UiT The Arctic University of Norway

Faculty of Humanities, Social Sciences and Education

The Effect of AI-Based Applications on EFL Writing Skill Development:

An Inquiry into Integration of AI into Language Learning

Fatemeh Etaat

Master's thesis in English Linguistics, ENG-3991, May 2024

Dedication

To my beloved ones, my adorable husband and simply perfect family for their unconditional love and support

Abstract

Aims and objectives: The main aim of the present study was to investigate the impact of AI-based writing applications on the development of EFL learners' writing skills. In addition, it was aimed to examine which aspects of writing- task fulfillment, organization, grammar, lexis, and mechanics- are more significantly affected. Furthermore, it was investigated whether using AI-based writing applications contributes to time efficiency in the writing process. Finally, the learners' and the teacher's attitude toward AIEd specially in the process of language learning was assessed.

Methodology: The study compared two groups of Iranian EFL learners in terms of writing skill development. The experimental group used two AI-based writing tools, *"Wordtune"*, and *"Insta text"* for doing the home assignments for 36 sessions over the study. While the control group followed the traditional pen-and-paper writing technique. Both groups were asked to record the average time spent per assignment, and also there were several essay writing tests including a pretest, 6 periodic tests, and a posttest. In addition, the learners' and the teacher's attitude toward AIEd were collected through questionnaires.

Data Analysis: The data collected from the scores of the several writing tests as well as the data related to the amount of time spent for the assignments were analyzed in RStudio, R version 4.1.2 (R Core Team, 2022). The lme4 R package (Bates et al., 2015) was used to fit the linear mixed-effects models to examine how the scores vary among different participants based on the type of group and kind of test.

Findings: The findings indicate that as a result of receiving automatic written corrective feedback (AWCF) provided by AI-based writing tools, the experimental group improved in overall writing skills, primarily in mechanics and lexis, followed by grammar, and finally task fulfillment and organization as the least affected feature. Moreover, analysis of time spent per assignment revealed that using AI-based writing tools accelerated the process of text generation. Furthermore, both the teacher in charge of the study and the students had a positive attitude toward AIEd.

Significance: the study adds to the existing data about the integration of AI into language learning specially for writing skills and contributes to the ongoing discussion on the role of technology in language learning specifically in an EFL context.

Keywords: FEL learners, writing skills, AI-assisted writing tools, AI, AIEd, AWCF, AWE

Acknowledgment

The process I went through in my M.A course and writing the present thesis helped me achieve many valuable things. First, I am deeply grateful to God for all the support he always provides me. I feel also obliged to express gratitude to a number of individuals whose contribution in assorted ways deserve special acknowledgement.

I would like to express my heartfelt thanks to my supervisors, Björn Lundquist and Gillian C Ramchand, whose exceptional supervision, guidance, and advice helped me navigate the process of the present thesis. I extend my deepest gratitude to the faculty members of the language department at UiT for their invaluable instructions during the course.

Words fail me to express my gratitude to my adorable husband, Hadi, who always believed me, and his boundless love, continuous support, and encouragement have paved the way toward my dreams. I have also been blessed to have such supportive family and friends. Dedicating this thesis, my heart goes to my beloved husband, my cherished family and all my dear friends.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

AES: Automated Essay Scoring

AI: Artificial Intelligence

AIEd: Artificial Intelligence in Education

AWCF: Automated Written Corrective Feedback

AWE: Automated Writing Evaluation

CF: Corrective Feedback

CGF: Computer-Generated Feedback

CMCF: Computer-mediated Corrective Feedback

EC: EssayCritic

EFL: English as a Foreign Language

ESL: English as a Second Language

ETS: Educational Testing Service

GRE: Graduate Record Examination

L1: First Language

LLS: Language Learning Strategy

PEG: Project Essay Grade

PPMH: Push-Pull Mooring-Habit

TOEFL: Test of English as a Foreign Language

WCF: Written Corrective Feedback

1 Introduction

Learning English as a global language is becoming increasingly essential for success in a variety of contexts including personal, professional, and academic. Effective writing is essential in both academic and professional environments (Lahuerta, 2017; McDonough & Crawford, 2018; Rosário, 2019). In academic environments, strong writing skills lead to successful communication of ideas as well as understanding resources and synthesizing information, which ultimately enhance critical thinking. Similarly, in professional environments, producing quality writing allows the clear expression of knowledge and leads to effective communication at work without misunderstandings. However, students always find writing skills demanding (Hamzaoui, 2021; Ruscetti et al., 2018), particularly EFL (English as a Foreign Language) learners find it a big challenge (Alsied & Ibrahim, 2017; Hanauer et al., 2019).

Based on Stern (1983), ESL (English as a Second Language) and EFL are different in terms of several aspects including learning environment, language functions, learning goals, teaching methods, and learning approaches. As Krashen (1985) argued, second language acquisition involves subconscious acquisition just like the process of mother tongue acquisition. In other words, ESL takes place during the course of social communication in the target language environment, while EFL occurs in learners' native language environment. Therefore, EFL learners use English only in specific circumstances and for particular purposes (Kirkpatrick, 2014; Kanchru, 1992). As a result, because of the limited role of English in the environment, the learning path is even harder, and numerous challenges are encountered.

From the teachers' perspective, ESL learners can simultaneously develop four skills of listening, reading, writing, and speaking, while it is impossible in EFL situation due to several reasons such as learning environment, teaching method, syllabus, content, and individual differences (Zhongde, 2001). To be more specific, ESL teachers benefit from the authentic sources around to pave the way for teaching, while EFL teachers more commonly face challenges of resource deficiencies. In addition, ESL teachers use the target language for teaching, while EFL practitioners have to use the source language to effectively convey the message, which makes the process of learning and teaching slower. Therefore, in countries such as Iran, where English is considered a foreign language, and it is a subject only taught within the educational curriculum, teachers face a lot of obstacles such as challenges with conveying the message properly, finding efficient resource, designing comprehensive lesson plans, and

devoting a lot of time for giving feedback to the learners individually, which is almost impossible to do so.

Generally, the EFL context has some specific socio-linguistic features reported in several studies such as: lack of sufficient exposure to the target language (Muñoz, 2014; Akbari, 2015); gaining instruction in the L1 of the learners (Liu et al., 2004); foreign language anxiety that prevents active participation in target language communities (Hsu, 2015; Gilakjani & Sabouri, 2016); low learning aptitude and lack of motivation since learning another language seems unnecessary (Mercer & Ryan, 2010; Akbari, 2015); and finally challenges related to teachers such as the shortage of competent teachers, lack of availability and time of the teachers practicing in the classroom, and using insufficient teaching methods (Kurniawan and Radia, 2017). In addition, there are several obstacles to effective writing, including: information organization (Phuong, 2021; Singh, 2017); lexis, grammar, and writing mechanics (Flowerdew, 2019; Komba, 2015; Mahammoda, 2016); and the rules of genres (Bitchener & Basturkmen, 2006; Finn, 2018).

Thus, EFL contains its own distinct set of difficulties related to possible gaps in syntax, lexis, grammar, pragmatic understanding, which make it relatively challenging. As a result, in the process of learning a foreign language, making errors is inevitable and teachers are responsible for correcting them by giving useful feedback. Based on Ur (2006, 242), feedback is "information that is given to the learner about his or her performance of a learning task, usually with the objective of improving this performance". He mentions that two elements of feedback are correction and assessment which are done with the aim of improvement. In other words, feedback is an important communication skill both inside and outside of the classroom (Dignen, 2014).

However, teachers often find it very time-consuming to provide corrective feedback (CF) on students' writings individually, particularly in the large classes with lower levels of proficiency which are mostly teacher centered. Consequently, giving effective CF is very demanding and challenging for many L2 teachers, and researchers debate how to provide learners with efficient CF and complain about not having enough time giving feedback individually (Goldin & Ashley, 2012; Bitchener & Ferris, 2012; Laflen, 2020). Moreover, traditional approaches to writing instructions and hand-written feedback by the teachers may no longer adequately address EFL

learners' needs and consequently may not help them improve in writing skills. In fact, it is necessary to extend reading and writing instructions even outside of the classroom using varied methods such as the use of digital resources (Liberg, 2007; Kress, 2010; Edwards-Groves, 2011; Åkerfeldt, 2014). Therefore, new and creative writing practices and tools are needed in today's writing pedagogy to help students improve their writing abilities more. In other words, more diverse approaches to reading and writing can both boost students' interest and motivation as well as giving all students chance to improve their skills (Agélii Genlott, & Grönlund, 2013).

According to studies, when compared to human raters and instructor-provided CF, automatic writing evaluation (AWE) tools can sometimes offer more detailed and consistent CF (Grimes &Warschauer, 2010; Hussein et.al, 2019). In addition, Written Corrective Feedback (WCF) has been studied by many researchers and it is confirmed that WCF will lead to the development of students' writing skills by letting them observe their errors and helping them to correct such errors (e.g., Karim & Nassaji, 2020; Zhai & Ma, 2021; Zhang & Zhang, 2018; Xu & Zhang, 2021). Similarly, teachers may find WCF demanding and time consuming to do for all students. In addition, most of the errors in writing specially those regarding the mechanics of a text such as punctuation, capitalization, and spelling are recurring, and teachers mostly prefer not to waste time on correcting them repeatedly. Considering that, new technological advancement can aid the students' writing enhancement by providing helpful automated feedback (Dikli, 2006; Kukich, 2000; Sireci & Rizavi, 2000; Lee, 2007; Mørch et al., 2005; Winerip, 2012; Lee, et al., 2013).

In recent years, Artificial Intelligence (AI), especially AI in education (AIEd), has emerged as an increasingly prominent technology for learning assistance, particularly in learning languages. Background studies reveal that integration of intelligent technology into education has been implemented with several goals in mind: to modernize education (Chiu, 2021; Chiu et al., 2022; Xia et al., 2022); to boost academic performance by increasing students' cognitive engagement; to reduce educational inequalities by helping under-privileged students (Boninger et al., 2020; Williamson & Eynon, 2020); and to improve efficacy in learning, teaching, assessment and administration (Gonzalez-Calatayud et al., 2021; Luckin, 2017). Furthermore, since AI-based writing tools, such as ChatGPT, Bing chat, Jasper.ai, and iwrite, have developed so quickly, studies have revealed positive outcomes regarding AI-based direct corrective feedback and its impact on the students' development of writing skills (Nobles & Paganucci,

2015; Dickson, 2017; Karyuatri, 2018; Azah, 2019; O'Neill & Russell, 2019; Koltovskaia, 2020; Dong & Shi, 2021), specifically AI-based feedback tools such as "Grammarly" and "Wordtune" (Qassemzadeh & Soleimani, 2016; Ghufron & Rosyida, 2018; Aljohani, 2021; Coenen et al., 2021; Zhao, 2022; Al Mahmud, 2023; Rad et al., 2023).

To elaborate more, AI-based feedback tools such as Grammarly, Wordtune, Insta text, and Quill Bot have several abilities more or less including: assisting the learners with generating a text sentence by sentence after giving the prompt first; predicting the next sentence based on the context; suggesting alternative sentences; detecting various types of errors and correcting them; offering several rewrites and paraphrases; and the ability to expand and summarize the text. Further explanation about the kinds of AI-based feedback tools and their potentials are discussed in the next chapter.

Considering the numerous abilities of such writing tools and also the studies mentioned above, further investigation is needed to show how such tools can be applied in EFL language classes to improve students' writing skills efficiently, and also *how* the process of writing development aided by AI happens. In addition, it is important to know whether all aspects of writing including task fulfillment, grammar, organization, lexis, and mechanics are influenced at the same time, since only a few researchers, and no one in Iran, have studied these detailed outcomes (Al Mahmud, 2023; Rad et al., 2023). Furthermore, almost no study has examined the amount of time spent working with such AI-based feedback tools to see whether the improvement in writing skills is simply due to the amount of time spent practicing writing or whether other factors are significant. Finally, only a few studies attempted to discover the teachers and the students' attitude toward applying such tools in language learning, whether they benefit long term from such applications or whether students simply get lazy and dependent on such tools and will fail in the absence of them. It must be noted that, it is important to know not only the students' but also the teacher's attitude toward technology tools, since they have conclusive impact on the effectiveness of such tools (Jiang et al., 2020).

Therefore, according to the review in Chiu, et al. (2023), research in AIEd domain is still in its exploratory stage, and studies have not yet shown exactly that how AI affects the process of learning and the attainment of students' outcomes, which is a gap the present study is designed to partially fill. According to Holmes et al. (2021), the impact of AI on education remains

unclear and more research is necessary to reveal whether and how such technology affects education. In other words, as they state it is not sufficient to argue that merely AI should be utilized in an educational context. It needs to be investigated that what type of AI technology is needed, how it should be implemented, and what the possible achievements for learners and teachers are (Holmes et al., 2021).

