed.), *Exploring the Dynamics of Second Language Writing* (pp. 315-331). New York: Cambridge University Press.
- Leki, I., Cumming, A., & Silva, T. (2010). *A synthesis of research on second language writing in English*. New York: Routledge.
- Lerner, N. D. (1996). *Teaching and learning in a university writing center*. Unpublished dissertation, Massachusetts: Boston University, Department of Humanities and Social Sciences.
- Li, J., Link, S., & Hegelheimer, V. (2015). Rethinking the role of automated writing evaluation (AWE) feedback in ESL writing instruction. *Journal of Second Language Writing*, 27, 1-18.

- Li, Z., Link, S., Ma, H., Yang, H., & Hegelheimer, V. (2014). The role of automated writing evaluation holistic scores in the ESL classroom. *System*, 44, 66-78.
- Lightbown, P., & Spada, N. (2006). *How languages are learned*. New York: Oxford University Press.
- Limpo, T., Nunes, A., & Coelho, A. (2020). Introduction to the special issue on technology-based writing instruction: A collection of effective tools. *Journal of Writing Research*, 12(1), 1–7.
- Link, S., Dursun, A., Karakaya, K., & Hegelheimer, V. (2014). Towards better ESL practices for implementing automated writing evaluation. *Calico Journal*, 31(3), 323–344.
- Linse, C. T. (2005). *Practical English language teaching: Young learners*. New York: McGraw-Hill Companies.
- Lombana, C. H. (2002). Some issues for the teaching of writing. *Profile: Issues in Teachers' Professional Development*, 3(1), 44-51.
- Long, M. H. (1984). Process and product in ESL program evaluation. *TESOL Quarterly*, 18(3), 409-425.
- Long, R. (2013). A review of ETS's Criterion online writing program for student compositions. *The Language Teacher*, 37(3), 11-18.
- Macken-Horarik, M. (2002). "Something to shoot for": a Systemic Functional approach to teaching genre in secondary school science. In A. Johns (ed.), *Genre in the classroom: multiple perspectives* (pp. 21–46). Mahwah, NJ: Erlbaum.
- Manchon, R. (2011). *Learning-to-write and writing-to-learn in an additional language*. Philadelphia PA: John Benjamins Publishing.
- Manchon, R., & Larios, J. R. (2008). Writing-to-learn in instructed language learning contexts. In E. Alcon Soler & M. P. Safont Jorda (Eds.), *Intercultural language use and language learning* (pp. 101–121). Dordrecht: Springer.
- Marshall, B., & Drummond, M. (2006). How teachers engage with assessment for learning: Lessons from the classroom. *Research papers in education*, 21(02), 133-149.
- Martin, H. J. (1994). *The history and power of writing*. Chicago: University of Chicago Press.

- Mason, B. J., & Bruning, R. (2001). Providing feedback in computer-based instruction: What the research tells us. Retrieved from <http://dwb.unl.edu/Edit/MB/MasonBruning.html>
- Matsuda, P. K. (2003). Process and post-process: A discursive history. *Journal of second language writing*, 12(1), 65-83.
- Matsumoto, K., & Akahori, K. (2008). Evaluation of the use of automated writing assessment software. In *E-Learn: World Conference on E-Learning in Corporate, Government, Healthcare, and Higher Education*. Japan: Tokai University.
- McNamara, T. (2000). *Language testing*. Oxford: Oxford University Press.
- Mehr, H. S. (2017). The impact of product and process approach on Iranian EFL learners' writing ability and their attitudes toward writing skill. *International Journal of English Linguistics*, 7(2), 158-166.
- Merriam, S. B., & Tisdell, E. J. (2015). *Qualitative research: A guide to design and implementation*. San Francisco: John Wiley & Sons.
- Miles, M. B., & Huberman, A. M. (1994). *Qualitative data analysis: An expanded sourcebook*. London: SAGE.
- Miri, T. (2014). Toward finding an approach for improving rhetorical organization of EFL learners' argumentative writing. *International Journal of Applied Linguistics and English Literature*, 3(3), 164-170.
- Mohsen, M. A., & Abdulaziz, A. (2019). The effectiveness of using a hybrid mode of automated writing evaluation system on EFL students' writing. *Teaching English with Technology*, 19(1), 118-131.
- Murray, L., & Hourigan, T. (2008). Blogs for specific purposes: Expressivist or socio-cognitivist approach?. *ReCALL*, 20(1), 82-97.
- Narciss, S. (2008). Feedback strategies for interactive learning tasks. *Handbook of research on educational communications and technology*, 3(1), 125-144.
- Neff-Lippman, J. (2012). Assessing writing. In I. L. Clark (Ed.), *Concepts in composition: Theory and practice in the teaching of writing* (pp. 145-167). New York: Routledge.
- Nordin, S. M. (2017). The best of two approaches: Process/genre-based approach to teaching writing. *The English Teacher*, 11(1), 75-85.

- Nordrum, L., Evans, K., & Gustafsson, M. (2013). Comparing student learning experiences of in-text commentary and rubric-articulated feedback: strategies for formative assessment. *Assessment & Evaluation in Higher Education*, 38(8), 919-940.
- Nunan, D. (1989). *Designing tasks for the communicative classroom*. Cambridge: Cambridge University Press.
- Nunan, D. (1999). *Second language teaching and learning*. USA: Heinle & Heinle Publishers.
- O'Brien-Moran, M., & Soiferman, L.K. (2010). Do first-year university students know what to expect from their first-year writing intensive course? *Paper presented at the 8th Annual Hawaii International Conference on Education*. Canada: University of Manitoba.
- Olson, D. R. (1996). *The world on paper: The conceptual and cognitive implications of writing and reading*. Cambridge: Cambridge University Press.
- Page, E. B. (1966). The imminence of grading essays by computer. *Phi Delta Kappan*, 48(1), 238-243.
- Page, E. B. (2003). Project Essay Grade: PEG. In M. D. Shermis & J. Burstein (Eds.), *Automated essay scoring: A cross-disciplinary perspective* (pp. 43–54). New Jersey: Lawrence Erlbaum Associates.
- Page, E. B., & Petersen, N. S. (1995). The computer moves into essay grading: Updating the ancient test. *Phi Delta Kappan*, 76(7), 561-565.
- Palermo, C., & Thomson, M. M. (2018). Teacher implementation of self-regulated strategy development with an automated writing evaluation system: Effects on the argumentative writing performance of middle school students. *Contemporary Educational Psychology*, 54(1), 255-270.
- Palermo, C., & Wilson, J. (2020). Implementing automated writing evaluation in different instructional contexts: A mixed-methods study. *Journal of Writing Research*, 12(1), 63-108.
- Parr, J. M., & Timperley, H. S. (2010). Feedback to writing, assessment for teaching and learning and student progress. *Assessing Writing*, 15(2), 68-85.
- Parra, G. (2019). Automated writing evaluation tools in the improvement of the writing skill. *International Journal of Instruction*, 12(2), 209-226.

- Patthey-Chavez, G. G., Matsumura, L. C., & Valdes, R. (2004). Investigating the process approach to writing instruction in urban middle schools. *Journal of Adolescent & Adult Literacy*, 47(6), 462-476.
- Peregoy, S., & Boyle, O. F. (2001). *Reading, writing, & learnign in ESL*. New York, NY: Longman.
- Persky, H. R., Daane, M. C, & Jin, Y. (2003). *The nation's report card: Writing 2002*. Washington, DC: U. S. Department of Education, Institute of Education Sciences, National Center for Education Statistics.
- Phuket, P. R. N., & Othman, N. B. (2015). Understanding EFL students' errors in writing. *Journal of Education and Practice*, 6(32), 99-106.
- Pica, T. (1994). Questions from the language classroom: Research perspectives. *Tesol Quarterly*, 28(1), 49-79.
- Pincas, A. (1982). *Teaching English writing*. London: Macmillan Press.
- Pintrich, P. R. (1999). The role of motivation in promoting and sustaining self-regulated learning. *International journal of educational research*, 31(6), 459-470.
- Powers, D. E., Burstein, J. C., Chodorow, M. S., Fowles, M. E., & Kukich, K. (2002). Comparing the validity of automated and human scoring of essays. *Journal of Educational Computing Research*, 26(4), 407-425.
- Prensky, M. (2001). Digital natives, Digital immigrants. *On the Horizon*, 9 (5), 1-6.
- Pritchard, R. J., & Honeycutt, R. L. (2007). Best practices in implementing a process approach to writing. In S. Graham, C. A. MacArthur, & J. Fitzgerald (Eds.), *Best practices in writing instruction* (pp. 28– 49). New York: The Guilford Press.
- Rahman, M. M. (2011). Genre-based writing instruction: Implications in ESP classroom. *English for Specific Purposes World*, 33(11), 1-9.
- Raimes, A. (1985). What unskilled ESL students do as they write: A classroom study of composing. *TESOL quarterly*, 19(2), 229-258.
- Raimes, A. (1991). Out of the woods: Emerging traditions in the teaching of writing. *TESOL quarterly*, 25(3), 407-430.

- Raissah, N., & Aziz, Z. A. (2020). An investigation of interlingual and intralingual interference found in English as a foreign language (EFL) students' composition of recount text. *English Education Journal*, 11(2), 251-275.
- Ramineni, C. (2013). Validating automated essay scoring for online writing placement. *Assessing Writing*, 18(1), 40-61.
- Ramineni, C., Trapani, C. S., Williamson, D. M., Davey, T., & Bridgeman, B. (2012). *Evaluation of the e-rater® scoring engine for the TOEFL® independent and integrated prompts: ETS Research Report*. Princeton, NJ: Educational Testing Service.
- Ranalli, J., Link, S., & Chukharev-Hudilainen, E. (2017). Automated writing evaluation for formative assessment of second language writing: investigating the accuracy and usefulness of feedback as part of argument-based validation. *Educational Psychology*, 37(1), 8-25.
- Reid, J. (2001). Writing. In: R. Carter & D. Nunan. (Eds.). *The Cambridge guide to teaching English to speakers of other languages* (pp. 28-33). Cambridge: Cambridge University Press.
- Richards, J. (1974) A non-contrastive approach to error analysis. In Richards, J. (ed.) *Error Analysis: Perspectives on Second Language Acquisition*. London: Longman.
- Rivers, W.M. (1968). *Teaching foreign language skills*. Chicago and London: The University of Chicago Press.
- Roscoe, R. D., Allen, L. K., Johnson, A. C., & McNamara, D. S. (2018). Automated writing instruction and feedback: Instructional mode, attitudes, and revising. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*. Los Angeles, CA: SAGE Publications.
- Roscoe, R. D., Wilson, J., Johnson, A. C., & Mayra, C. R. (2017). Presentation, expectations, and experience: Sources of student perceptions of automated writing evaluation. *Computers in Human Behavior*, 70 (1), 207-221.
- Ruegg, R. (2010). Interlanguage development: The effect of unfocused feedback on L2 writing. *Intercultural communication studies*, 12(1), 247-254.