The aim of the present thesis is to investigate the effect of using AI-based writing tools on EFL learners' writing development. The further goal is to examine which aspects of writing- task fulfillment, organization, grammar, lexis, and mechanics- are affected most by use of these tools. In addition, the effect of time spent working with AI-based writing applications as well as the teacher's and the students' attitude towards using feedback tools are included. Two AI-based writing tools are the focus of the investigation: 1) "*Wordtune"*, which is an AI-based text generator and text evaluator, to facilitate writing with suggesting ideas, predicting the next sentence based on the context, and providing AWCF correcting the errors of different types such as grammar, spelling, capitalization, and punctuation; and 2) "*Insta text"*, which is an automatic text evaluator used for the final refinement and correction. The participants in the experimental group are instructed to take the advantage of using "Wordtune" and "Insta text" for writing home assignments over 36 sessions, while those in the control group followed the traditional pen-and-paper technique with teacher-correction feedback. During the course, there are 6 tests distributed throughout the span of three months as well as a pre and a posttest to investigate how the students' writing skills improve over the study, considering five aspects of a text: task fulfillment, grammar, organization, lexis, and mechanics.

This thesis is significant in terms of testing two distinct AI-based tools geared to improving EFL learners' writing development to the greatest extent, which is missed in the previous studies. The immediate feedback and accessibility of such tools are potentially beneficial for EFL students who may suffer from geographical limitations, resource restrictions, or cultural barriers. The findings of the present study will be beneficial for EFL learners, allowing them to take control of their language learning in general, and particularly their writing skills. In addition, it will be useful for the language teachers who can save the time spent on correcting the writings and giving feedback individually, add variety to their teaching plans by applying AI technology in the classroom, motivate students to engage in learning more, and to generally accelerate the learning process.

The present thesis is structured as follows: the next chapter is devoted to the theoretical background about the integration of technology into learning, AI in education (AIEd) paradigm, the related review of literature about using AI in teaching, learning, assessment, and administration, and also the importance of feedback in EFL context along with its different types. In addition, in chapter 3, the methodology section, information about the research questions and assumptions, research design, procedure, participants, materials, and data analysis are outlined. Chapter 4 is devoted to the results of this study a long with the statistical analysis and several tables and graphs. In chapter 5 the findings are discussed thoroughly, and in chapter 6 the conclusion is presented.

1.1 Chapter Summary

This chapter highlights the special challenges that are involved in learning English as a foreign language, and the importance of developing writing skills for which new tools are needed to overcome the shortcomings of traditional methods. In addition, the challenges for EFL learners and teachers are explained where corrective feedback (CF) is essential for writing improvement. However, in an EFL context, giving such feedback can be demanding for teachers, and AWCF can potentially act as supplement to CF provided by teachers. Recent technology advances like AI-based writing applications have emerged to meet the learners' needs and address the possible challenges when learning languages, especially with regard to writing skills.

As discussed, AI-powered writing tools have the potential to develop writing skills by providing instant corrective feedback, personalized learning experiences, creative ideas, and suggesting context-based sentences. However, use of AI in EFL context is in its initial stages and further research is required to explore how it can be integrated into language learning.

2 Background

In this chapter the main focus is first on the integration of technology into language learning specifically in writing skills. Next, the scope of AI in education is discussed in domains of learning, teaching, assessment, and administration. Finally, the role of feedback in increasing language proficiency, types of feedback for writing, computer generated corrective feedback, and automatic written corrective feedback a long with AI-based corrective feedback providers are discussed. In addition, the related background studies are discussed throughout this chapter.

2.1 Integration of Technology into Language Learning

Researchers and practitioners have examined language learning through technology from different perspectives and have stated several functions of technology in: helping the actual learning process happen faster (Rodinadze & Zarbazoia, 2012); increasing social interactions, motivation and engagement of the students (Godzicki, 2013; Baytak et al., 2011; Hennessy, 2005; Arifah, 2014); increasing students' confidence which leads to more cooperation between learners and teachers (Mouza, 2008; Sabzian et al., 2013; Lee, 2001); experiencing real learning and increasing learners' responsibility (Drayton et al., 2010); and increasing exposure to English in a meaningful authentic context (Warschauer, 2000; Parvin & Salam, 2015; Zhao, 2013). Considering this, integration of technology into teaching plans can lead to a more effective and vibrant class than those traditional teacher-centered lecture-based classes.

As Biggs (1996) believes, due to integration of technology, the responsibility for learning has been shifted away from instructors to the learners. Similarly, Brown (2002) states that as a result of incorporation of technology into the curriculum, learners take on the responsibility for most of the work that was previously done in the classroom. Similarly, in EFL classes one of the main goals is to increase self-regulated learning and to encourage the learners to take responsibility for learning, and technological advances can help meet this goal. Drayton et.al (2010) have found out that utilizing technological innovations such as computers and internet in language classes increases learners' responsibilities by making them experience true learning and also encourages them toward self-directed learning. According to, Warschauer (2000), the integration of technology into language learning can happen in the classrooms through two approaches: first, the cognitive approach through which the learners have opportunities to increase their exposure to language meaningfully which leads to an increase in their knowledge of language skills; second, the social approach, which provides more social interactions in

authentic situations so learners can practice real life skills. In relation to this, Lin and Yang (2011) carried out research about the influence of Wiki technology on students' writing skills where the students were invited to join a Wiki page and start interactions with their fellow classmates through writing passages to them and responding back. The findings revealed that the learners significantly improved in terms of vocabulary, spelling, and grammar by reading the writings of their peers and by receiving immediate feedback using such technology. Similarly, in another study by Alsaleem (2014), students experienced a significant improvement in terms of vocabulary, word choice, writing and speaking abilities through using WhatsApp. Thus, integration of technology into language learning can provide more exposure to language which leads to contextual learning and can lessen the difficulties EFL learners confront.

In spite of the wide range of studies, there are still open questions of how to implement technology in language classes to achieve the greatest benefit. According to Ahmadi (2018) who did a literature review on the use of technology for language learning, there are some recommendations for the successful integration of learning into technology: 1) institutions should encourage and support teachers personally and professionally in the training and use of technology; 2) the particular aid should be in line with the educational goals and desired outcomes; 3) institutions should regard technology as a significant part of teaching and learning programs, and students should be encouraged to use it for increasing their language abilities; 4) technology should be implemented for helping teachers towards learner-centered instruction as opposed to teacher-centered instruction.

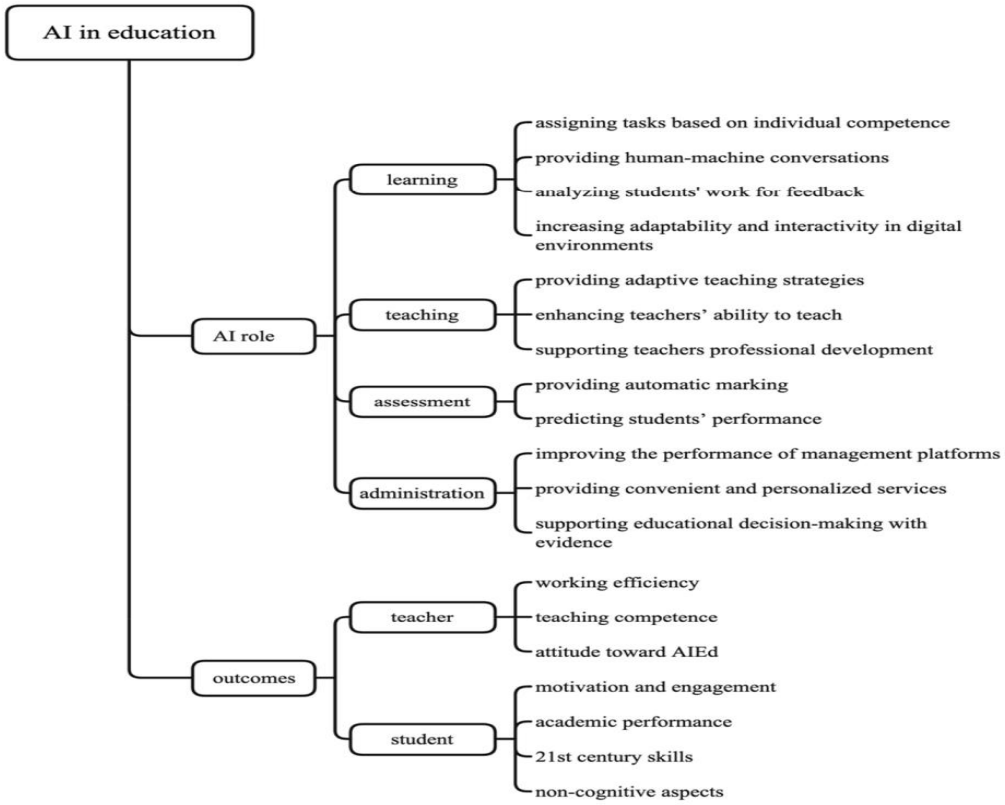
Considering integration of technology into language learning elaborated above, integration of AI into education specially AI in language learning is receiving increased attention these days. The following section presents the domains of integration of AI into education.

2.2 AI in education (AIEd)

Applications of artificial intelligence in education (AIEd) are new to most researchers, teachers, and learners. To describe AIEd, three paradigms- AI-directed (learner as recipient), AI-supported (learner as collaborator), and AI-empowered (learner as leader)- have been proposed by Ouyang and Jiao (2021). In other words, AI-directed paradigm allows the learners to receive personalized learning; to have automated feedback based on their needs; to experience having

interactions with virtual tutors or assistants; and to enhance their understanding via AI-driven instructions. In addition, based on AI-supported paradigm, AI acts as a partner collaborating with the learners to enhance skills like critical thinking and problem solving in a collaborative learning environment. Moreover, the AI-empowered paradigm enables the learners to take control of their own learning in a more self-directed way, to innovate latest ideas, explore possibilities, and pursue their goals.

In addition, Chiu et al., (2023) present a systematic literature review on opportunities, challenges, and future research recommendations of AIEd. By employing matrix coding and text analysis techniques to evaluate the literature over the last 10 years (2012–2021), they seek to identify the prospects and difficulties of AIEd. As a result of this study, the role of AI in education was divided into four main categories: learning, teaching, assessment and administration, which is depicted as follows:



The roles and outcomes of AI in education (Chiu et al., 2023)

2.2.1 AI in Learning and Teaching

Researchers have studied the benefits of employing AI in education from various aspects. Regarding the use of AI in language learning, practitioners have employed it in several ways with controversial results. According to the figure by Chiu et al. (2023), the main functions of AI in learning include: (1) task assignment based on individual competence; (2) human-machine conversations; (3) feedback analysis of student work; and (4) enhancing adaptability and interactivity in digital environments.

One of the emerging kinds of AI is chatbots that are being widely used in the language learning process. Researchers focusing on chatbots have pointed out their several potentials for: learning languages (Ayedoun et al., 2019); creating group activities for developing learning skills (Tegos et al., 2014; Xu et al., 2021); providing necessary interventions (Lin & Chang, 2020); and presenting relevant content (Ruan et al., 2019). It must be noted that, based on Wegerif (2004), using chatbots for language learning supports the constructivist learning theory where the students should take the control of their own learning and develop their knowledge by constructing dialogues and interactions. As previously elaborated, due to difficulties they confront, EFL learners need to take the control of their own learning and lessen the burden on teachers. In EFL contexts, Goda et al. (2014) did experimental research to find out how EFL learners contributed to group discussion following their chatbot conversations. According to the study, the students who had conversations with chatbots were more willing to contribute to the discussions in the group. As a result, learners having conversations with chatbots, gain more confidence and motivation to participate in discussions and interactions. This may be due to one of the common obstacles EFL confront such as fear of making mistakes and feeling embarrassed when speaking in a foreign language which can be overcome in interaction with a chatbot.

In addition, students considered chatbots as a valid resource for language learning that can solve the deficiencies of peer conversations (Thompson et al., 2018). In other research, Kim (2018) conducted an experimental study and investigated the effects of using chatbots on English listening and speaking skills, where he reported that the students' proficiency level increased from intermediate to advanced level through using the chatbot named Elbot. It can be inferred that the proficiency level of the learners, using chatbot in the learning process, needs more research, since the findings are controversial. In a study by Qinghua and Satar (2020), students

with high level of proficiency demonstrated dissatisfaction with chatbots while students with low level of proficiency benefited most. In addition, beside the pros of using chatbots mentioned above, there are some cons like feeling alienation and having artificial interaction without emotions that practitioners and researchers concern about. In a study by Annamali et.al (2023) about the use of chatbots for learning English, the focus was on four main aspects: performance expectation, effort expectation, social isolation, and covid-19 fear. Adapting the Push-Pull Mooring-Habit (PPMH) theoretical framework, the investigation concluded that performance and effort expectations serve as pull factors because they cause the users to have a pleasant experience with using chatbots for language learning. However, the push factors are social isolation, feeling of artificial interaction, emotionlessness, and lack of flow in conversation which are undeniable.

Another common use of AI is implementing AI-assisted software for development of different skills in language classroom. Educators are acquainted with a variety of computer software tools and use them in the process of teaching and learning, such as Ms.word, Grammarly, Wiki, Wordtune, etc. (Kabilan et al., 2010; Kuteeva, 2011; Yunus et al., 2011; Melor & Salehi, 2012; Yunus et al., 2012). For example, Kangasharju (2022) used AI approaches to examine how the digital “Poetry Machine” affects students' poetry writing. The students were invited to try out a variety of poetic devices in the drafts to write their own poems. The data points to a correlation between the final poem's quality and the quantity of altered versions. The findings imply that students who participated in the writing process are inspired and supported by a co-creative AI-based tool, and also that their poems improve from their initial drafts.

In addition, a number of research findings confirm that EFL learners benefit from using certain type of computer software for the enhancement of their learning process (Chappelle, 2004; Naba’h et al., 2009; Barani, 2011; Fageeh, 2011; Razak et al., 2013; Daniels & Leslie, 2013; Qassemzadeh & Soleimani, 2016). Based on the study by Ghufron and Rosyida (2018) about the impact of Grammarly on EFL learners’ writing development, it was revealed that three aspects of writing such as grammar, mechanics, and diction improved significantly, while organization and content of the writings were less affected. Ghufron and Rosyida (2018) found as a result of working with Grammarly software several times, that the learners were able to select not only the correct form of mechanics but also the correct grammar and lexis, which significantly improved their writing skills. This can be due to instant CF that the students

received by Grammarly which results in reduction in errors along with several learning outcomes. In other words, based on the findings of Ghufron and Rosyida (2018), the students who evaluate their works by computer software tools like Grammarly improve their writing skills and reduce their errors more significantly than those who receive teacher's CF.

As previous studies have shown, the integration of technology into learning can aid learners in improving their writing skills through automated corrective feedback (Dikli, 2006; Kukich, 2000; Sireci & Rizavi, 2000; Lee, 2007; Mørch et al., 2005; Winerip, 2012). In fact, based on these studies, when students have the opportunity to evaluate their work through the online learning system, they are more motivated and generally have positive attitude toward the integration of technology into the learning process (Fageeh, 2011). Kabilan et al., (2010) also confirms that online learning motivates students well. Therefore, some studies took advantage of using AI in curriculum to increase motivation of the students along with the potential pedagogical benefits. Annamalai et al (2023) used the Self-Determination study to analyze the motivation of 25 college students to learn English via a chatbot. According to the three psychological demands of learners—autonomy, competence, and relatedness—the information gathered from interviews and found that chatbots promote relatedness, competence, and autonomy. Since in the research the learners subjectively reported that chatbots lack an emotional environment, they suggested a hybrid learning strategy to benefit as much as possible.

AI has been integrated into teaching and instruction as well. Based on Chiu et al. (2023), AI in teaching has three functions such as (1) providing adaptive teaching methodologies, (2) improving teachers' instructional skills, and (3) promoting teachers' professional development. The ultimate goal of intelligent teaching is to provide instructional materials and assignments that best fit the students' need s (Aldeman et al., 2021; Bellod et al., 2021; McCarthy et al., 2016; Weragama & Reye, 2014). In fact, AI-assisted mediums have been adapted in educational contexts specially language classes to enhance learning such as assigning tasks. In addition, based on the research, application of AI in classrooms have made class management more effective for teachers (Gupta & Bhaskar, 2020; Huang et al., 2021; Jarke & Macgilchrist, 2021; Rapanta & Walton, 2016).

In relation to the effect of AI on EFL instructions, Zhai and Wibowo (2023) did a systematic review of AI dialogue systems for enhancing EFL students' interactional competence in the university. This systematic review found six main factors that affect how AI dialogue systems are used for EFL instruction, which includes: task designs, student involvement, learning objectives, technological limitations, technological integration, and novelty effect. In addition, some gaps are found in the AI dialogue system design where (1) elements of debate and problem-solving skills in EFL acquisition in university education appeared to be neglected, and (2) the significance of embedding culture, humor, and empathy functions were not taken into consideration. The results of this study show that the creation and application of an AI dialogue system in EFL are still in their infancy stage. Further research is needed how to practically implement AI in the curriculum to foster self-controlled learning that will lessen the burden on the EFL teachers.

2.2.2 AI in Assessment and Administration

Based on the figure by Chiu et al. (2023), two roles of AI in assessment are automatic marking and prediction of students' performance. In addition, there are three main roles assigned to AI in administration including (1) improving the performance of management platforms, (2) providing convenient and personalized services, and (3) supporting educational decision-making with evidence Chiu et al. (2023).

Automated Essay Scoring (AES) which refers to using technology and computer programs to analyze and score essays, traces back to Page (1967) as the pioneer figure of automatic grading Project Essay Grade (PEG) which predict the scores by using the measurable characteristics of texts, such as average sentence length, the number of words, prepositions, commas, etc. This method was often criticized because of focusing only on surface structure, not considering the semantic features and content (Attali, 2013; Dikli, 2006). In other words, considering only the surface structure such as grammar and spelling, AES was neglecting the richness and complexity of the content leading to inaccurate scoring. Later, more accurate AES tools such as e-raters were created that have the ability to provide analysis regarding sentence structure (syntax), word structure (morphology), and meaning (semantic) (Burstein et al., 2013). Therefore, e-raters were trying to fill the gap and consider a broader range of linguistic and semantic features providing more comprehensive analysis of essays. Having high reliability and validity, e-raters in combination with human raters are being widely used in international

proficiency tests like GRE and TOEFL (Attali & Burstein, 2006). This hybrid technique ensures excellent reliability and validity in evaluating students' proficiency by using the advantages of both human and technology. In addition to scoring, e-raters are equipped with automated written corrective feedback which has a profound impact on the students' writing development (e.g., Cotos, 2014; Koltovskaia, 2020; Ranalli, 2013). This outstanding feature can target different aspects of a text such as grammar, spelling, punctuation, capitalization, choice of word and sentence structure. Through immediate automated written corrective feedback, learners can see their strengths and weaknesses and take action for improving them.

Regarding the use of AI in assessment, Gonzalez-Calatayud´et al., (2021) has done a systematic review which has revealed that AI has two main functions in assessment, including formative evaluation and automatic student grading, which generally assist both the teachers, especially when they have a large number of students, and the learners with providing immediate feedback. They have also mentioned that most of the authors rather focus on analyzing the tools, explaining the platform used as well as the algorithm, but they do not emphasize on the educational rational behind the use of particular activities. They have concluded that the approach to AI is mainly technical, and the educational approaches are secondary. In addition, it is stated that not only do teachers need to be trained in how to apply AI, but also further research is needed to understand the scope of integration of AI into educational assessment. In fact, for successful implementing AI in assessment, it is essential to research more to find out how to take advantage of AI in language classes specially in EFL contexts. In addition, more research is needed about effective training of teachers using AI, examining fairness of assessment, ethic of data privacy and collaboration of data. Likewise, Casal and Kesler (2023) did a study, which considered the problems with accuracy, human judgment, and research ethics. In particular, they looked into three areas: 1) how well linguists and reviewers from prestigious journals can differentiate between writing generated by AI and human writing; 2) the rationale behind the decisions made by reviewers; and 3) how much editors of prestigious Applied Linguistics journals think AI tools are appropriate for use in research. In the study, reviewers evaluated research abstracts produced by AI and humans, and some reviewers took part in follow-up interviews to discuss their decisions. In a similar vein, editors answered questions on their ideas and participated in interviews. Results indicate that even though reviewers used many criteria to evaluate texts, they were mainly unable to detect AI compared

to handwritten text, showing a mere 38.9% overall rate. Furthermore, many editors thought AI tools could be used ethically to speed up research procedures, but others didn't agree.

In addition, Mizumoto and Eguchi (2023) did a study in which the objective was to assess the accuracy and reliability of AES using ChatGPT. They automatically scored each of the 12,100 essays in the ETS Corpus of Non-Native Written English (TOEFL11) using the GPT-3 text-davinci-003 model, and then they compared the scores to benchmark levels. The study also investigated how much linguistic characteristics affect AES, using GPT. The outcomes demonstrated that AES with GPT has a certain degree of accuracy and dependability and may be a useful supplement to human evaluations. Additionally, the investigation showed that the accuracy of the scoring might be improved by using linguistic features. These results imply that AI language models, like ChatGPT, can be used as powerful AES instruments, potentially changing the way that evaluation and feedback are written in both practice and research.

In another study, Li (2023) did a review on an AI-based online writing assessment tool called "Peerceptiv" which encourages students to construct, exchange, and improve their writing by giving and receiving peer comments. To elaborate more, after students upload their work to Peerceptive, receive written feedback and grade from their peers first, followed by further evaluation from Peerceptive that shows how accurately the review by peer is. Studies have shown that this process is particularly helpful for students' writing and feedback literacy improvement. This method seems appropriate for writing instructors, and it can be integrated with AI-driven automated feedback systems in their virtual writing programs, fostering a technology- and peer-mediated interactive learning environment where students can review and edit a range of writing issues, thereby enhancing self-directed learning. More elaboration about feedback on writing is provided in the following section.

2.3 Feedback

One of the main factors in enhancing learning specially in EFL is receiving feedback. The term feedback is defined as any comments on test or learning task, either from a teacher or peers, which is concerned about the learners' success (Richards & Schmidt, 2010). Generally, feedback is an important communication skill both inside and outside of the classroom (Dignen, 2014). In fact, feedback is a crucial part of both learning and teaching which significantly improves the performance of teachers as well as students. In the process of learning a foreign

language, making errors is inevitable and teachers are responsible to correct them by giving useful feedback. Based on Ur (2006, 242), feedback is "information that is given to the learner about his or her performance of a learning task, usually with the objective of improving this performance". He mentions that two elements of feedback are correction and assessment which are done with the aim of improvement. He further warns about being judgmental in giving feedback which is inevitable and trying to be nonjudgmental is unrealistic and impossible.

As mentioned by the researchers like Black and Wiliam (2009), there are several key strategies in feedback: defining the aims of learning and setting criteria for success; creating engaging classroom activities and discussions that elicits students' understanding; providing feedback that helps students make progress; activating students to take control of their own learning; and empowering them to act as structural resources for one another. In addition, there are four levels of feedback defined by Hattie and Timperley (2007): feedback about the task itself, feedback about the processing of it, feedback about the self-regulation, and feedback about the person itself like praising which is the least effective.

In EFL setting, as mentioned before, due to lack of exposure outside of the class, teacher's role is dominant in the class; although, students have to take the control of their own learning as well. Writing as one of the main skills has several challenges in EFL context and needs a great attention. It must be emphasized that the EFL learners face more significant writing challenges than in their first language (Lin & Morrison, 2021; Hanauer et al., 2019). Based on Murray (1972), the writing process consists of three steps: prewriting, writing, and rewriting which have been considered a complex process where feedback, as an active intervention, facilitates learning in the process of writing (Graves, 1982; Thompson, 2013).

Feedback in writing may include several aspects such as grammar, lexis, organization, content, and mechanics of a text, which will be significantly effective in the development of EFL learners' writing skills. According to Ferreira (2006) the main goal of feedback in writing is to provide useful information so that the writers can modify their mistakes. However, as Ur (2006) confirms, teachers find correction of mistakes, especially the recurring ones, a waste of time and energy, and they mostly feel correction of such errors are not fruitful, instead they need to make opportunities for students to get things right as much as possible. The same happens in

EFL contexts, especially in large classes where it is almost impossible to give efficient feedback to all individuals for every piece of writing task.

2.3.1 Written Corrective Feedback (WCF)

According to Ellis (2008) error correction contributes to success in second language acquisition. In fact, error correction refers to identifying and addressing the error made by the learners in their language productions. Through error correction and receiving feedback, the learners increase their awareness about the language and work towards improving their proficiency. In other words, corrective feedback (CF) may contain responses about students' sentences including the occurrence of the error, the correct target language form, metalinguistic analysis of it, or a combination of all of these (Soori et al., 2011).

Generally, based on Ellis (2008) there are six types of (CF) strategies: direct, indirect, metalinguistic, focused/ unfocused, electronic, and reformulation. Direct feedback refers to indicating the errors explicitly and providing the correct form. For example, a teacher may underline a spelling mistake of a word and write the correct form of it for the student. While indirect feedback involves encouraging the learners to identify and correct the errors themselves without providing the right form. Indirect feedback can occur by asking questions or giving suggestions. The third type, metalinguistic CF engages explanations about the nature of language and linguistic norms toward how and why it is called error. Such type of CF contributes to awareness of the underlying grammar and linguistic rules. The fourth type, focused CF refers to a specific type of error or feature, while unfocused CF involves the errors broadly without addressing specific feature. Another type of CF is electronic which involves using technology such as computers and online platforms to provide feedback for the learners' language productions. Finally, reformulation CF helps the learners see the right form in the context by restating the erroneous utterance in the correct form. In line with the objectives of the present thesis, the direct and electronic CF in written form is elaborated.

Written Corrective Feedback (WCF) has been studied by many researchers and it is confirmed that WCF leads to the development of students' writing skill by letting them observe their errors and helping them to correct such errors (e.g., Karim & Nassaji, 2020; Zhai & Ma, 2021; Zhang & Zhang, 2018; Xu & Zhang, 2021). On the other hand, there are several barriers to WCF in that teachers may find it demanding and time consuming to do so for all students. In addition,

most of the errors in writing specially those regarding the mechanics of a text- punctuation, capitalization, and spelling- are recurring and teachers mostly prefer not to waste time on correcting them. Currently, with the advent of technology, teachers are getting increasingly aware of electronic CF which has come to supplement teachers' CF in mostly beneficial ways.