- Rusinovci, X. (2015). Teaching writing through process-genre based approach. *US-China Education Review*, 5(10), 699-705.
- Ryan, L. (2012). Students' attitudes towards corrective feedback in the second language classroom. Unpublished Bachelor Dissertation. Dublin: University of Dublin, Department of Linguistics and Language.
- Sadler, D. R. (2009). Transforming holistic assessment and grading into a vehicle for complex learning. In G. Joughin (Ed.), *Assessment, Learning and Judgement in Higher Education* (pp. 1-19). Netherlands: Springer.
- Sadrolmomaleki, V. (2009). Electronic feedback and teaching writing. *International Journal of Arts and Sciences*, 3(5), 36-48.
- Sağlamel, H., & Kayaoğlu, M. N. (2015). English major students' perceptions of academic writing: A struggle between writing to learn and learning to write. *Journal of History Culture and Art Research*, 4(3), 37-52.
- Sallies, T. G. (1998). *Towards communicative measurement of writing: Where are we now?* Rio de Jenerio: Pontificia Universidade.
- Schaefer, E. (2008). Rater bias patterns in an EFL writing assessment. *Language Testing*, 25(4), 465-493.
- Scharber, C., Dexter, S., & Riedel, E. (2008). Students' experiences with an automated essay scorer. *Journal of Technology, Learning, and Assessment*, 7(1), 4-44.
- Schultz, J. M. (1991). Mapping and cognitive development in the teaching of foreign language writing. *The French Review*, 64(6), 978-988.
- Schultz, J. M. (2011). Foreign language writing in the era of globalization. In T. Cimasko & M. Reichelt (Eds.), *Foreign language writing instruction: Principles and practices* (pp. 65-82). Anderson, SC: Parlor Press.
- Schultz, M. T. (2013). The IntelliMetrix automated essay scoring engine: A review and an application to Chinese essay scoring. In M. D. Shermis, & J. Burstein (Eds.), *Handbook of automated essay scoring: Current applications and future directions* (pp. 89-98). New York: Routledge.
- Scriven, M. (1991). *Evaluation thesaurus*. Los Angeles, CA: Sage.
- Scriven, M. (1996). Types of evaluation and types of evaluator. *Evaluation practice*, 17(2), 151-161.

- Senel, E. (2018). The integration of creative writing into academic writing skills in EFL classes. *Online Submission*, 6(2), 115-120.
- Shermis, M. D., & Burstein, J. (2003). *Automated essay scoring: A cross-disciplinary perspective*. New York: Routledge.
- Shermis, M. D., & Burstein, J. (2013). *Handbook of automated essay evaluation: Current applications and new directions*. New York: Routledge.
- Shermis, M. D., Burstein, J., & Bursky S. A. (2013). Introduction to automated essay evaluation. In M. Shermis & J. Burstein (Eds.), *Handbook of automated essay evaluation: Current applications and new directions* (pp. 1-15). New York: Routledge.
- Shermis, M. D., Burstein, J., Higgins, D., & Zechner, K. (2010). Automated essay scoring: Writing assessment and instruction. *International encyclopedia of education*, 4(1), 20-26.
- Shermis, M. D., Garvan, C. W., & Diao, Y. (2008). The impact of automated essay scoring on writing outcomes. *Paper presented at the annual meeting of the National Council on Measurement in Education*, New York, NY: University of Florida.
- Shermis, M. D., Koch, C. M., Page, E. B., Keith, T. Z., & Harrington, S. (2002). Trait ratings for automated essay grading. *Educational and Psychological Measuremnt*, 62(1), 5-18.
- Shute, V. J. (2008). Focus on formative feedback. *Review of Educational Research*, 78(1), 153-189.
- Silva, T. (1990). Second language composition instruction: developments, issues, and directions. In B. Kroll (ed.), *Second language writing: research insights for the classroom* (pp. 11–23). Cambridge: Cambridge University Press.
- Silva, T. (1993). Toward an understanding of the distinct nature of L2 writing: The ESL research and its implications. *TESOL quarterly*, 27(4), 657-677.
- Smyth, T. (2004). *The principles of writing in psychology: Macmillan study skills*. London: Palgrave Macmillan.
- Sprague, A. (2017). *Analyzing the feedback preferences and learning styles of second-language students in esol writing courses at Bowling Green State University*. Unpublished Doctoral Dissertation. Ohio: Bowling Green State University, Department of English.
- Spratt, M., Humphreys, G., & Chan, V. (2002). Autonomy and motivation: Which comes first?. *Language teaching research*, 6(3), 245-266.

- Srichanyachon, N. (2014). EFL learners' perceptions of using lms. *Turkish Online Journal of Educational Technology-TOJET*, 13(4), 30-35.
- Stevenson, M. (2016). A critical interpretative synthesis: The integration of automated writing evaluation into classroom writing instruction. *Computers and Composition*, 42(1), 1-16.
- Stiggins, R.J., Arter, J.A., Chappuis, J., & Chappuis S. (2004). *Classroom assessment for learning: Doing it right—using it well*. Princeton, NJ: Educational Testing Service.
- Stone, G. (2017). Implementation of critical literacy for English writing classes in the Thai context. *The New English Teacher*, 11(2), 65-65.
- Susser, B. (1994). Process approaches in ESL/EFL writing instruction. *Journal of Second language writing*, 3(1), 31-47.
- Swain, M. (1985). Communicative competence: Some roles of comprehensible input and comprehensible output in its development. *Input in second language acquisition*, 15(1), 165-179.
- Swain, M., & Lapkin, S. (1995). Problems in output and the cognitive processes they generate: A step towards second language learning. *Applied linguistics*, 16(3), 371-391.
- Swales, J. (1990). *Genre analysis: English in academic and research settings*. Cambridge: Cambridge University Press.
- Tangpermpoon, T. (2008). Integrated approach to improve students' writing skills for English major students, *ABAC Journal*, 28(2), 1-9.
- Taras, M. (2005). Assessment—summative and formative—some theoretical reflections. *British journal of educational studies*, 53(4), 466-478.
- Tawney, J., & Gast, D. (1984). *Single-subject research in special education*. New York: Merrill.
- Teng, F. (2016). Immediate and delayed effects of embedded metacognitive instruction on Chinese EFL students' English writing and regulation of cognition. *Thinking Skills and Creativity*, 22(1), 289-302.

- Tom, A. A., Morni, A., Metom, L., & Joe, S. (2013). Students' perception and preferences of written feedback in academic writing. *Mediterranean Journal of Social Sciences*, 4(11), 72-72.
- Tribble, C. (1996). *Writing*. Oxford: Oxford University Press.
- Tuffs, R. (1993). A genre approach to writing in the second language classroom: the use of direct mail letters. *Revue belge de philologie et d'histoire*, 71(3), 699-718.
- Ur, P. (2012). *A course in English language teaching*. Cambridge: Cambridge University Press.
- Üstünbaş, Ü., & Çimen, S. (2016). EFL learners' preferences for feedback types for their written products. *The Online Journal of New Horizons in Education*, 6(4), 68-74.
- Van Beuningen, C. (2010). Corrective feedback in L2 writing: Theoretical perspectives, empirical insights, and future directions. *International Journal of English Studies*, 10(2), 1-27.
- Vandergrift, L. (2005). Relationships among motivation orientations, metacognitive awareness and proficiency in L2 listening. *Applied linguistics*, 26(1), 70-89.
- Vantage Learning. (2006). *Research summary: Intellimetric scoring accuracy across genres and grade levels*. Retrieved from http://www.vantagelearning.com/docs/intellimetric/IM_ReseachSummary_InteliMetric_Accuracy_Across_Genre_and_Grade_Levels.pdf
- Vygotsky, L. (1978). Interaction between learning and development. *Readings on the development of children*, 23(3), 34-41.
- Vyncke, M. (2012). *The concept and practice of critical thinking in academic writing: an investigation of international students' perceptions and writing experiences*. Unpublished MA thesis. London: King's College, Graduate Faculty of Social Science.
- Wang, F., Chen, X., & Fang, W. (2011). Integrating cell phones into a Chinese high school EFL classroom: Students' attitudes, technological readiness, and perceived learning. *Journal of Educational Technology Development and Exchange*, 4(1), 91-102.

- Wang, Y. J., Shang, H. F., & Briody, P. (2013). Exploring the impact of using automated writing evaluation in English as a foreign language university students' writing. *Computer Assisted Language Learning*, 26(3), 234-257.
- Wang, Z. Q. (2014). The application of process writing in Chinese EFL classrooms in higher education. *International Journal of English Linguistics*, 4(3), 88.
- Ware, P. (2011). Computer-generated feedback on student writing. *Tesol Quarterly*, 45(4), 769-774.
- Warschauer, M. (2010). New tools for teaching writing. *Language learning & technology*, 14(1), 3-8.
- Warschauer, M., & Grimes, D. (2008). Automated writing assessment in the classroom. *Pedagogies: An International Journal*, 3(1), 22-36.
- Warschauer, M., & Ware, P. (2006). Automated writing evaluation: Defining the classroom research agenda. *Language teaching research*, 10(2), 157-180.
- Webster, L., & Ammon, P. (1994). Linking written language to cognitive development: Reading, writing, and concrete operations. *Research in the Teaching of English*, 28(1), 89-109.
- Weigle, S. C. (2002). *Assessing writing*. Cambridge: Cambridge University Press.
- Weigle, S. C. (2013). English as a second language writing and automated essay evaluation. In M.D. Shermis, J.C. Burstein (Eds.), *Handbook of Automated Essay Evaluation: Current Applications and New Directions*, (pp. 36–54). New York: Routledge.
- Weigle, S. C. (2016). Second language writing assessment. In R. Manchon & P. K. Matsuda (Eds.), *Handbook second and foreign language writing* (pp. 473-493). Berlin: De Gruyter Mouton.
- White, E. M. (1994). Issues and problems in writing assessment. *Assessing Writing*, 1(1), 11-27.
- White, E. M. (2009). *Putting assessment for learning into practice in a higher education EFL context*. Sydney: Universal-Publishers.

- White, J., & Ihasz, P. L. (2015). The use of formal english language and ict input sources among japanese university students. *Paper presented in International Conference on Culture and Computing (Culture Computing)*. Japan: IEEE.
- White, L. (1987). Against comprehensible input: the input hypothesis and the development of second-language competence¹. *Applied linguistics*, 8(2), 95-110.
- White, R. (1988). Academic writing: process and product”. In P. C. Robinson (Ed.), *Academic writing: process and product. ELT documents* (pp. 4-16). London: Modern English Publications and The British Council.
- White, R., & Arndt, V. (1991). *Process writing*. London: Longman.
- William, D. (2011). *Embedded formative assessment*. Bloomington, IN: Solution Tree Press.
- William, G., & Kaplan, R. B. (1996). *Theory and practice of writing: An applied linguistic perspective*. London: Longman.
- Wilson, J., & Czik, A. (2016). Automated essay evaluation software in English Language Arts classrooms: Effects on teacher feedback, student motivation, and writing quality. *Computers & Education*, 100 (1), 94-109.
- Wilson, J., & Roscoe, R. D. (2020). Automated writing evaluation and feedback: Multiple metrics of efficacy. *Journal of Educational Computing Research*, 58(1), 87-125.
- Wilson, J., Roscoe, R. D., & Ahmed, Y. (2017). Automated formative writing assessment using a levels of language framework. *Assessing Writing*, 34(1), 16-36.
- Winsor, D. A. (2012). *Writing power: Communication in an engineering center*. New York: State University of New York Press.
- Xi, X. (2010). Automated scoring and feedback systems: Where are we and where are we heading? *Language Testing*, 27, 291–300.
- Yağiz, O., Yiğiter, K. & Genç, G. (2009). Academic writing in english: Approaches, processes and challenges in higher education. *Paper presented at the 1st International Symposium on Sustainable Development*. Sarajevo: International Burch University.
- Yalçın, I. (2010). *Turkish speaking first year and third year ELT students’ syntactic errors in their argumentative essays*. Unpublished doctoral dissertation. Eskişehir: Anadolu University, Institute of Educational Sciences.

- Ying, Z. (2018). Exploring construction of college English writing course from the perspective of output-driven hypothesis. *English Language Teaching*, 11(2), 188-192.
- Yuhong , J. (2005). The role of online peer review in learners writing ability development. *Foreign Language Teaching and Research*, 3 (3) : 226—230.
- Yurdugül, H., & Çetin, N. M. (2015). Investigation of the relationship between learning process and learning outcomes in e-learning environments. *Eurasian Journal of Educational Research*, 15(59), 57-74.
- Zamel, V. (1983). The composing processes of advanced ESL students: Six case studies. *TESOL quarterly*, 17(2), 165-188.
- Zeng, S. (2020). The potential of online technology for language learning. *English Language Teaching*, 13(10), 23-37.
- Zhang, L. J., & Qin, T. L. (2018). Validating a questionnaire on EFL writers' metacognitive awareness of writing strategies in multimedia environments. In A. Haukas, C. C. Bjorke, & M. Dypedahl (Eds.), *Metacognition in language learning and teaching* (pp. 157-179). London, : Routledge.
- Zhang, Z. (2020). Engaging with automated writing evaluation (AWE) feedback on L2 writing: Student perceptions and revisions. Elsevier Inc: *Assessing Writing*, 43, 1-14.
- Zhang, Z. V., & Hyland, K. (2018). Student engagement with teacher and automated feedback on L2 writing. *Assessing Writing*, 36(1), 90-102.
- Zhang, Z., & Zhang, Y. (2018). Automated writing evaluation system: Tapping its potential for learner engagement. *IEEE Engineering Management Review*, 46(3), 29-33.
- Zheng, S., & Dai, W. (2012). Studies and suggestions on prewriting activities. *Higher Education Studies*, 2(1), 79-87.
- Zheng, Y., & Yu, S. (2018). Student engagement with teacher written corrective feedback in EFL writing: A case study of Chinese lower-proficiency students. *Assessing Writing*, 37, 13-24.
- Zhou, D. (2015). An empirical study on the application of process approach in non-English majors' writing. *English Language Teaching*, 8(3), 89-96.

- Zimmerman, B.J., Bonner, S., & Kovach, R. (1996). *Developing self-regulated learners: Beyond achievement to self-efficacy*. Washington, DC: American Psychological Association.
- Zupanc, K., & Bosnic, Z. (2015). Advances in the field of automated essay evaluation. *Informatica*, 39 (4), 383–395.
- Zupanc, K., & Bosnic, Z. (2017). Automated essay evaluation with semantic analysis. *Knowledge-Based Systems*, 120 (1), 118-132.

Appendix A – Students’ Consent Form

Öğrenciler için Gönüllü Katılım Formu

Bu çalışma,” **Otomatik Yazma Becerisi Değerlendirme Sisteminin Öğrencilerin Yazma Performansı Üzerindeki Etkisi ve Öğretmenler ile Öğrencilerin Bu Sistem Hakkındaki Görüşleri**” başlıklı bir araştırma çalışması olup **Otomatik yazma becerisi değerlendirme sistemlerinin yabancı dil olarak İngilizce öğretimi bağlamında kullanımına uygunluğunu ortaya çıkarma** amacını taşımaktadır. Çalışma, **Hilal YILDIZ** tarafından yürütülmekte ve sonuçları ile **öğretmenlerin ve öğrencilerin otomatik yazma becerisi değerlendirme sistemlerinin kullanımı hakkında görüşlerine ve bu sistemlerden sağlanan değerlendirmenin etkinliğine** ışık tutulacaktır.

- Bu çalışmaya katılımınız gönüllülük esasına dayanmaktadır.
- Çalışmanın amacı doğrultusunda, otomatik yazma becerisi değerlendirme sistemleri kullanımı hakkında kısa bir eğitim alacaksınız ve haftalık yansıtma raporu ile görüşmeler yapılarak sizden veriler toplanacaktır. Sekiz hafta boyunca bu sistemleri kullanarak yazacağınız denemeler incelenerek elde edilen veri araştırmaya dahil edilecektir.
- İsminizi yazmak ya da kimliğinizi açığa çıkaracak bir bilgi vermek zorunda değilsiniz/araştırmada katılımcıların isimleri gizli tutulacaktır.
- 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.
- İstemeniz halinde sizden toplanan verileri inceleme hakkınız bulunmaktadır.
- Sizden toplanan veriler araştırma sonunca kadar korunacak ve araştırma bitiminde arşivlenecek veya imha edilecektir.
- Veri toplama sürecinde/süreçlerinde size rahatsızlık verebilecek herhangi bir soru/talep olmayacaktır. Yine de katılımınız sırasında herhangi bir sebepten rahatsızlık hissederseniz çalışmadan istediğiniz zamanda ayrılabilirsiniz. Çalışmadan ayrılmanız durumunda sizden toplanan veriler çalışmadan çıkarılacak ve imha edilecektir.

Gönüllü katılım formunu okumak ve değerlendirmek üzere ayırdığınız zaman için teşekkür ederim. Çalışma hakkındaki sorularınızı Antalya Bilim Üniversitesi Yabancı Diller Yüksekokulu’ndan Hilal Yıldız’a yöneltebilirsiniz.

Araştırmacı Adı: Hilal Yıldız

Adres :

E-posta adresi : hilalyildiz@antalya.edu.tr

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

(Lütfen bu formu doldurup imzaladıktan sonra veri toplayan kişiye veriniz.)

Katılımcı Ad ve Soyadı:

İmza:

Tarih:

Appendix B – Instructors’ Consent Form

Öğretmenler için Gönüllü Katılım Formu

Bu çalışma, “**Otomatik Yazma Becerisi Değerlendirme Sisteminin Öğrencilerin Yazma Performansı Üzerindeki Etkisi ve Öğretmenler ile Öğrencilerin Bu Sistem Hakkındaki Görüşleri**” başlıklı bir araştırma olup **Otomatik yazma becerisi değerlendirme sistemlerinin yabancı dil olarak İngilizce öğretimi bağlamında kullanımına uygunluğunu ortaya çıkarma** amacındadır. Çalışma, **Hilal YILDIZ** tarafından yürütülmekte ve sonuçları ile **öğretmenlerin ve öğrencilerin otomatik yazma becerisi değerlendirme sistemlerinin kullanımı hakkında görüşlerine ve bu sistemlerden sağlanan değerlendirmenin etkinliğine** ışık tutulacaktır.

- Bu çalışmaya katılımınız gönüllülük esasına dayanmaktadır.
- Çalışmanın amacı doğrultusunda, otomatik yazma becerisi değerlendirme sistemleri kullanımı hakkında kısa bir eğitim alacaksınız ve bu sistemlerin sağladığı değerlendirme ile öğretmenlerin sağladığı değerlendirme arasındaki benzerlik ve farklılıkları görebilmek adına değerlendirmeniz istenecektir. Sonrasında kendi değerlendirmeniz ve sistem değerlendirmesi hakkında görüşlerinizi öğrenebilmek için yansıtma sorularına sözlü cevap vermeniz istenecek ve görüşme yapılacaktır. Kaydedilen görüşmeden ve yaptığınız değerlendirmeden elde edilen veriler araştırmaya dahil edilecektir.
- İsminizi yazmak ya da kimliğinizi açığa çıkaracak bir bilgi vermek zorunda değilsiniz/araştırmada katılımcıların isimleri gizli tutulacaktır.
- 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.
- İstemeniz halinde sizden toplanan verileri inceleme hakkınız bulunmaktadır.
- Sizden toplanan veriler araştırma sonunca kadar korunacak ve araştırma bitiminde arşivlenecek veya imha edilecektir.
- Veri toplama sürecinde/süreçlerinde size rahatsızlık verebilecek herhangi bir soru/talep olmayacaktır. Yine de katılımınız sırasında herhangi bir sebepten rahatsızlık hissederseniz çalışmadan istediğiniz zamanda ayrılabilirsiniz. Çalışmadan ayrılmanız durumunda sizden toplanan veriler çalışmadan çıkarılacak ve imha edilecektir.