2.3.2 Automatic Written Corrective Feedback

Considering the goals of language learning and the context, teachers can take advantage of providing CF through technology, using computers, specific software, and applications. Computer-based WCF or computer-mediated corrective feedback (CMCF) has come to EFL teachers' attention in language classes. Accordingly, Engeness and Morch (2016) did a comparison study about developing writing skills in English using content-specific CGF with *EssayCritic* in Norwegian upper-secondary school wherein the target class was receiving feedback from an essay critiquing system while the comparison class was receiving feedback from peers. The result demonstrated that although both classes improved dramatically; in the target class, the students implemented more ideas in their essays than the students in the comparative class who had trouble providing each other comments because the feedback from EssayCritic provided content-specific hints. The results highlight the function of EC in continuously aiding students during the writing process by simultaneously delivering individualized feedback to every learner on many occasions and drafts. Additionally, EC helped the students to analyze the writings and identify the subthemes that were and were not addressed in the text, which was difficult for the students in the comparison class to do. Similarly, in another study about different mediums for feedback in writing by Lee et al. (2013), students in one group received feedback from both CGF and teacher, while the other group received feedback from only the teacher. The findings revealed that the students receiving two types of feedback outperformed the second group and could write essays with richer content.

Some writing assistant tools help students during writing process such as "Wordtune" by generating context-based sentences and offering various suggestions, while some others offer feedback after the writing process like "InstaText and QuillBot" by providing accurate suggestions and correcting the errors. In fact, CMCF can not only contribute to the learning process, but can also facilitate the process of feedback and save time so that the teacher can focus on more prominent issues in language classes. Providing immediate, personalized and automated feedback, CMCF gives the opportunity to the learners to see where they have

deviated from the linguistic norms, including grammar inaccuracies, spelling, punctuation and capitalization errors.

Alharbi (2023) categorizes AI-based writing assistant tools into four groups: 1) Automated writing evaluation tools (AWE), 2) automated writing corrective feedback providers (AWCF), 3) AI-based machine translators (AMT), 4) GPT-3 automatic text generators (GPT-3). Some widely used AWCF tools like Grammarly, Ginger, and ProWritingAid are able to provide feedback simultaneously and continuously while the text is generated by the writer (Dale & Viethen, 2021).

"The biggest change in writing since the invention of the word processor" has occurred with the appearance of AI-based text generators (Floridi & Chiriatti, 2020, p.69). AI systems are increasingly getting capable of generating texts based on the prompt given to them using text prediction technology. According to Kasirzadeh and Gabriel (2022), large language models, which are AI systems relying on vast datasets, trained on next word prediction, have become excellent at generating grammatical, well-formulated text in English. However, it must be considered that the popularity of such tools may increase the risk of abuse, whereby they can be used to generate content directly instead of the students producing it themselves. Since by the rise of such tools it is inevitable that learners mostly tend to use them, so it is necessary to direct them in a right way. The main challenge is to find a way of using the functionality of these tools to improve writing skills rather than to do it for them, and also it is important to avoid plagiarism and cheating by implementing such tools as a kind of language learning material and as an advantage toward achieving ultimate learning goals.

The positive features of AWCF tools such as Wordtune and Grammarly include: it delivers feedback fast and it is available in two versions free and paid (Nova, 2018; Huang, et, al. 2020); it improves the quality of writing (Ghufron, 2019); it increases lexical diversity (Dizon & Gayed, 2021); it categorizes errors based on the nature and provides personalized feedback (O'Neill & Russell, 2020); it prevents plagiarism by identifying textual borrowing (Dodigovic &Tovmasyan, 2021); develops self-corrections and self-regulations (O'Neill & Russell, 2020). In addition, based on Dale and Viethen (2021), automatically completing sentences and phrases as well as suggesting alternative wording features are among the greatest developments that AI has brought to the writing domain. In addition, such AI-powered tools are built on massive

linguistic models which provide a package with full range of language assistance services from MT to sentence generation (Dale & Viethen, 2021).

EFL learners obviously need more tools to be integrated in education in order to accelerate the process of learning, especially when it comes to writing skills. Consequently, due to the challenges mentioned above, applying new strategies, use of technological advancements and integration of AI into education can come to effectively aid EFL learners in the process of learning language and particularly in writing skill development.

However, it has been controversial how AWCF should be implemented in the studies, since some researchers offer it to the learners with lower proficiency level (Nova, 2018); while others recommend it to be implemented for English learners with advanced level of proficiency (Koltovskaia, 2020). It seems that, the impact of such tools on writing process still needs to be further analyzed to unlock their full potentials for different tasks and different levels of student ability so as to implement them appropriately in the writing classrooms (Grimes & Warschauer, 2010; Huang & Wilson, 2021; Vinall & Hellmich, 2022). Furthermore, only a few studies have so far considered the impact of AI writing assistants on different aspects of writing- task fulfillment, organization, lexis, grammar, and mechanics- where the suggestion so far is that organization, argumentation strength, and coherence are among the areas that AWCF may fail to help (Al Mahmud, 2023; Rad et al., 2023). In addition, no studies to date have examined the correlation of time spent working with the applications and the outcome of writing development. Since improvement in writing is likely to be correlated with the amount of time spent working on texts, the time spent on using writing tools need to be factored into any experimental study that seeks to show the efficacy of the tools themselves in improving students' writing skills. Assessing the impact of AI writing assistants on different aspects of writing is one of the aims of the present study, and the time the student spend on training and task completion will be explicitly measured during the course of the experimental in order to ensure that this factor is not a confound factor for our conclusions.

Overall, as mentioned by Chiu et al., (2023), the scope of AIEd is too scattered and most of the research is done about engineering aspects like creating new algorithms and developing machine learning techniques, so that the impact of AI on education has remained unclear (Holmes et al., 2021). Obviously, more research is required to clarify whether and how these

emerging technologies can benefit education in general and language learning in EFL contexts in specific. Additionally, more research is needed about the different available AI-based tools, how they work, how it is possible to integrate them in language learning contexts, and which aspect of language skills is most likely to be affected by them. Based on Godwin-Jones (2021) it is necessary for the teachers to design tasks that implement automatic text generators along with students' effort. In other words, students should not merely become dependent on such tools and get distracted from the main objectives of learning. In fact, the use of AI-based writing tools should be integrated within a comprehensive language curriculum with the value of communication at the core (Link et, al. 2020).

Therefore, addressing the current gaps in AIEd, the purpose of the present study is to investigate the effect of two AI-based writing tools, "Wordtune" and "Insta text" on EFL learners' writing development. Specifically, the study focuses on evaluating the impact of these tools on students' writing in terms of five writing aspects including task fulfillment, organization, grammar, lexis, and mechanics, which are elaborated more in chapter three. It must be emphasized that the study seeks to implement these tools in writing home assignments, not in any of the assessments or in class test, in order not to get distracted from the classroom communication goals. Although AI-based writing tools can be utilized to improve learners' writing skills, they should play a supporting role in the writing process not a leading role (Huang & Wilson, 2021). It must be noted, time is allocated in the classroom to exchange experiences and discuss the challenges with peers and the teacher. Plus, the study involves questionnaires for both the teacher and the students to express their attitude toward AIEd in language classroom. We have implemented this as part of the study since it is important to know not only the students' but also the teacher's attitude toward technology tools, since these attitudes have conclusive impact on the effectiveness of such tools (Jiang et, al. 2020).

The following chapter is devoted to the research questions, predictions, and methodology in line with the objectives of the present study.

2.4 Chapter Summary

This chapter discusses the background studies related to: The integration of technology into the language learning process; use of AI in education in domains of learning, teaching, assessment,

and administration; and the role of feedback and its diverse types in learning specially in the EFL learning environment.

Accordingly, it is elaborated that EFL learners, facing a lot of challenges when learning a foreign language, need a learning environment different from ESL with increased need for control of their own learning. Therefore, the role of feedback is significantly highlighted in facilitating the language learning process. As a result of integration of technology into education, learners can now benefit from computer generated instant feedback provided by AI-based writing tools to observe the errors as they occur, to increase their awareness, and to take action toward correcting them. However, effective integration of technology requires further research, efficient teacher training, curriculum adaptations, and training about ethics of data privacy.

As discussed, integration of AI into education in learning, teaching, assessment, and administration is increasing rapidly, encouraging personalized self-directed learning. However, challenges remain with the need for further research especially in an EFL context to understand how AI impacts learning languages and which methods are more effective. Furthermore, concerning EFL learners' writing skills, it is needed to know how AI-based writing applications work to facilitate writing skills development, and also which aspects of writing are most influenced. The following chapter will be devoted to the specific research questions, hypotheses and predictions, along with information about methodology.

3 Methodology

The aim of the present thesis is to investigate the effect of using AI-based writing applications such as "*Wordtune"* and "*Insta text"*, on EFL learners' writing development. In addition, the thesis investigates which of the five key aspects of writing_ task fulfillment, organization, mechanics, lexis and grammar_ are most influenced by use of these applications. Another aim is to examine whether using AI-based writing applications contributes to time efficiency in the process of writing. Finally, the attitudes of the teacher and the students toward AIEd particularly in the process of language learning is assessed. Therefore, in this section, the following research questions, hypotheses and predictions are introduced along with information about the procedure, the materials, the participants and data analysis.

3.1 Research Questions

Regarding the aims of the present study which is to investigate the influence of using AI-based applications on EFL learners' writing skills development, the following research questions are raised:

1. What is the impact of using AI-based writing applications on the EFL learners' writing skills?
2. Which aspect of writing among task fulfillment, organization, grammar, lexis, and mechanics are most affected by the use of AI-based writing tools?
3. Does using AI-based writing applications contribute to time efficiency in the writing process?
4. What are the learners' and the teacher's attitudes toward AIEd specially in the process of language learning?

3.2 Research Hypotheses and Predictions

According to the nature of EFL learning and numerous challenges that EFL learners experience, which was discussed in chapter 2, taking control of their own learning as well as having more exposure to English are two main goals which should lead to better writing skills. Since AI tools facilitate both these factors, it is hypothesized that AI-based applications will have a positive effect on EFL learners' overall language proficiency, especially when it comes to writing skill. As discussed in chapter 2, AI-based writing tools provide AWCF which is beneficial for the learners from several aspects, but how about "Wordtune" and "Insta Text",

which were implemented in the study, whether they have the same effects as Grammarly, and also whether the learners genuinely become better at doing the writing themselves benefiting from AWCF. Thus, the following predictions are listed:

1. Based on the model for AIEd proposed by Chiu et al., (2023) and the previous research findings, although it is predicted that using AI-assisted writing tools improves overall writing skills, it is not yet clear how it works so, especially in an EFL context.
2. Based on Ellis (2008) strategy of direct CF, and also based on CGF and AWCF, it is predicted that as a result of observing errors corrected automatically, grammar, lexis and mechanics of writing are affected and improved more than other aspects.
3. Using computer generated sentences and benefiting from AWCF, it is predicted that the students in the experimental group spend less time on generating texts.
4. Based on the previous research, it is predicted that students experience a new learning environment through which they have more confidence and motivation toward learning a foreign language. In addition, based on the previous research, EFL teachers suffer from a great burden of giving feedback and error corrections. If AI assisted feedback is even equally efficient as human feedback, this will be a clear advantage to teachers.

The first prediction is built upon the challenges discussed in chapter 2, which EFL learners confront when learning English. One of the most highlighted challenges in previous research is lack of exposure for which AI-based writing applications are providing authentic sentence suggestions based on the context (Muñoz, 2014; Akbari, 2015). In addition, according to findings of Kukich (2000), Sireci and Rizavi (2000), Mørch et al. (2005), Dikli (2006), Lee (2007), and Winerip (2012), due to providing automated error correction, such applications can be helpful in writing development. Moreover, based on the findings by Ranalli (2013), Cotos (2014), and Koltovskaia (2020), immediate automated written corrective feedback helps students see their strengths and weaknesses, and encourages them to take action.

Prediction 2 is based on Elis' (2008) corrective feedback which will result in more successful second language acquisition. Direct corrective feedback is one of the possible strategies that indicates the error explicitly and provides the correct form. According to the research done by Zhang and Zhang (2018), Karim and Nassaji (2020), Zhai and Ma (2021), Xu and Zhang (2021), AWCF improves writing skills through giving the learners opportunities to observe

their errors and how to correct them. Since errors regarding the mechanics of writing-capitalization, spelling, and punctuation- more frequently occur, it is predicted that they improve mechanical errors first. In addition, based on the findings by Ghufron and Rosyida (2018), Al Mahmud (2023), and Rad et al., (2023), beside the mechanics, lexis and grammar improve due to AWCF.

Regarding prediction 3, Since the condition for both groups in terms of doing home assignments are not the same, and using a writing application can be a facilitator of the writing process, it is predicted that the students in experimental group do the home assignments faster and spend less time on the process of creating a text.

Prediction 4 and 5 are based on the findings by Mørch et al. (2005) that using computer generated feedback can supplement teacher's feedback. In other words, the burden on the teacher is lessened and the students are benefiting from the AWCF and are getting more self-directed in the process of learning. In addition, based on the findings of Kurniawan and Radia (2017), the problems such as lack of teachers' time and availability are solved by implementing AI in the curriculum. Furthermore, implementation of AI in the curriculum increases the learners' motivation which can result in more active participation in the classroom that all the teachers wish for (Annamalai, et al. 2023). In sum, outcomes 4 and 5 concern the teacher's and the students' attitude toward using AI in writing practices and generally in the process of learning a foreign language.