Gönüllü katılım formunu okumak ve değerlendirmek üzere ayırdığınız zaman için teşekkür ederim. Çalışma hakkındaki sorularınızı Antalya Bilim Üniversitesi Yabancı Diller Yüksekokulu’ndan Hilal Yıldız’a yöneltebilirsiniz.

Araştırmacı Adı: Hilal Yıldız

Adres :

E-mail adresi : hilalyildiz@antalya.edu.tr

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

(Lütfen bu formu doldurup imzaladıktan sonra veri toplayan kişiye veriniz.)

Katılımcı Ad ve Soyadı:

İmza:

Tarih:

Appendix C- Holistic Writing Criteria of ABU SOFL

	- Content fulfills the requirements of the task - Ideas are totally relevant, well-developed, supported with relevant examples - Effective organisation of ideas / effective use of cohesive devices
18-20	- Minor grammatical errors - Wide range of appropriate sentence structures - Minor capitalization / punctuation errors - Very effective use of target vocabulary – Wide range - Minor spelling mistakes
15-17	- Content fulfills the requirements of the task - Ideas are relevant and usually well-developed or supported - Good organisation of ideas / good use of cohesive devices - Few grammatical errors - Good range - Very few capitalization / punctuation errors - Effective use of target vocabulary - Good range - Very few spelling mistakes
12-14	- Content shows some attempt to fulfill the requirements of the task - Most of the ideas are relevant and there is some support - Ideas are adequately organised - Adequate use of cohesive devices - Some grammatical errors which do not obscure meaning - Adequate range - Few capitalization / punctuation errors - Adequate use of target vocabulary - Adequate range - Few spelling mistakes
9-11	- Content shows limited attempt to fulfill the requirements of the task - Some ideas are relevant but there is inadequate support and development - Some ideas are not organised effectively -Inadequate use of cohesive devices - Some grammatical errors which sometimes obscure meaning –Limited range - Some capitalization / punctuation errors - Inadequate use of target vocabulary – Limited range - Some spelling mistakes

6-8	- Content shows very limited attempt to fulfill the requirements of the task - Ideas are somewhat relevant but poorly developed or supported - Ideas are not organised effectively - Poor use of cohesive devices - Frequent grammatical errors which sometimes obscure meaning – Very limited range - Frequent capitalization / punctuation errors - Lexical mistakes often obscure meaning– Very limited range - Frequent spelling mistakes
3-5	- Content shows almost no attempt to fulfill the requirements of the task - Ideas are often irrelevant and not developed or supported - Almost no organisation of ideas – Very poor use of cohesive devices - Frequent grammatical errors which often obscure meaning – Almost no range - Very frequent capitalization / punctuation errors - Lexical mistakes obscure meaning – Almost no range - Very frequent spelling mistakes
0-2	- No response or totally irrelevant

Appendix D- Criterion Six-Point Holistic Rubric

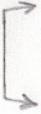
ETS Criterion Score Guide

6 | 5 | 4 | 3 | 2 | 1

Score = 6	Excellent - Develops ideas well and uses many specific, relevant details throughout the essay. - Is well organized with clear transitions; maintains focus. - Sustains varied sentence structure. - Exhibits many specific word choices. - Contains little or no errors in grammar and conventions; errors do not interfere with understanding. Persuasive mode - Clearly states the position and effectively persuades the reader of validity of argument.
Score = 5	Skillful - Develops ideas with some specific, relevant details. - Is clearly organized; information is presented in an orderly way, but essay may lack transitions. - Exhibits some variety in sentence structure. - Displays some specific word choices. - May contain some errors in grammar and conventions; errors do not interfere with understanding. Persuasive mode - Clearly states the position and persuades the reader.
Score = 4	Sufficient - Provides clear ideas, but sparsely developed; may have few details. - Provides a clear sequence of information; provides pieces of information that are generally related to each other. - Generally has simple sentences; may exhibit uneven control over sentence structure. - Consists mainly of simple word choices, but may contain some specific word choices. - Contains errors in grammar and conventions that generally do not interfere with understanding. Persuasive mode - States a position and adequately attempts to persuade the reader.
Score = 3	Uneven - Provides limited or incomplete information; may be list-like or have the quality of an outline. - Is disorganized or provides a disjointed sequence of information. - Exhibits uneven control over sentence structure. - May have some inaccurate word choices. - Contains errors in grammar and conventions that sometimes interfere with understanding. Persuasive mode - While a position is stated, either it is unclear or undeveloped.
Score = 2	Insufficient - Provides little information and makes little attempt at development. - Is very disorganized or too brief to detect organization. - Exhibits little control over sentence structure. - Contains inaccurate word choices in much of the essay. - Is characterized by misspellings, missing words, and incorrect word order; errors in grammar and conventions are severe enough to make understanding very difficult in much of the essay. Persuasive mode - Either a position is not clearly given or little attempt is made at persuasion.
Score = 1	Unsatisfactory - Attempts a response, but may only paraphrase the prompt or be extremely brief. - Exhibits no control over organization. - Exhibits no control over sentence structure. - Contains inaccurate word choices throughout most of the essay. - Is characterized by misspellings, missing words, and incorrect word order; errors in grammar and conventions severely impede understanding throughout the essay. Persuasive mode - Little effort is made to persuade, either because there is no position taken or no support is given.

Appendix E- Error Codes

CORRECTION SYMBOLS			
THE SYMBOL	WHAT THE SYMBOL MEANS	EXAMPLE	
		FALSE	CORRECT
Λ	missing word	It <u>is</u> interesting book. Λ	It is an interesting book.
S / P	Singular / plural mistake	I really hate insect. s/p	I really hate insects.
T	Tense mistake	I go to school yesterday morning. T	I went to school yesterday morning.
?	I do not understand this	I school very go good fine. ? ?	
S / V	Subject and verb do not agree.	My brother and I <u>am</u> interested in technology. s/v	My brother and I are interested in technology.
C / UC	Countable / uncountable mistake	There are many kinds of <u>coffees</u> . c/uc	There are many kinds of coffee.
IR	Irrelevant information	Weather in Scotland at this time of the year is chilly. <u>We always wear shorts when we work outdoors.</u> IR	Weather in Scotland at this time of the year is chilly.
WO	Word order mistake	Usually you <u>do go</u> to the cinema? w/o	Do you usually go to the cinema?
WW	Wrong word	Jane's 1m 80 cm, which is very <u>long</u> for a woman. ww	Jane's 1m 80cm, which is very tall for a woman.
WF	Wrong word form	Could you find the solve of this problem? w/f	Could you find the solution of this problem?
Sp.	Spelling mistake	I don't like my neighbourhood because it's so <u>noisy</u> . s.p.	I don't like my neighbourhood because it's so noisy.

C	Capitalization mistake	I don't have classes on tuesdays.	I don't have classes on Tuesdays.
X	This is one word	house X wife	housewife
/	Separate these words	post / office	post office
P	Punctuation	Do you want some coffee. <u>P</u>	Do you want some coffee?
SF	Sentence fragment	Because I don't like spinach. SF	
RO	Run-on sentence	I did shopping, ate lunch. <u>RO</u>	I did shopping and ate lunch.
X	Remove this word	English language X is an important language.	English is an important language.
PP	Wrong preposition	Our parents are interested with our problems. PP	Our parents are interested in our problems.
Tr.	Transition signal	Pat was sick. He stayed at home. <u>Tr.</u>	Pat was sick, so he stayed at home.
	These words must change place	I easily can draw a cat. 	I can easily draw a cat.
	Move this word	I do, like Mary not. 	I do not like Mary.

Appendix F- Weekly Written Reflection Questions

HAFTALIK YAZILI YANSITMA SORULARI

1. Otomatik yazma becerisi değerlendirme sisteminin İngilizce yazma becerisine sağladığı düzeltici geri bildirim hakkında ne düşünüyorsun?

.....

.....

.....

.....

.....

.....

2. Bu hafta otomatik yazma becerisi değerlendirme sistemini kullanırken herhangi bir sorunla karşılaştın mı? Karşılaştıysan, sorunun ne olduğunu açıklar mısın?

.....

.....

.....

.....

.....

3. Otomatik yazma becerisi değerlendirme sistemini kullanarak ne tür hatalarını düzeltebildin? Örnekleyebilir misin?

.....

.....

.....

.....

.....

4. Otomatik yazma becerisi değerlendirme sisteminin sağladığı düzeltici geri bildirim ile düzeltilmediğin hatalar oldu mu? Olduysa bu hatalar nelerdir?

.....

.....

.....

.....