3.3 Research Design

The present study is an experimental approach, using between-subjects design, to compare two groups of EFL learners in terms of their writing development during the study. The experimental group is exposed to the potential benefits of utilizing AI-based writing applications for doing home assignments during the experimental period (or 'training' phase). The control group, on the other hand, does not have access to these tools and follows the traditional pen-and-paper writing techniques. The two groups are tested both before and after the experimental training phase on a range of different language criteria. This controlled research design allows for comparison between these two groups in order to evaluate the effect of AI-based writing tools, such as "Wordtune" and "Insta text", on the students' writing development. In this regard, the evaluation of writing development in general, as well as

evaluation of five key aspects of writing are dependent variables, while the use of AI-based writing tools in the training phase or not are the independent variables.

It must be mentioned that the participants in this study passed the language proficiency test in advance and were qualified to enter the early intermediate level in EFL classes. Both groups took part in the writing pretest which was implemented with the same conditions for all the participants, without using AI-based writing applications, to evaluate their writing proficiency prior to the study. Over the study (the experimental, or 'training' phase), which lasts for almost three months, the students deliver written home assignments every session and discuss their challenges in the process of writing in the classroom. To evaluate their progress in writing, there are six periodic writing tests plus a final posttest. Data is collected in the form of test scores on their writing which is assessed based on the rubrics, and these scores are analyzed statistically to evaluate whether/how the students' performance differ between the two groups.

3.4 Procedure

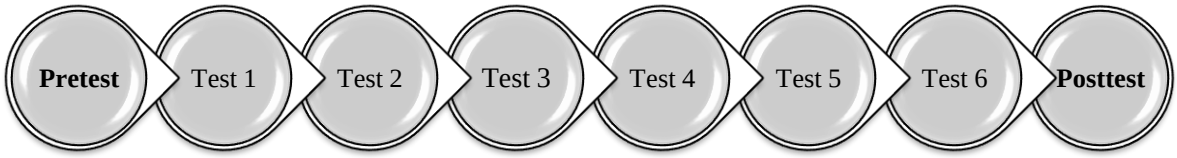
Since the present study is aimed to the evaluate the impact of using AI-based application on writing skill development for EFL learners, the first step was holding a general English proficiency test to ensure the homogeneity of the participants. It must be noted that prior to the study, the eligible participants signed the consent form confirmed by SIKT organization in Norway, in line with the ethic of data collection and privacy.

The study started with an essay writing pretest which aimed to assess the students' baseline writing skills. The pretest was done based on standard essay writing with a topic selected from the previous student textbook. Next, the participants were divided in to two groups, one as an experimental group (n=16), and another as a control group (n=18). The participants attended the course in the English Language Academy in Iran in two different branches and were not in contact with each other. The study lasted for almost three months, and the classes were held for three sessions per week, from early November 2023 until late January 2024, for 36 sessions in total.

Throughout the study, the students in both groups were assigned to write one essay as a home assignment per session, in which the experimental students were supposed to use "Wordtune" and "Insta text", two AI-based writing applications that are explained more in section 3.7. It must be noted that the experimental group had received enough instructions about these tools

and had practiced in the classroom in advance. However, the control students were not introduced to these tools or the possibility of using them and wrote all the assignments through the pen-and-paper technique without using the applications. In addition, both groups were asked to note the amount of time spent on each assignment. The assignments were delivered and the teachers in both classes were responsible for giving feedback on them and the students could discuss their challenges in writing in the classroom as well.

During the study, there were also six periodic essay writing tests, one test at the end of each unit of the assigned coursebook, held in the class for both groups under the same condition, and without using the applications. Finally, at the end of the study, both groups had an essay writing posttest in the class, once again without using the applications. The data gathered as a result of the writing scores of the pretest, the six periodic tests, and the posttest was then used to evaluate and to compare the students' progress in both groups. The following flowchart gives an overview of the procedure of the study:



3.5 Pilot Study

In order to test how the process of practicing writing through AI-based writing application goes on, prior to the main experiment, a pilot study was conducted. Accordingly, seven students of the same level, who were taking private English classes, took part in the pilot study. In fact, the pilot study was carried out to test how the applications work for them, and what type of instructions they need in order to benefit fully. In addition, it was needed to ensure that they understand how to work with the applications to be on the right track. Moreover, we wanted to to predict what types of challenges might arise so that they could be prevented for the main experiment.

During the pilot study, the participants were given a topic to write an essay using "Wordtune" to generate sentences related to the topic and to use "Insta text" for a final evaluation of the

text. It was revealed that they needed detailed instructions about how to use the applications properly, how to set the context to be understandable and to lead the flow of the text, and also how to choose among the suggested sentences. One of the surprising results of the pilot study was that the students could observe how their grammatical, spelling and capitalization errors were corrected automatically which made them more engaged in the test. In addition, sentence suggestions based on the topic and context were amazing for them, since they complained of always having a lack of ideas. They expressed that it was a very new experience for them, and that it would be interesting to continue using the applications even outside of the curriculum. In addition, they were motivated enough to recommend other members of their peer group to use the applications as well. As a result, there was no need to apply any significant changes to the plan for our study, but the instructions were made longer and clearer. It must be noted that, due to the pilot study, the topics assigned for the essays were slightly modified in order to appropriately set the context in the applications, so that AI could suggest more context-based sentences.

3.6 Participants

Participants in this study are the students learning English as a foreign language at the early intermediate level (A2-B1 based on CEFR), in Safir English language academy in Iran. All the participants passed the final exam of pre-intermediate level and gained the target score which is more than 70 out of 100 and have permission to take part in the intermediate course. In fact, there is a proficiency test after each level that students have to pass in order to be eligible to participate in the next level. Therefore, the participants are homogeneous in terms of proficiency level.

Next, the participants were divided in to two groups based on conditions such as access to a laptop, flexible time, etc. In this way, 18 participants attended the classes in one branch as the control group and 16 students participated in another branch as the experimental group. The age range of the students was about 18-23 years old, and all were female. In addition, all the participants signed the consent letter to take part in this study and agreed to share their home assignments as well as the tests' results. For the purpose of ethical consideration and data privacy, each student was assigned a number which was kept confidential. Therefore, the essays were anonymous for the scorers to eliminate the biases, to increase fairness, and to raise reliability and validity.

It must be mentioned that the participants were allowed to withdraw at any time in this study and as they were informed in the consent form, there was no penalty for withdrawal. In addition, they were aware that this study was for their own sake to try innovative approaches toward improving the process of learning English.

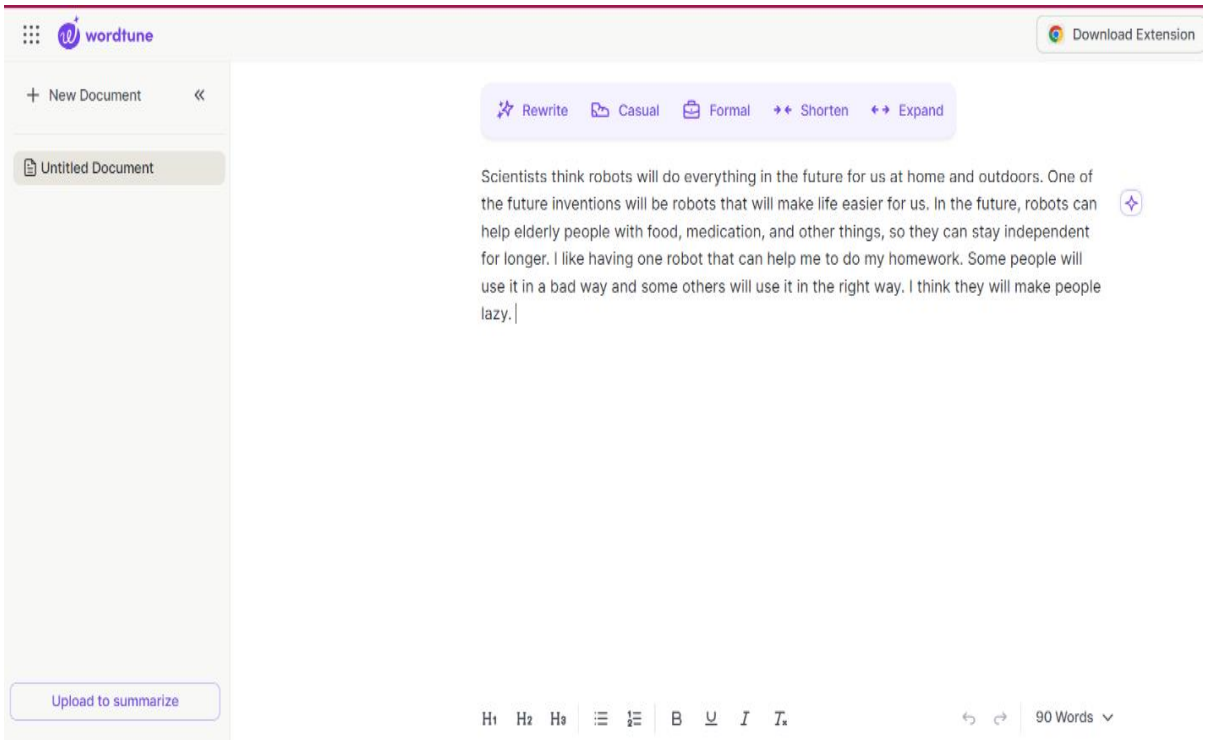
3.7 Materials

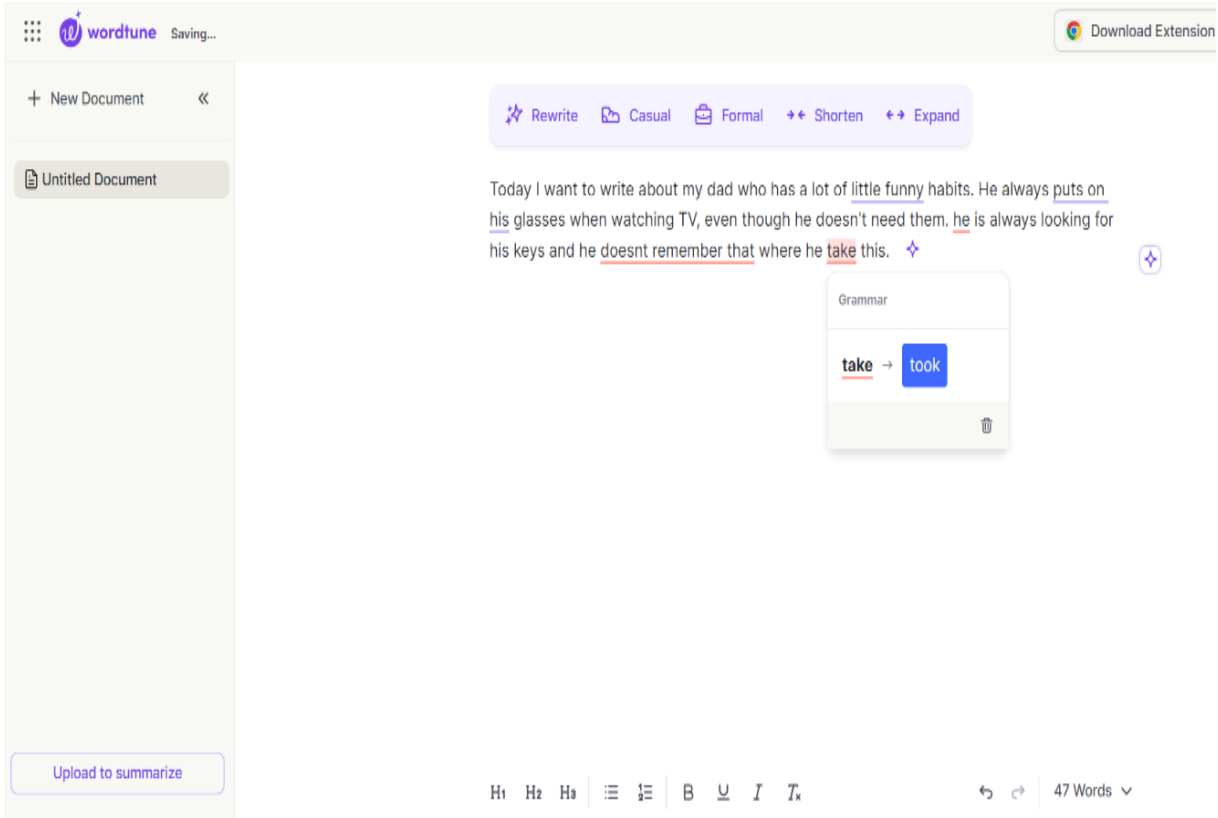
In this study, the main material used in both classes, for the experimental and control groups, was the English textbook 'Touchstone 3', for intermediate level, by Michael McCarthy, Jeanne McCarten, and Helen Sandiford (2014). In fact, during the study six units of the book were taught in both classes over 36 sessions. As the book is structured, each unit is devoted to a specific topic and there are listening, reading, speaking, and writing exercises as well.