Appendix G- Semi- structured Interview Questions for Students

ÖĞRENCİLER İÇİN GÖRÜŞME SORULARI

1. Bu araştırmadan önce otomatik yazma becerisi değerlendirme sistemini kullanmış mıydın?
2. Otomatik yazma becerisi değerlendirme sisteminin İngilizce yazma becerisine sağladığı geri bildirim hakkında ne düşünüyorsun?
3. Bu sistemin sağladığı geri bildirim doğrultusunda İngilizce yazarken yaptığın hatalarını düzeltmek hakkında ne düşünüyorsun?
4. Bu sistemin sağladığı düzeltici geri
5. bildirim doğrultusunda tüm hatalarını düzeltebildin mi? Eğer tüm hatalarını düzeltemediysen ne tür hataları düzeltemediğini düşünüyorsun?
6. Otomatik yazma becerisi değerlendirme sisteminin sağladığı düzeltici geri bildirim yazma becerini geliştirmendeki rolü nedir?
7. Otomatik yazma becerisi değerlendirme sistemini kullanırken herhangi bir sorunla karıştın mı? Eğer karşılaştıysan, bu sorunlar nelerdir?
8. Otomatik yazma becerisi değerlendirme sisteminin güçlü ve zayıf yanları nelerdir?
9. Otomatik yazma becerisi değerlendirme sisteminin sağladığı geri bildirim ile öğretmenlerin yazma becerisi için sağladığı geri bildirim arasında benzerlikler var mı? Varsa nelerdir?
10. Otomatik yazma becerisi değerlendirme sisteminin sağladığı geri bildirim ile öğretmenlerin yazma becerisi için sağladığı geri bildirim arasında farklılıklar var mı? Varsa nelerdir?
11. Yazma becerini geliştirmek için otomatik yazma becerisi değerlendirme sisteminin sağladığı geri bildirimi mi yoksa öğretmenlerin sağladığı düzeltici geri bildirimi mi tercih edersin? Neden?
12. Otomatik yazma becerisi değerlendirme sistemini gelecekte yazma becerini geliştirmek ve hatalarını düzeltmek için kullanmak ister misin? Neden?

Appendix H- Semi-structured Interview Questions for Teachers

ÖĞRETMENLER İÇİN GÖRÜŞME SORULARI

- 1) Daha önce hiç otomatik yazma becerisi değerlendirme sistemi kullandınız mı?
- 2) Otomatik yazma becerisi değerlendirme sistemini kullanırken herhangi bir sorunla karşılaştınız mı? Eğer karşılaştıysan, ne tür sorunlarla karşılaştığını açıklar mısınız?
- 3) Otomatik yazma becerisi değerlendirme sisteminin sağladığı geri bildirim hakkında ne düşünüyorsunuz?
- 4) Bir akademik yazma becerisi öğretmeni olarak sizin sağladığınız düzeltici geri bildirim ve otomatik yazma becerisi değerlendirme sisteminin sağladığı geri bildirim arasında benzerlikler var mı? Varsa nelerdir?
- 5) Bir akademik yazma becerisi öğretmeni olarak sizin sağladığınız düzeltici geri bildirim ve otomatik yazma becerisi değerlendirme sisteminin sağladığı geri bildirim arasında farklılıklar var mı? Varsa nelerdir?
- 6) Otomatik yazma becerisi değerlendirme sisteminin sağladığı düzeltici geri bildirimin süreç-odaklı yazma aşamalarındaki yeterliliği hakkında ne düşünüyorsunuz?
- 7) Öğrencilerin yazma becerilerini geliştirmek için otomatik yazma becerisi değerlendirme sistemini kullanmaları hakkında ne düşünüyorsunuz?
- 8) Yazma becerisi derslerinde otomatik yazma becerisi değerlendirme sistemi kullanmayı düşünür müsün? Eğer kullansaydın, bu sistemi ne şekilde kullanırdınız?
- 9) Otomatik yazma becerisi değerlendirme sisteminin sağladığı yazma becerisi değerlendirme puanları hakkında ne düşünüyorsunuz?
- 10) Otomatik yazma becerisi değerlendirme sisteminin herhangi bir değişikliğe ya da gelişime ihtiyacı olduğunu düşünüyor musunuz? Eğer ihtiyacı varsa, ne tür bir değişiklik veya gelişime ihtiyacı olduğunu açıklar mısınız?

Appendix I – Approval from the Ethical Committee of Anadolu University

Evrak Kayıt Tarihi: 14.06.2019

Protokol No: 48348

Tarih: 24.06.2019



ANADOLU ÜNİVERSİTESİ
SOSYAL VE BEŞERÎ BİLİMLER BİLİMSEL ARAŞTIRMA VE YAYIN ETİĞİ KURULU
KARAR BELGESİ

ÇALIŞMANIN TÜRÜ:	Yüksek Lisans Tez Çalışması
KONU:	Eğitim Bilimleri
BAŞLIK:	Otomatik Yazma Becerisi Değerlendirme Sisteminin Yabancı Dil Olarak İngilizce Öğretilen Ortamlarda Kullanılması: Sistemin Etkinliği ve Öğretmenler ile Öğrencilerin Görüşleri
PROJE/TEZ YÜRÜTÜCÜSÜ:	Dr. Öğr. Üyesi S. İpek KURU GÖNEN
TEZ YAZARI:	Hilal YILDIZ
ALT KOMİSYON GÖRÜŞÜ:	-
KARAR:	Olumlu
Prof.Dr. Emel ŞIKLAR (Başkan-İkt. ve İdari Bil. Fak.)	
Prof.Dr. T. Volkan YÜZER (Başkan Yardımcısı-Açıköğretim Fak.)	Prof.Dr. Esra CEYHAN (Eğitim Fak.)
KATILMADI Prof.Dr. Münevver ÇAKI (Güzel Sanatlar Fak.)	Prof.Dr. M. Erkan ÜYÜMEZ (İkt. ve İdari Bil. Fak.)
Prof.Dr. Handan DEVECİ (Eğitim Fak.)	Prof.Dr. Hasan TUTAR (İkt. ve İdari Bil. Fak.)

Appendix J – Sample Materials Used in the Training Session of Students

Log In Screen

[HELP](#) | [CLIENT SERVICES](#)



Sign In to Your Criterion® Account

The Criterion® Online Writing Evaluation service from ETS is a web-based instructional writing tool that helps students, plan, write and revise their essays guided by instant diagnostic feedback and a Criterion score. [Learn more.](#)

- **Students:** Have more opportunities to practice writing at their own pace, get immediate feedback and revise essays based on the feedback.
- **Teachers:** Can decrease their workload and free up time to concentrate on the content of students' work and teach higher level writing skills.
- **Administrators:** Can make data driven decisions and easily monitor writing performance across and classroom writing performance.

Quick Links

- [System Requirements \(PDF\)](#)

System Announcement

The new version of Criterion will launch on Monday, August 19, 2013. In preparation for the launch, Criterion will be unavailable beginning Monday, August 12, 2013 at 9:00 a.m. Eastern through Monday, August 19, 2013 at 9:00 a.m. Eastern. Please plan accordingly.

Thank you for your patience.

Returning User

User name:

Password:

[Sign In](#)

[Forgot User Name](#) | [Forgot Password](#)

New User

- You will need an access code from your Administrator or instructor to create an account.
- If you used Criterion with a previous institution or class, please sign in with that user name and password. There is no need to create a new account.

[Create Account](#)

By logging in, subscriber agrees that the privacy of student and instructor information, essays, and score data are the responsibility of the instructor and client institution. Student and instructor names, user identification, passwords, essays, and score data are maintained by ETS on a secure server. ETS does not disseminate student or instructor information and it is only accessible to the instructor, designated Subscriber Administrator, and ETS Account Manager. The Subscriber acknowledges that prior to using Criterion, said Subscriber is obligated to obtain any written parental consent that may be required in accordance with COPPA, Section 6.002.

Criterion® home page

[HOME](#) | [CLIENT SERVICES](#) | [RESOURCES](#) | [HELP](#)

Jan Strohl | Student | [Sign Out](#)

Got a new Access Code? Connect to an additional institution, a higher role in your institution, or a new class as an instructor or student: [Enter Access Code](#)

- [Registration Instructions for Bookstore Students](#)
- [Student Quick Access Guide](#)
- [Student Guide for Beginners](#)

– Notices

Oct 28, 2013 05:23:16 PM CDT	You are now connected to class Strohl-S0-11-3.
Oct 28, 2013 04:42:15 PM CDT	You are now connected to class Strohl-S0-11-3.
Oct 23, 2013 06:29:52 PM CDT	A new assignment Exchange Student(2) is added in class Criterion Training
Oct 23, 2013 06:28:28 PM CDT	A new assignment Exchange Student is added in class Criterion Training
Oct 17, 2013 02:37:08 PM CDT	You are now connected to class Strohl-EngB-4.

More notices...

Classes [Print](#)

Class	Grade	Instructors	End Date	Account Owner	Status
CWET-Strohl, Jan, Strohl District, Strohl School					
Criterion Training	None	Ms. Jan Strohl	Jul 01, 2014	CWET-Strohl, Jan	Active

Begin an assignment

ETS Criterion

HOME | CLIENT SERVICES | RESOURCES | HELP

Nala Strohl Student Sign Out

English 9 Class: English 9, CWET-Strohl, Jan. Strohl District, Strohl Sch

Activity Assignments Click the Assignment tab.

View assignment text for any un-timed assignment. Start an assigned plan or view a saved plan. Start a response or view a submitted response attempt. View available results for submitted attempts. Select a peer name to review and comment on their submitted attempt.

Print

Assignment	Plan	Response	Results	Peer Review	Start Date	End Date	Status
Leadership	Start Plan	Start Response			Jun 26, 2013 06:11:11 PM CDT	Jul 02, 2013 08:59:59 AM CDT	4 of 4 Attempts Available

Leadership

A leader is defined as a person who directs, commands or guides a group or activity. Think about someone you know who has proven to be a great leader, either in leading a country or team or organization. Write an essay about this leader, explaining the characteristics they have which would encourage people to follow them. Be as specific as possible in illustrating the attributes that this leader has exhibited which would lead others to respect, admire and follow them.

Legal | Privacy & Security | About Criterion

Copyright © 2013 by ETS. All rights reserved.
ETS, the ETS logo, LISTENING, COGNITIVE, LEADERSHIP, and CRITERION are registered in the United States of America and other countries throughout the world.

ETS Learning Learning

View a topic.

Create a plan.

Begin a Draft.

Create a plan

ETS Criterion

HOME | CLIENT SERVICES | RESOURCES | HELP

Nala Strohl Student Sign Out

English 9 Class: English 9, CWET-Strohl, Jan. Strohl District, Strohl Sch

Activity Assignments

View Student plan, response, and results for this assignment. View Student Assignment page. Export Student assignment.

Leadership

Export

Plan Response

Leadership

A leader is defined as a person who directs, commands or guides a group or activity. Think about someone you know who has proven to be a great leader, either in leading a country or team or organization. Write an essay about this leader, explaining the characteristics they have which would encourage people to follow them. Be as specific as possible in illustrating the attributes that this leader has exhibited which would lead others to respect, admire and follow them.

Additional Material:

Outline List Idea Tree Free Writing Idea Web Compare & Contrast Cause & Effect Persuasive Choose a plan type.

Using an outline is a good way to help organize your ideas about the topic. You can plan your main ideas first and then, under each main idea, you can list some examples, reasons, or details that help support this main idea.

Clear All Start a New Plan

Thesis

Main Idea

Support

Support

Support

Save

Click Save when your plan is complete.

Start Response

Enter text into your plan.

View your completed plan

ETS Criterion

HOME | CLIENT SERVICES | RESOURCES | HELP

Nala Strohl Student Sign Out

English 9 Class: English 9, CWET-Strohl, Jan. Strohl District, Strohl Sch

Activity Assignments

View Student plan, response, and results for this assignment. View Student Assignment page. Export Student assignment.

Leadership

Export

Plan Response

Leadership

A leader is defined as a person who directs, commands or guides a group or activity. Think about someone you know who has proven to be a great leader, either in leading a country or team or organization. Write an essay about this leader, explaining the characteristics they have which would encourage people to follow them. Be as specific as possible in illustrating the attributes that this leader has exhibited which would lead others to respect, admire and follow them.

Additional Material:

Outline List Idea Tree Free Writing Idea Web Compare & Contrast Cause & Effect Persuasive

Using an outline is a good way to help organize your ideas about the topic. You can plan your main ideas first and then, under each main idea, you can list some examples, reasons, or details that help support this main idea.

Edit Start a New Plan

Outline

Abraham Lincoln was a great leader of our country.

He was a great thinker.

He listened not only to his fellow politicians of the time but to common people.

He sought other opinions.

He thought deeply about issues.

He set a great example.

He bore personal sacrifices and grief with dignity.

He encouraged cooperation.

He was a father figure for the country.

He took action.

He made and executed difficult decisions about the war.

He took necessary steps to end the war.

He worked with compassion to bind the nation's wounds.

Click here to begin your draft, or click the Response tab.

Start Response

Craft your response

Click the plan tab to toggle between your plan and the response. Copy and paste from your plan.

The screenshot shows the 'Leadership' assignment page. At the top, there are tabs for 'Plan' and 'Response'. A red arrow points to the 'Plan' tab with the text 'Click the plan tab to toggle between your plan and the response. Copy and paste from your plan.' Below the tabs, there is a 'Leadership' section with a description of a leader and a 'Response' section with a large text area. A red arrow points to the 'Response' section with the text 'Review the instructions.' Another red arrow points to the text area with the text 'Type your essay directly into this box, or copy and paste your response from a word processing document.' At the bottom, there are 'Save' and 'Submit' buttons. A red arrow points to the 'Save' button.

Remember to Save frequently as you type. Click Submit when the essay is ready for evaluation.

Copy and Paste an Essay

The screenshot shows a 'Paste Essay' dialog box. It has a title bar that says 'Paste Essay'. Below the title bar, there is a text area with the instruction 'Please paste your essay into the text box below. Click OK to continue.' Below the text area, there are 'OK' and 'Cancel' buttons.

You may copy and paste an essay from a word processing program. Press Ctrl-V, and a text box will appear. Paragraphing may be stripped out during copy and paste, so you will need to check this after pasting text. To paste the text into the *Criterion* writing screen, click OK.

Before you click Submit

Submit

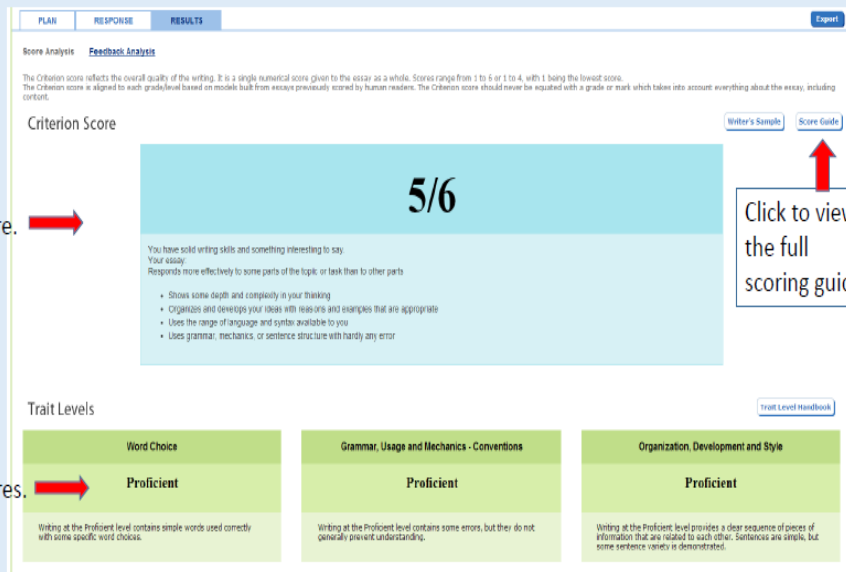
Recheck the content of your essay by asking yourself these questions:

1. Do I have an introductory paragraph that creates interest in the essay, lets the reader know the general topic and tells the reader the main idea (thesis) of the essay?
2. Have I used topic sentences to address each of my main ideas and to announce the idea of each paragraph, tell the reader where the essay is going or to make a bridge between two paragraphs or main ideas?
3. Do my supporting ideas help to convince the reader that my main ideas are good ones? Have I told a story, given examples, reasons, details, or used metaphors or analogies to help the reader understand my ideas?
4. Do I have a concluding paragraph that restates my thesis, summarizes my main points and leaves the reader thinking about my topic?
5. Did I use transitional words and phrases to take the reader smoothly from one idea to the next?

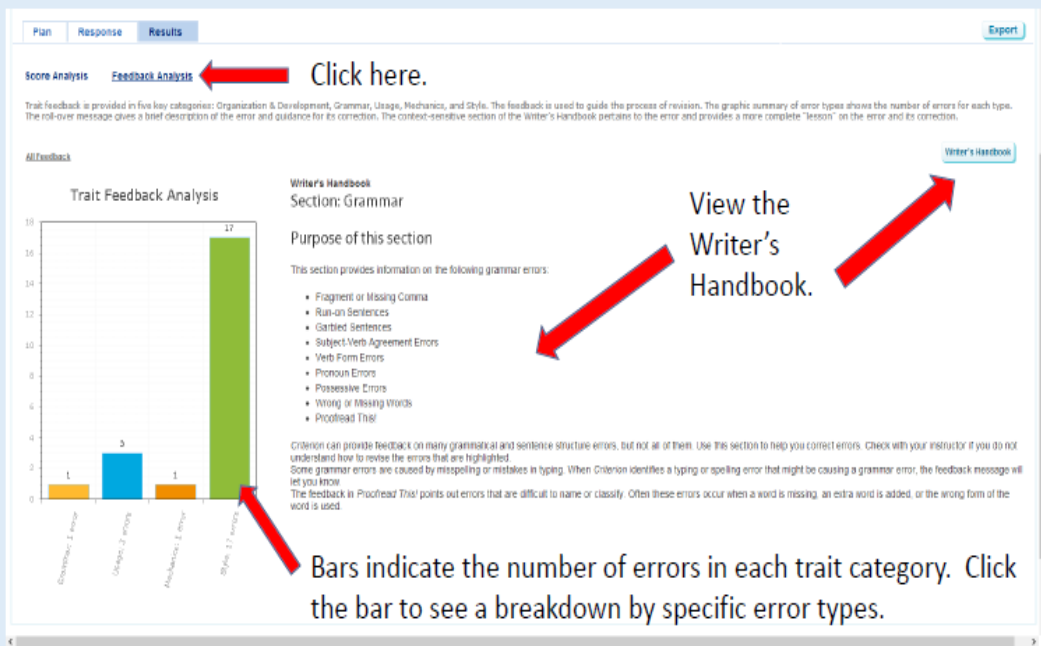
View your feedback

View your Criterion Score.

View your Trait Level scores.