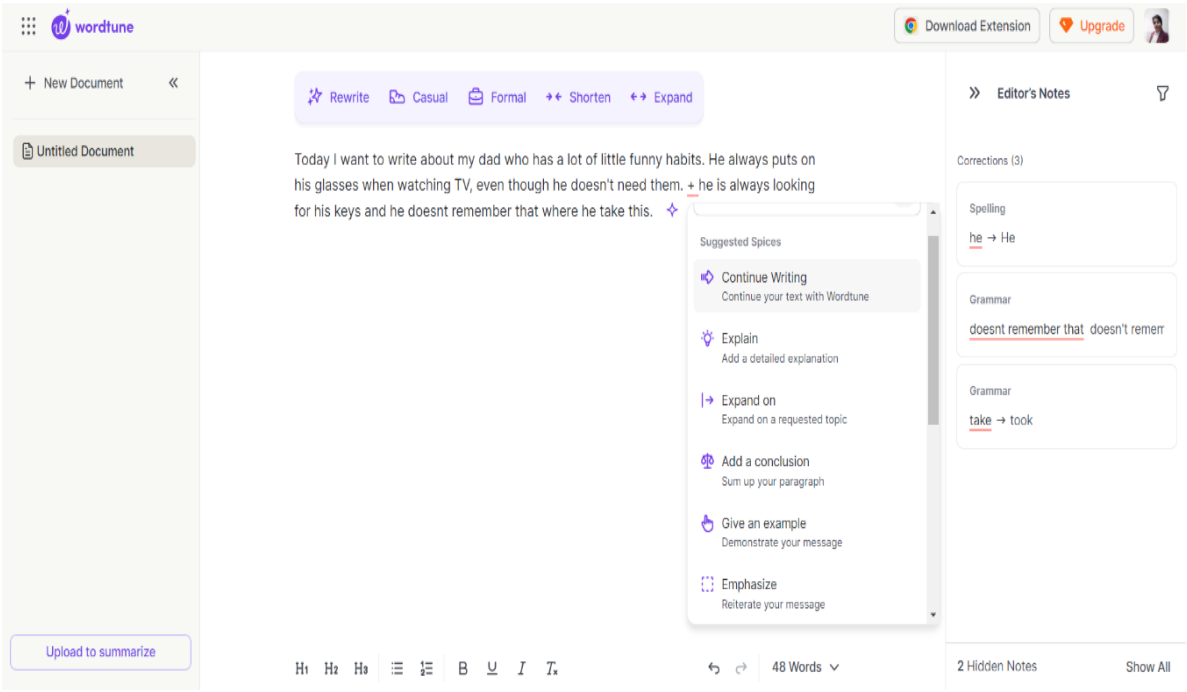
In addition, during the study, for both groups, there was a home assignment every session to write an essay about a specified topic related to the lesson taught, about 140-190 words. The students in the control group followed the traditional pen-and-paper writing technique and the teacher was responsible for correcting the assignments and giving them feedback. However, the students in the experimental group were supposed to do home assignments using two AI-based writing application, "Wordtune" and "Insta text", which are free and available to easily sign in with an e-mail address. As mentioned in chapter 2, one of the main AI-based writing assistants is "Wordtune" which has several specific functionalities such as: 1) providing error corrections; 2) providing rewriting and paraphrasing features; 3) the ability to shorten and expand text; and 4) the ability to provide several context-specific sentence suggestions.

Similarly, "Insta text" has several features and provides improvements of a text in terms of error corrections, sentence suggestions, tone, and dialect adjustment. In fact, they were asked to generate the text by "Wordtune" as an AI-based text generator and polish the final draft with "Insta text" as an automatic text evaluator. The experimental learners were free to use AI generated sentences and correct the text several times before delivering the home assignments and the teacher was responsible for checking them.

To be more specific about the AI-based applications, how they appear, and how they work in terms of error correction and sentence suggestions, the following pictures depict the design of "Wordtune":

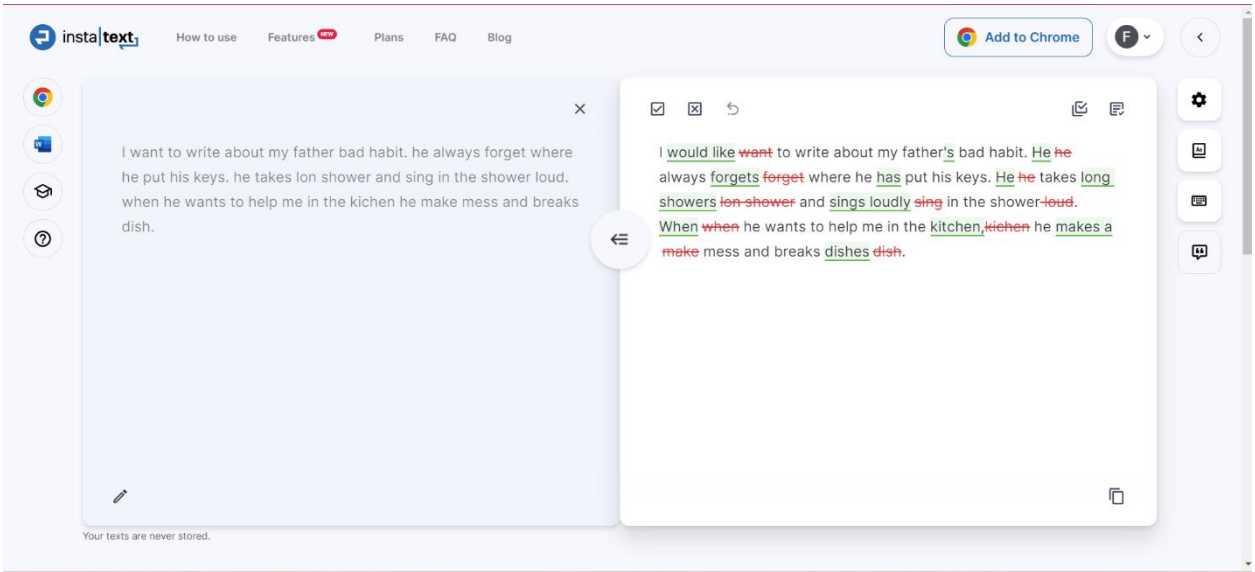

wordtune
Download Extension
New Document
Untitled Document
Rewrite
Casual
Formal
Shorten
Expand
Scientists think robots will do everything in the future for us at home and outdoors. One of the future inventions will be robots that will make life easier for us. In the future, robots can help elderly people with food, medication, and other things, so they can stay independent for longer. I like having one robot that can help me to do my homework. Some people will use it in a bad way and some others will use it in the right way. I think they will make people lazy.
Upload to summarize
90 Words


wordtune
Saving...
Download Extension
New Document
Untitled Document
Rewrite
Casual
Formal
Shorten
Expand
Today I want to write about my dad who has a lot of little funny habits. He always puts on his glasses when watching TV, even though he doesn't need them. he is always looking for his keys and he doesnt remember that where he take this.
Grammar
take → took
Upload to summarize
47 Words



In Wordtune after logging in, you can start writing by setting the context, and the application then predicts the next sentences and provides alternative suggestions. There are also other options such as expanding, rewriting, and shortening. Even you can paste your ready text here and the application detects the errors, provides the corrections, and helps to continue the text. Wordtune helps during the process of writing, while Insta text helps after writing.

In Insta text, you should paste your text in the left page and then the application detects the errors and provides the corrections in the right side. The following picture shows the design of "Insta text" and how it works in terms of error corrections and word suggestion:



The other materials to mention here are some essay writing tests that the participants took part in. First, for the sake of ensuring student homogeneity in writing proficiency level, before the beginning of the study, an essay writing pretest was held, in which the topic was adapted from the previous term's material 'Touchstone 2'. The students were asked to write a short essay, about 140-190 words in 40 minutes, about "a future invention". In addition, there were 6 periodic essay writing tests, at the end of each unit, related to the content of the units taught in both classes. Finally, at the end of the study, there was an essay writing posttest under the same conditions, 140-190 words in 40 minutes. All of the essays were scored by a group of teachers who were trained scorers and educated in English Teaching. It must be mentioned that the scorers were independent and did not have information about the groups that the essays belonged too. Moreover, the papers were anonymous to assure validity and reliability. The essays were scored based on the designed rubric adapted from IELTS task 2 essay writing, ranging from 0 to 4 points for each aspect of writing, giving a total of 20. The rubric consists of five aspects of writing: task fulfillment, organization, grammar, lexis, and mechanics. The definition of each aspect of writing is as follows:

- Task fulfillment: refers to the extent to which the students address the requirements of a writing task including topic related content, format of the text, as well as the length of the writing. In other words, it is related to how much the objective of the task is achieved.

- Organization: refers to the overall structure of the text, arrangement of the ideas, use of cohesive devices to connect the ideas, and coherence of paragraphs. In other words, it is related to how the sentences and ideas are arranged.
- Grammar: evaluates how much the sentences are structured grammatically correctly in terms of the tense, subject-verb agreement, place of nouns, adjectives, and adverbs.
- Lexis: refers to the range of vocabulary used by the students based on the material taught and their level of proficiency, choice of words, variety of phrases, and expressions. This aspect assesses to what extent the text is rich in terms of vocabulary.
- Mechanics: includes technical aspects of writing, such as punctuation, spelling, and capitalization. This aspect examines to what extent the learners have written with accurate mechanics.

To be more specific, table 1 shows the scope of analytic rubric along with the scores assigned for each aspect of writing. As it can be seen, there are five key aspects of writing, each of which is ranged from 0 to 4:

Table 1. *Analytic Rubrics for Scoring the Writings*

	Key aspects of writing	*0*	*1*	*2*	*3*	*4*
1	Task Fulfillment (topic, format, length)					
2	Organization (coherence, cohesion)					
3	Grammar					
4	Lexis					
5	Mechanics (errors of capitalization, spelling, and punctuation)					

To elaborate more on the scoring based on the rubric mentioned above, a short excerpt from the pretest (about a future invention), and an excerpt from the posttest of one of the participants in the experimental group (about time management) is presented.

1- Pretest writing:

« one of the invention of the future will be space elevaitor which make our life better. It will make dreams become true. It will take people to space for vacation to others planets for visit the life outside of world. It will be great feeling and interesting. I think there will be no polloutions and any accidents with the travel by space elevaitor. Also the sientists will build restaurant and caffe in moon and mars for people enjoy…»

2- Posttest writing:

« I am a multitasker but I can manage my time successfully. Multitasker means when you can handle many tasks at the same time. I always try to plan and focus on the most important tasks. First, I set deadlines and try to keep track of my progress. I've always tried to stick to my plan and set reminders for myself. Next, I take regular breaks to stay productive, motivated and energetic. I also reward myself for example I buy gifts or go to the restaurant with my friends... »

Obviously, the excerpt from the pretest has a number of spelling errors as well as grammatical issues such as tense inconsistency, errors in subject-verb agreements, and word order inadequacy. In addition, there are some issues with lexical accuracy, word choice, and variety of vocabulary which have resulted in lack of clear text. Moreover, in terms of organization, the text lacks a clear structure, flow, coherence, and paragraphing. In contrast, the excerpt from the posttest contains various appropriate vocabulary, complex sentences, error free mechanics and grammar. These issues regarding the five aspects of the essays based on the participants' proficiency level were considered in scoring. Therefore, the first example got lower scores compared to the second one.

Finally, another piece of material which was used at the end of the study was a survey about integration of technology into AI in language learning. In this regard, the students in the experimental group were asked to fill out a questionnaire regarding their attitude toward the assigned writing application; how AI was helpful in improving their writing skills; what challenges they faced using applications; what features they preferred; and whether they will

recommend using the applications to other peers or not. Similarly, there was another questionnaire for the teacher's attitude toward AIEd in language learning.

3.8 Data Analysis

Data collected throughout the present study is categorized in to three groups. First, there is some data related to the scores of the pretest, tests 1-6, and the posttest of essay writing assessed based on the rubrics mentioned above in table1. Second, some data was collected regarding the time spent by the students, in both control and experimental group, for doing the home assignments. Finally, the descriptive data was collected as the result of questionnaires regarding the experimental students' attitude toward using technological advancements such as AI-based tools for the purpose of language learning. In addition, there was another questionnaire for the teacher's attitude toward integration of technology into education and implementation of writing applications in the lesson plans.

In order to analyze the data related to the scores of the writings, a linear mixed-effects model is run to compare how the students in the experimental group have improved in comparison to those in the control group in terms of overall effect on writing skills. In addition, a series of statical analysis are applied and graphs are produced to depict how using AI-based application has affected different individual aspect of writings including task fulfillment, organization, grammar, lexis, and mechanics over the time in two groups. For the purpose of gradual improvement visualization line plots are produced. Moreover, to compare the amount of time spent by the students of the two groups, for the home assignment, two sets of time recorded for 36 sessions were compared. These results are demonstrated through graphs as well. Finally, the descriptive data referring to the teacher's and the students' attitude, collected by the questionnaires, is demonstrated in tables and charts referring to the AI-based applications, including the following factors:

1) AI-based application usage
2) Specific application usage
3) Frequency of usage
4) Impact on English skills
5) Specific ways of improvement
6) Liked features

7) Challenges faced
8) Suggestions for improvement
9) Satisfaction level
10) Recommendation

3.9 Chapter Summary

In this chapter, the objectives of the study were presented, which is primarily to investigate the impact of technology, such as AI-based applications, on EFL learners' writing skills. In addition, the question of which aspect of writing is mostly influenced is raised, and also what the students' and the teacher's attitude toward the use of AI in this context are. Moreover, the factor of time for doing assignments is measured to compare between the groups. Accordingly, there are some research questions and predictions which are made based on previous studies.