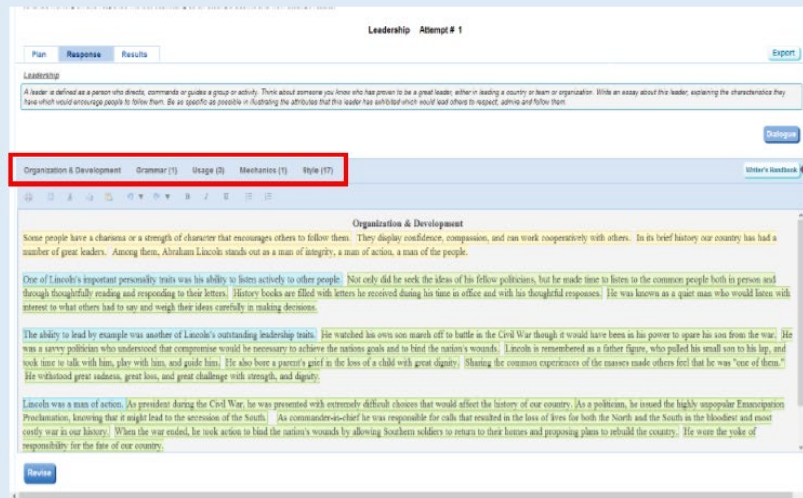


View the Feedback Analysis Chart



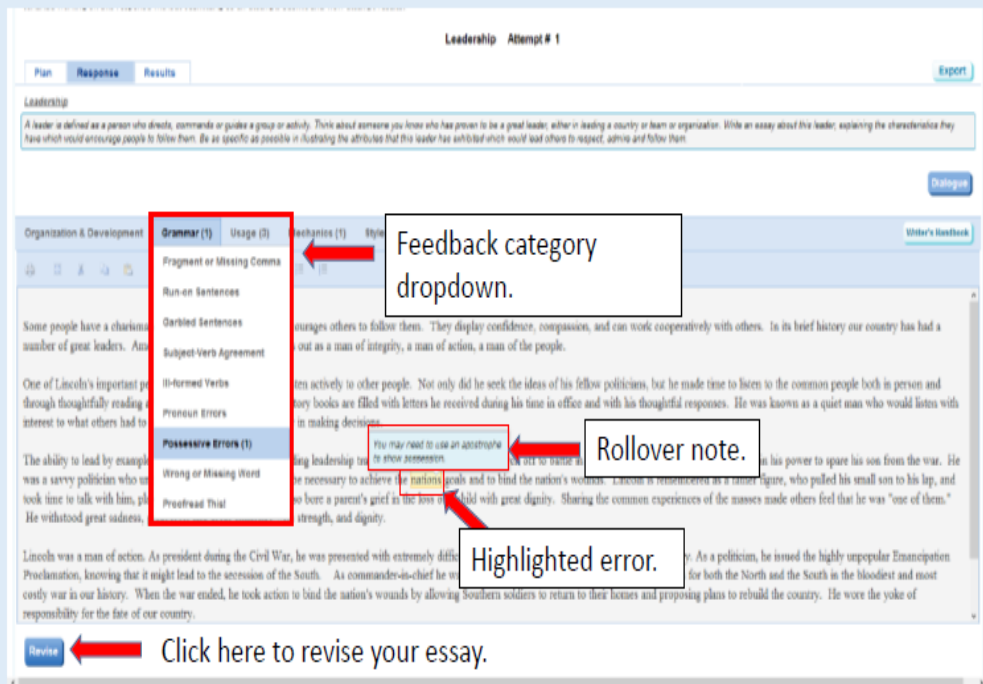
View specific trait feedback for your essay

From the Response tab, click on each of the five categories to view specific feedback on your essay. Errors will be highlighted and roll over notes will provide information on each error.



Click the Writer's Handbook for more information on any identified error.

Review your errors



Feedback category dropdown.

Roll over note.

Highlighted error.

Click here to revise your essay.

Revise your essay

When all errors have been corrected, review the content of your essay for further needed improvements. Check the Score Guide for guidelines.

Click here to submit your revised draft.

View error feedback on left and enter your corrections on the right.

Before you click Submit

Submit

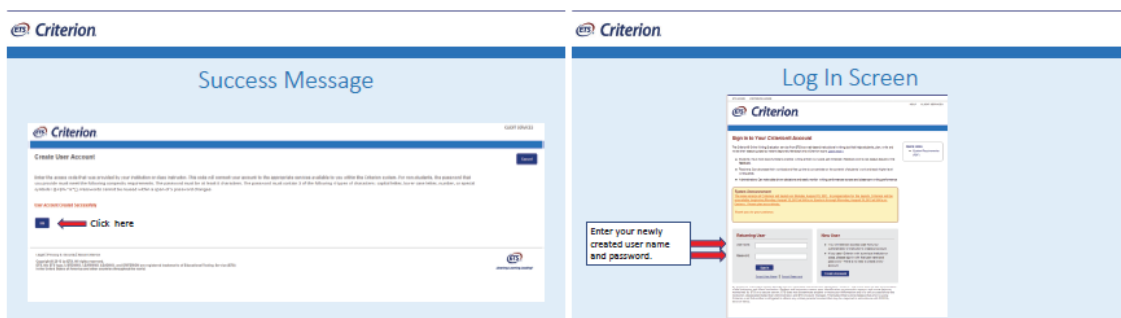
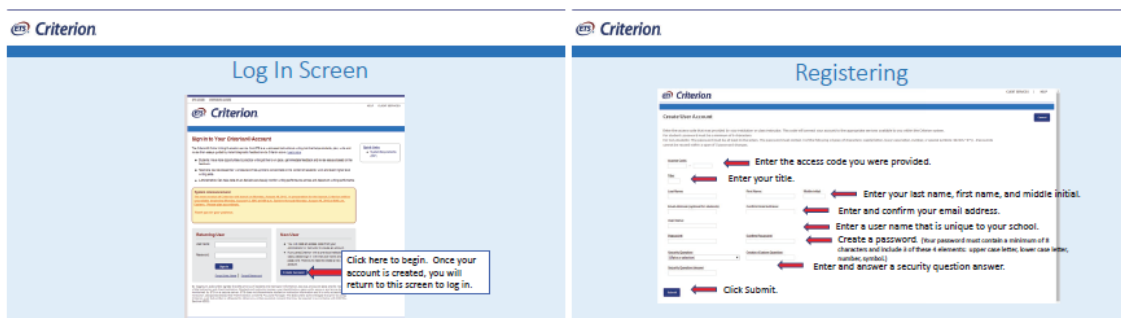
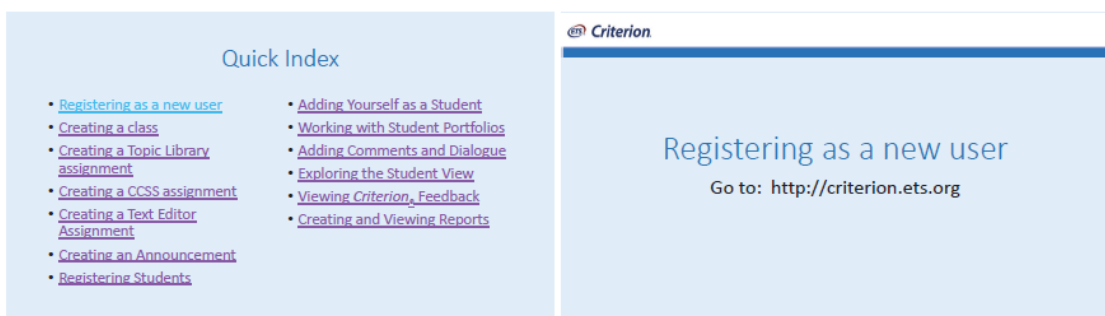
Recheck the content of your essay by asking yourself these questions:

1. Do I have an introductory paragraph that creates interest in the essay, lets the reader know the general topic and tells the reader the main idea (thesis) of the essay?
2. Have I used topic sentences to address each of my main ideas and to announce the idea of each paragraph, tell the reader where the essay is going or to make a bridge between two paragraphs or main ideas?
3. Do my supporting ideas help to convince the reader that my main ideas are good ones? Have I told a story, given examples, reasons, details, or used metaphors or analogies to help the reader understand my ideas?
4. Do I have a concluding paragraph that restates my thesis, summarizes my main points and leaves the reader thinking about my topic?
5. Did I use transitional words and phrases to take the reader smoothly from one idea to the next?

Review a Peer's Essay

Click the name of a member of your peer review group to comment on his or her submitted essay.

Appendix K – Sample Materials Used in the Training Session of Instructors



Home Page

As you create classes and add assignments, they will be listed under notices.

From the row of icons you can perform the following tasks:
View - Check the box in front of the name of the class, and view current activity within the class.
Add - Create a class. Criterion automatically generates the record codes for instructors and students.
Edit - Edit class information.
Announcements - Generate a dated and/or default announcement to be sent to all class members.
Connect - Search for an additional instructor from your school by name and connect him/her to your class.
Disconnect - Disconnect an instructor currently registered to your class.

Adding a Class

Adding a Class

Click here to add a new class.

Entering a Class Description

1. Select your school from the Level drop down.
2. Enter a class name. This must be unique to your school.
3. Select the grade level of the class.
4. Select the Writer's Handbook that best meets the needs of the entire class.
5. Check the appropriate dictionary.
6. Add an optional Class End Date.
7. Check your time zone. This should come through by default.
8. Enter the courtesy name by which students address you.
9. Click Save when all information is complete.

Creating Assignments

Choosing a Class

Click on the name of the class.

Adding an Assignment

1. Click the Assignments tab.
2. Click Add.

Creating an Assignment

1. Choose whether or not to select an administrator assignment.
2. Choose an assignment type from the drop down menu: Topic Library, CSD (Common Core State Standards), Scored Instructor Topic or Test Editor.
3. Choose Level, Mode, and Prompt from the dropdown menus. Click View Topic Library to view full text of library prompts.
4. If you choose a prompt from the Topic Library dropdown, the name and text of the prompt will autofill Assignment Name and Assignment Text.
5. Add up to 7 URL reference links to any assignment for multi-stimulus prompts, custom plan templates, etc. Include the URL and a display name for each.

Registering Students

Registering Students

Select the Roster tab.

There are three ways to register students:

1. To allow students to self-register, click Get Access Code. This generates a student access code for the class. Provide this to students so that they can self-register. If students are already registered for another class or have been imported into the school, click Connect Student. You can then enter the student last name in the search box. Check the box in front of the name of a student and click Connect. Leave the search box blank to pull up a list of all registered students in your school. Check all boxes and click Connect to connect students all at once.
2. Register students by clicking Add Student. Enter a generic password such as "student" that students will be prompted to change at their first login.

Adding Yourself as a Student

Getting the Student Access Code

If you have already created a class and recorded or printed the access code, you can easily connect to a class as a student. This makes it easy to experiment with Criterion from a student point of view and to demonstrate for your students.

If you do not know the access code for your class, check the box in front of the name of your class and click the Get Access Code button.

Connecting as a Student

Enter your student access code and click Connect. Confirm you are being added as a student to the class, and click Connect again.

Connect

Moving Between Instructor and Student Views

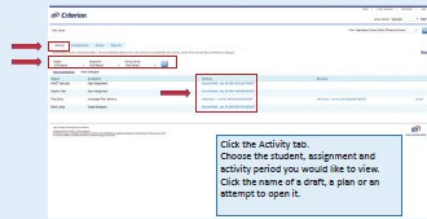
To switch from instructor to student (or vice versa), simply choose your role from the drop-down menu at the top right of your home screen.

Working with Student Portfolios

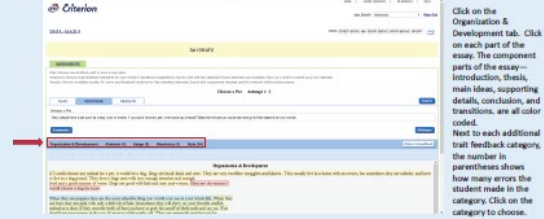
Selecting a Class

Click on the class name to select a class.