In addition, the chapter describes in detail the procedure of conducting the study_ a pretest, 6 periodic tests, a posttest_ along with information about the participants who are divided into two groups, the experimental (n=16) and the control (n=18). The chapter presents the five key aspects considered when scoring the essays: task fulfillment, organization, grammar, lexis, and mechanics. This leads to a writing rubric for assessing the various writing tests, where each of the key aspects is scores from 0 to 4. Information about two AI-based applications used in the study are also described.

Furthermore, information about data collection and data analysis is given. Data is divided into three types: 1) scores of the pretest, 6 periodic tests, and the posttest; 2) data related to the average time spent pre assignment, recorded by the participants in both groups; 3) data related to the questionnaires answered by the experimental participants and the teacher in charge. The results of the gathered data along with several statistical analyses are discussed in the following chapter.

4 Results

In this chapter, the data related to the proficiency level of the participants, as well as the results gained from analysis of the collected data based on the scores of different essay writing tests are presented. In addition, some data related to the amount of time spent on writing home assignments recorded by the students, in both the experimental and control group, are included. The analysis was done within R studio, using a series of statistical models like linear mixed-effects models which are explained in detail in the following sections. The main focus of this chapter is to see: 1) how the students in both groups performed in the writing pretest, the 6 periodic tests, and the posttest; 2) which aspect/aspects of writing- task fulfillment, organization, grammar, lexis and mechanics- have been affected more as a result of using AI-based writing tools; 3) whether using AI-based writing tools contributes to time efficiency in the process of writing; and 4) what the attitude of the students and the teacher in charge of the study are toward integration of technology into the process of language learning.

4.1 The General Proficiency

As explained in chapter 3, prior to the experiment, there was a general proficiency test. In fact, all the students in the language academy have to pass a proficiency test to be able to attend the next level. Due to ethical considerations, the test could not be added here, but the scores were available for the purpose of demonstrating the homogeneity of the students. Generally, the test was scored out of 100, and the students who gained more than 69 were eligible to take part in the next level. It must be noted that this proficiency test was structured the same as the standardized Oxford proficiency test. Basically, there were combinations of questions testing three language skills: listening, reading, and writing. Moreover, there were some questions concerning vocabulary and grammar adapted to the level. The participants in this study could gain scores ranging from 70 to 98 which would confirm they were eligible to attend the intermediate level.

To be more specific, the scores' ranges in the experimental group were between 70-95, with mean=83.06250, median= 83; and the scores in the control group ranged between 70-98, with mean=82.55556, median= 80.5. Based on this data, figure 1 shows the comparison of proficiency level between two groups.

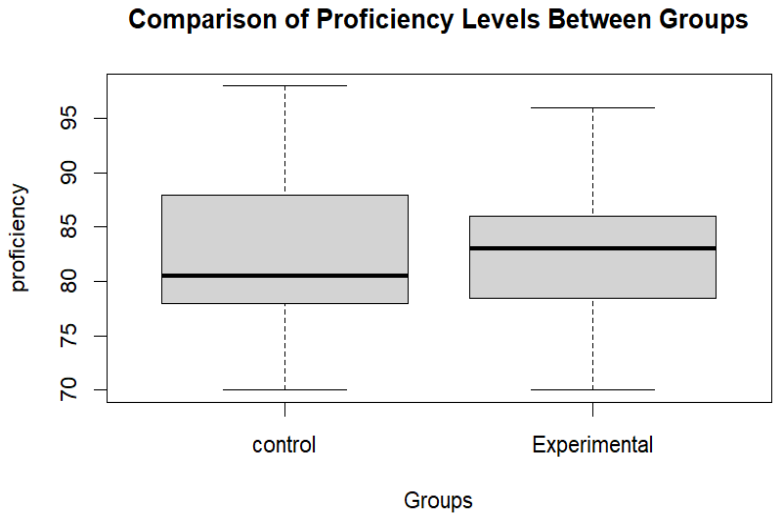


Figure 1*.General Proficiency Level of the Participants*

In addition, a Welch two sample t-test was run to compare the means of proficiency test between the two groups. Accordingly, the result showed that t = -0.19673, df= 31.807, and p-value =0.8453 with 95% confidence interval. The p-value (0.8453 > 0.05) indicates that there is no significant difference between the means of scores in the proficiency test in both groups. Therefore, the two groups were balanced in terms of general proficiency level.

4.2 Writing Proficiency Pretest

As mentioned in chapter 3, after the general proficiency test, to ensure that the participants were also homogeneous in terms of writing skills, an essay writing pretest was conducted for both groups with the same conditions. Accordingly, they were given a topic, selected from the previous student book, to write an essay between 140 to 190 words in 40 minutes. The essays were scored based on the rubrics discussed in chapter 3, considering all five aspects of writing, which had 4 points each so that the total score for each essay was out of 20.

Regarding the scores of the writing pretest in the experimental group, the scores ranged from 9.5 to 12.5, with the mean= 10.9, and median= 10.8. In addition, the scores of the writing proficiency pretest, in the control group, ranged from 8.5 to 12.5, with mean= 10.6, and median= 10.5. Accordingly, figure 2 presents the comparison between the control and the

experimental group in terms of writing proficiency before the main experiment, and it confirms their homogeneity in writing proficiency.

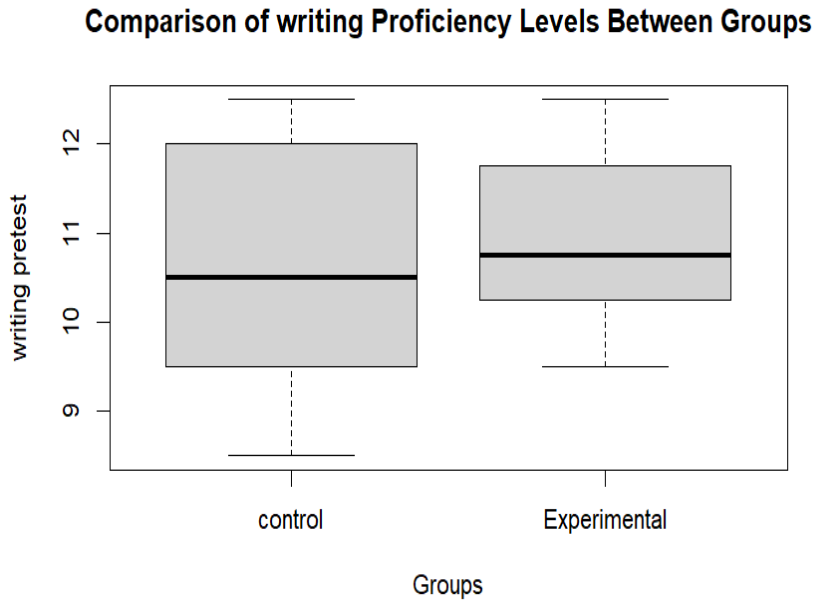


***Figure 2**. Writing Proficiency Level*

Moreover, to compare the means of scores of writing pretest in both groups, a Welch two sample t-test was run. The result was in this way: t = -0.58456, df = 31.167, and p-value = 0.5631 with 95% confidence interval. As a result, the p-value (0.5631), which is significantly higher than 0.05, indicates that there is no significant difference between the means of scores in writing pretest in both groups. Therefore, the two groups were balanced in terms of writing proficiency level.

4.3 Comparison Between the Pretest, the 6 periodic tests, and the Posttest

In addition to the writing proficiency pretest in both groups, there were six periodic tests throughout the study, and a posttest at the end of the study. It must be emphasized that conditions for both control and experimental groups were the same for the tests: essays were written using pen and paper, and the use of writing applications was not allowed. For all of the essay tests, the participants were supposed to write about 140-190 words in 40 minutes about the assigned topic. After the essays were scored based on the rubrics, the data was collected.

Table 2 shows the means of the total scores from the pretest, the 6 periodic tests, and the posttest in both groups.

Table 2. *Means of the Tests' Scores for the two Groups*

Tests	***Control Group***	***Experimental Group***
Pretest	10.66	10.90
Test1	11.02	11.71
Test2	11.44	12.71
Test3	11.52	13.93
Test4	12.33	14.84
Test5	12.41	16.09
Test6	12.97	16.93
Posttest	13.13	17.68

Accordingly, the means of all tests in experimental group range from 10.9 to 17.68, while the means of tests in control group range from 10.6 to 13.1. In this way, based on the data presented in table 2, figure 3 illustrates the interaction of the means of scores based on test and group:

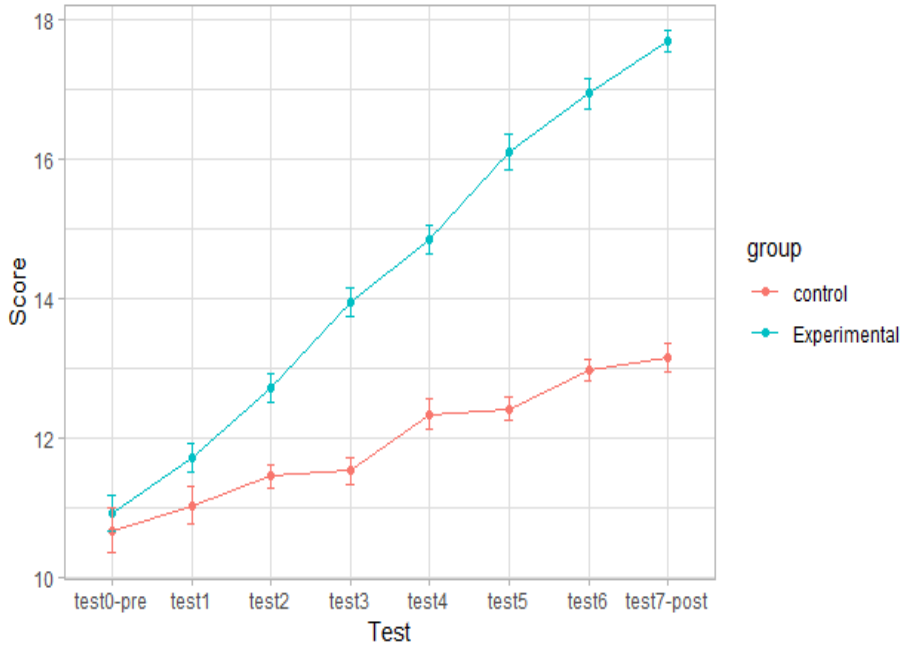


***Figure 3**. Interaction Between Means of Scores and Types of Tests for the two Groups*

In figure 3, the different tests ranged chronologically are displayed on the x-axis, while the mean score is given on the y-axis. The data for the control group is in red, while the experimental group´s data is pictured in blue. Based on the figure, we can see that both groups experienced an upward trend from test 0 (pretest) to test 7 (posttest). In fact, there is a slight increase in the means of scores in the control group which has occurred gradually over time. The experimental group, on the other hand, experienced a sharper increase in the means of the scores from test 0 to test 7 ending up with a mean score on the posttest which is significantly higher than the mean for the control group. In other words, while the means of the pretest scores are almost in the same level, the means of posttest scores are highly different, with the experimental group much higher than the control group. Obviously, the experimental group has experienced a sharp growth in terms of overall writing development over the course of the study.