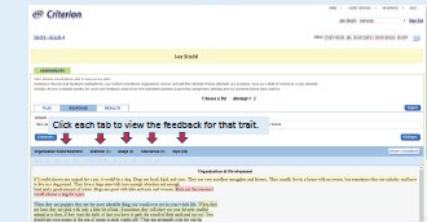
Viewing Activity



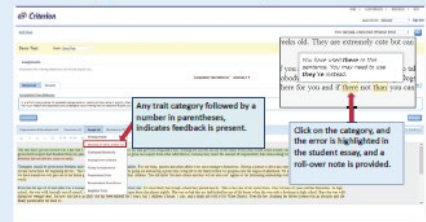
Reviewing Criterion Results



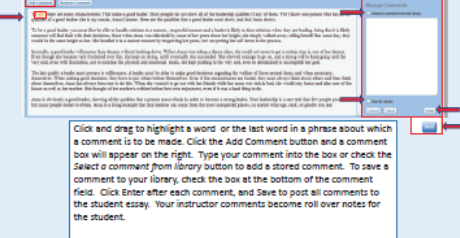
Reviewing Trait Feedback



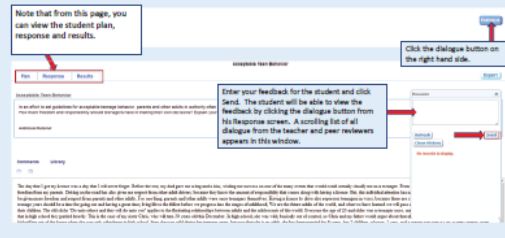
Using Feedback Drop Down Menus



Adding Comments



Entering Dialogue

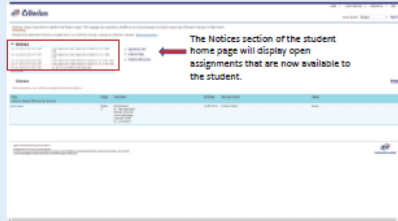


Exploring the Student View

Viewing Class Activity

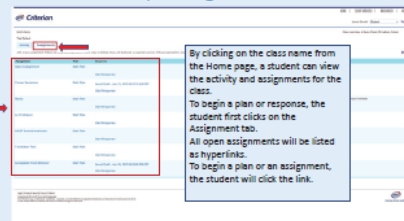


The Student View



The Notices section of the student home page will display open assignments that are now available to the student.

Opening the Class



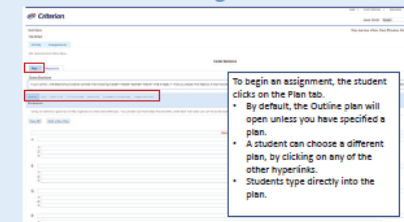
By clicking on the class name from the Home page, a student can view the activity and assignments for the class. To begin a plan or response, the student first clicks on the Assignment tab. All open assignments will be listed as hyperlinks. To begin a plan or an assignment, the student will click the link.

Viewing Assignments



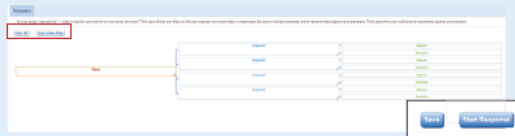
By clicking the Assignments tab, a student can see a list of all assignments currently available. Results from submitted work, peer review, start and end dates and status are displayed.

Creating a Plan



To begin an assignment, the student clicks on the Plan tab. By default, the Outline plan will open unless you have specified a plan. A student can choose a different plan, by clicking on any of the other hyperlinks. Students type directly into the plan.

Saving the Plan



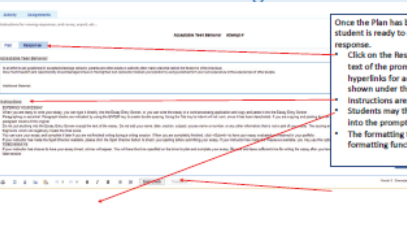
Students type directly into the plan. Buttons allow them to Clear All or Start a New Plan. Once the Plan is complete, the student clicks the Save button in the lower right corner. Only the current plan can be saved for an essay.

Using the Saved Plan



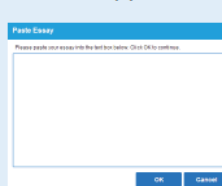
Students can toggle between the Plan and Response tabs to view, copy and paste from the plan.

Student Assignment View



Once the Plan has been saved, the student is ready to construct his response. Click on the Response Tab. The full text of the prompt and a list of hyperlinks for additional material are shown under the assignment name. Instructions are provided. Students may then enter text directly into the prompt box. The formatting tool bar provides formatting functions.

Copy and Paste an Essay



A student may copy and paste an essay from a word processing program. In that case, a text box will appear. Paragraphing may be stripped out during copy and paste, so students will need to check this after pasting text. To paste the text into the Criterion writing screen, click OK.

Criterion

Formatting Tool Bar

Print Select All Cut Copy Paste Undo Redo Bold Italic Underline Bulleted List Numbered List Spell Check Thesaurus

Criterion

Starting a Response

Students begin composing in text area. They may also copy and paste from a word processing document. They will paste first into a pop-up box.

- Note the Save button on the far left. Students should Save frequently while working to prevent data loss.
- When the essay is complete, the student clicks the Submit button on the lower right to receive Criterion feedback on his work.

Criterion

Viewable Autosave for Essay Responses

Timestamps of the last 'Autosave' of a Response is Now Viewable to Students!

Saved: Oct 22, 2013 04:14:44 PM EDT

Criterion

Viewing feedback

View Criterion Score: 5/6

Click to view the full scoring guide.

Click for more information.

View Trait Level scores.

Criterion

Viewing the Feedback Analysis Chart

Click here.

View the Writer's Handbook.

Click here to drill down further to specific errors.

Criterion

Viewing specific trait feedback for an essay

Click on each of the five categories to view specific feedback on the essay. Errors will be highlighted and roll over notes will provide information on each error.

Clicking the Writer's Handbook provides more information on any identified error.

Click here to save during revision.

Criterion

Reviewing errors

Feedback category dropdown.

Rollover note.

Highlighted error.

Click here begin revising an essay.

Criterion

Making corrections

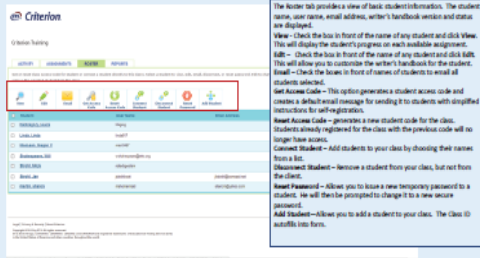
Students click here for comments and dialogue from instructor and peer reviewers.

Click here to save during revision.

Click here to submit a revised draft.

View error feedback on left and enter corrections on the right.

Exploring the Roster Tab



The Roster tab provides a view of basic student information. The student name, user name, email address, writer's handbook version and status are displayed.

View - Check the box in front of the name of any student and click View. This will display the student's progress on each available assignment.

Edit - Check the box in front of the name of any student and click Edit. This will allow you to customize the writer's handbook for the student.

Email - Check the boxes in front of names of students to email all students selected.

Reset Access Code - This option generates a student access code and creates a default email message for sending it to students with simplified instructions for self-registrars.

Reset Access Code - generates a new student code for the class. Students already registered for the class with the previous code will no longer have access.

Connect Student - Add students to your class by choosing their names from a list.

Disconnect Student - Remove a student from your class, but not from the class.

Reset Password - Allow you to issue a new temporary password to a student. He will then be prompted to change it to a new secure password.

Add Student - Allows you to add a student to your class. The Class ID autofills into form.

Creating Reports

Report Types

- **Score Analysis Report**—shows student identifying information, *Criterion*, score, question, essay, dialogue and comments from instructor and peer reviews.
- **Expanded Performance Detail Report**—shows identifying information, *Criterion*, score, question, response and footnotes identifying all *Criterion*, identified errors within each trait level type.
- **Expanded Performance Summary**—shows identifying information, *Criterion*, score, question, summary of trait feedback analysis by error type, essay.

Report Types

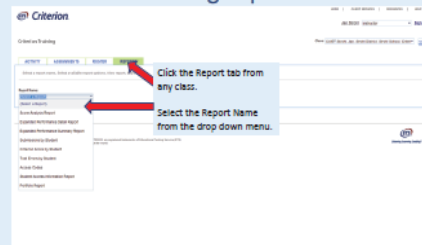
- **Submissions by Student**—shows all submissions made by the student to date
- **Criterion Score by Student**—shows student *Criterion* scores for designated assignments, number of attempts, average score, percent of attempts and advisories
- **Trait Errors**—shows how many errors and what types from the diagnostic trait feedback, the number of attempts with specific errors, and the percentage of attempts with a specific error type.
- **Access Codes**—shows current instructor and student access codes for the selected class
- **Student Access Information Report**—set a date range to view user information and last sign in date
- **Portfolio Report**—Links to the student portfolio and assignments

Creating Reports: Selecting a Class



Click on the class name to select a class.

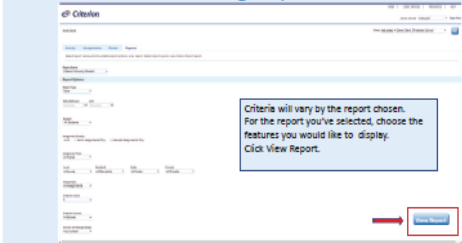
Selecting Reports



Click the Report tab from any class.

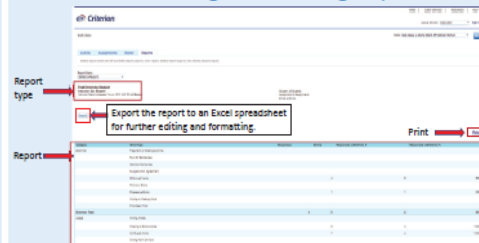
Select the Report Name from the drop down menu.

Selecting Report Criteria



Criteria will vary by the report chosen. For the report you've selected, choose the features you would like to display. Click View Report.

Viewing and Printing Reports



Export the report to an Excel spreadsheet for further editing and formatting.

Print