Moreover, to test if the increase in scores from the pretest to the posttest was significantly larger for the experimental group compared to the control group, we fitted a linear mixed-effects model (score~ group * test+ (1| participants)), in R using the lme4 package. The dependent variable is the test score, and the predictors are the tests (Pre/Post), and the groups

(Control/Experimental) and the interaction between tests and groups. In addition, the model includes a random intercept for the participants. In general, the linear mixed-effects model was conducted to examine how the scores vary based on the type of group, kind of test, and different participants. Table 3 presents the summary of linear mixed-effects model using the lme4 R package (Bates et al., 2015):

Table 3. *Linear Mixed-Effects Model*

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: score ~ group * test + (1 | participants)
Data:
## Subset (big.data, test == "test0-pre" | test == "test7-post")
## REML criterion at convergence: 193.9
## Scaled residuals:
##    Min       1Q    Median       3Q        Max
## -2.13780 -0.67330   -0.01425   0.64617   1.82442
## Random effects:
## Groups        Name         Variance    Std.Dev.
## participants (Intercept)   0.01921      0.1386
## Residual                   0.99563      0.9978
## Number of obs: 68, groups:  participants, 34
## Fixed effects:
                                    Estimate Std. Error  t value  Pr(>|t|)
## (Intercept)                       10.6667     0.2374  44.923  4.069995e-50
## groupExperimental                  0.2396     0.3461   0.692  4.914403e-01
## testtest7-post                     2.4722     0.3326   7.433  3.221401e-10
## group Experimental:testtest7-post 4.3090     0.4849   8.887  8.910890e-13
## Correlation of Fixed Effects:
                  (Intr)  grpExp  tstt7-
## grpExprmntl   -0.686
## tsttst7-pst   -0.700   0.480
## grpExprm:7-    0.480  -0.700  -0.686
```

To elaborate more about the table, REML criterion at convergence of 193.9 confirms that the model fits well with the data. In addition, scale residuals suggests that there is a good fit between model and data, since the residuals are ranged from -2.13780 to 1.82442, which is almost close to zero. The intercept value for the control group in the pretest is 10.6. The group experimental coefficient gives the difference between the control and the experimental group in terms of the pretest which is not significantly different (as presented in figure 2). The value of 2.472 shows that the control group scored 2.472 higher in the posttest in comparison to the pretest. In addition, the difference between the control and the experimental is 4.309, which shows that the experimental group scored 4.309 higher than the control group. In sum, the exceptionally low value of p (3.221401e-10 <0.05) associated with t-value of 7.433 indicates that there is a significant difference in posttests in experimental group in comparison to the control group.

To be more specific, there was a comparison only between the pretest and the posttest in both groups, without considering the periodic tests. Based on the data presented in table 2, figure 4 shows a comparison between the means of the pretest and the posttest, and figure 5 shows the interaction plot of the pretest and the posttest in both control and experimental groups. In both figures, x-axis refers to different tests and Y-axis refers to the means of the scores in both groups.

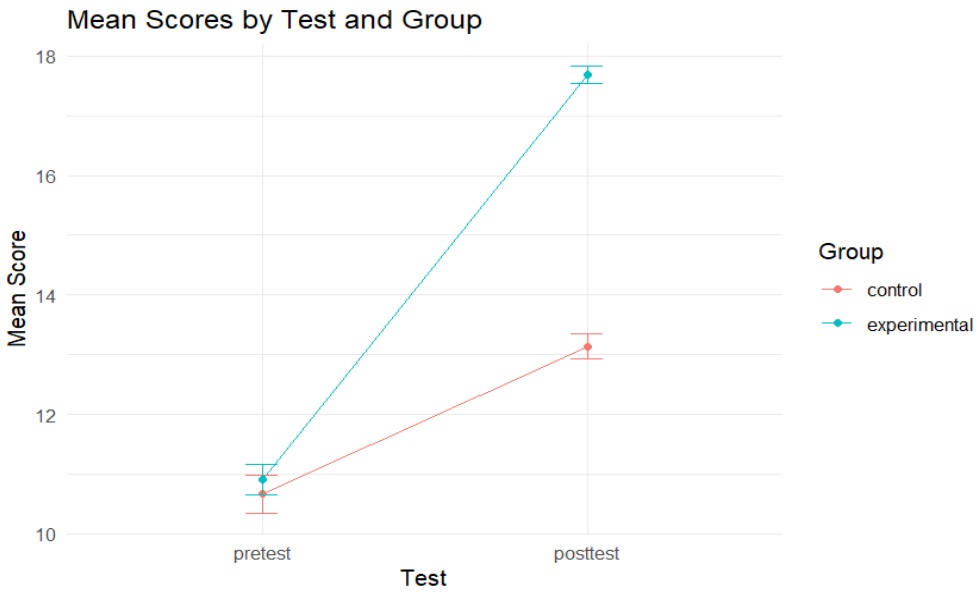


Figure 4. *Means of Scores in the Pretest and the Posttest*

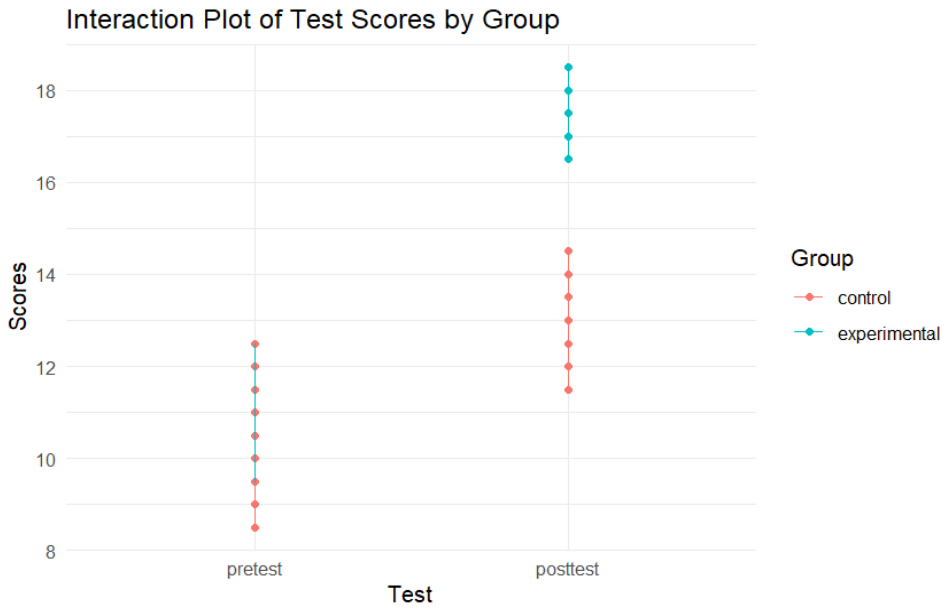


***Figure 5**. Interaction Plot of the Pretest and the Posttest*

Based on figure 4 there is an upward growth for both groups from pretest toward posttest, while the growth in the experimental group is so much higher and incredibly significant in comparison to the control group. Therefore, the graph shows a highly significant improvement in the experimental group in terms of writing skills. In addition, as figure 5 presents the interaction plot of pretest and posttest, both groups have the same level in terms of the pretest, while in the posttest they are noticeably different, with the experimental group outperforming the control group.

4.4 The Analysis of Five Key Aspects of Writing

As discussed in chapter 3, the essays were scored based on the rubrics presented in table 1. Accordingly, there were five key aspects in writings- task fulfillment (TF), organization (O), grammar (G), lexis (L), and mechanics (M)- each of which was assigned points from 0 to 4. Table 4 presents the means of the scores of all five key aspects in the pretest, the six periodic tests, as well as the posttest for both groups. In addition, following the data in table 4, there are some figures depicting the comparison of the means for each aspect in the two groups.

***Table 4**.Means of the Writings' Key Aspects*

	Experimental					*Control*				
	TF	**O**	**G**	**L**	**M**	**TF**	**O**	**G**	**L**	**M**
Pretest	2.7	2.2	2.1	2	1.6	2.5	2.2	2.1	1.9	1.7
Test 1	2.8	2.4	2.2	2.2	1.9	2.6	2.3	2.2	2.1	1.6
Test 2	2.9	2.5	2.5	2.3	2.2	2.7	2.4	2.3	2.1	1.6
Test 3	3	2.7	2.9	2.5	2.6	2.8	2.4	2.3	2.1	1.6
Test 4	3	2.9	3.1	2.8	2.8	2.9	2.6	2.6	2.3	1.7
Test 5	3.3	3.1	3.2	3.2	3.1	2.8	2.6	2.7	2.3	1.7
Test 6	3.3	3.3	3.1	3.5	3.5	2.9	2.7	2.7	2.3	2.1
posttest	3.5	3.2	3.1	3.8	3.8	2.9	2.8	2.7	2.4	2.2

According to table 4, the lowest point in the experimental group in the pretest belongs to mechanics followed by lexis, grammar, organization, and finally task fulfillment which is the highest. The same hierarchy is true for the control group in the pretest. However, regarding the posttest in the experimental group, compared to the pretest, both mechanics and lexis have significantly improved, followed by grammar and organization which improved one point similarly, and finally the task fulfillment has improved slightly from 2.7 to 3.5. Regarding the posttest in the control group, the scores follow the same hierarchy as the pretest, and they have slightly increased. Further explanations and illustrations regarding the trend of each aspect of writing is presented in the following sections.

4.4.1 Task Fulfillment

As mentioned above, one of the key aspects of writing, which was considered in scoring the writings, was task fulfillment. In fact, task fulfillment assesses how much the text produced by

the learners addresses the requirements of the task in terms of topic, content, length, and format. Based on the means of the scores related to task fulfillment in each test, the following figure depicts the results of the comparison between the two groups:

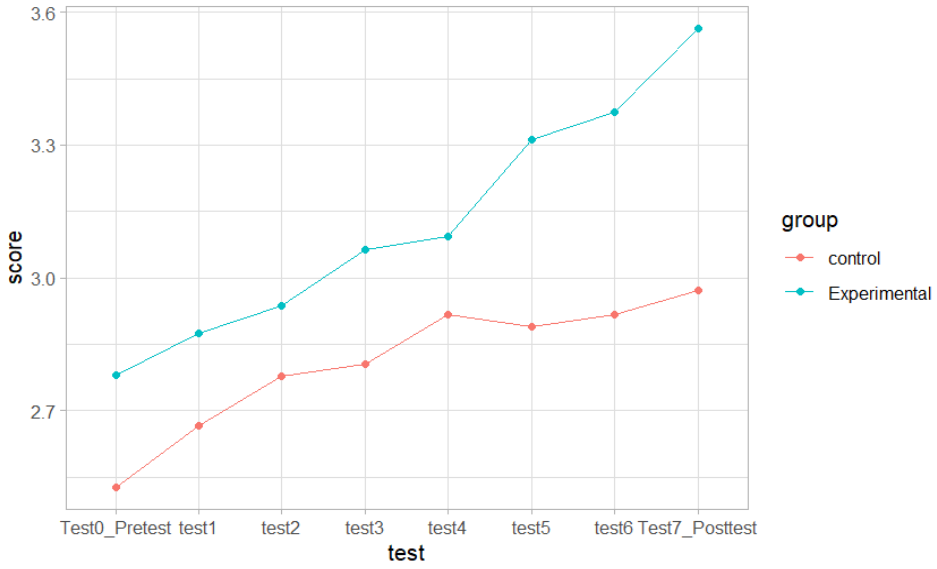


Figure 6. *Comparison of Task Fulfillment Development in the two Groups*

As it can be seen, task fulfillment in the experimental group has increased from approximately 2.7 in the pretest to 3.5 in the posttest, and in the control group the range is from 2.5 to 2.9. Although both groups met the growth in task fulfillment, the growth in the experimental group is slightly higher than the control group. Improvement in task fulfillment confirms that they could produce longer text and richer content meeting the requirements of the task properly.

4.4.2 Organization

The second textual aspect assessed was organization that considered cohesion and coherence. In other words, it refers to how the ideas are arranged. In fact, use of cohesive devices, linking words, and solid flow of the text were given higher points. The following figure compares the means of scores related to this aspect throughout the study for both groups.

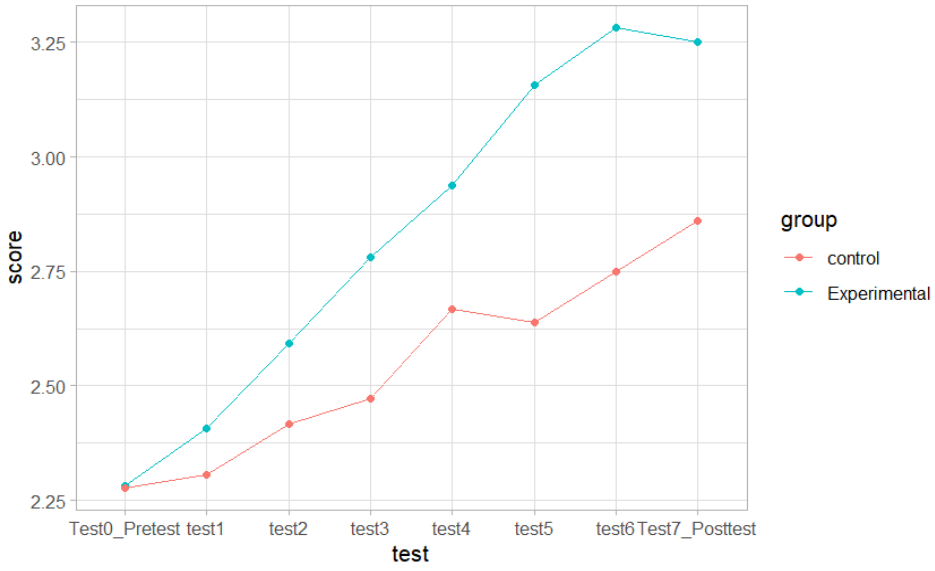


Figure 7*. Comparison of Organization Development in the two Groups*

As figure 7 shows, the means of scores related to organization aspect of writing in the experimental groups has increased from 2.2 to 3.2, while in the control group the growth is from 2.2 to 2.8. Therefore, both groups made improvements in their writing production when it comes to text organization and devices promoting coherence and cohesion. However, the experimental group could produce texts with somewhat higher scores in this area in comparison to the control group.

4.4.3 Grammar

The third aspect in rubrics for scoring the writings was grammar, which considered grammar accuracy and the extent to which learners could produce sentences with correct structure based on the lessons taught in the class and their level. As with the other aspects, grammar was assigned points from 0 to 4 for each test. Figure 8 presents the comparison of means of grammar scores in the two groups.

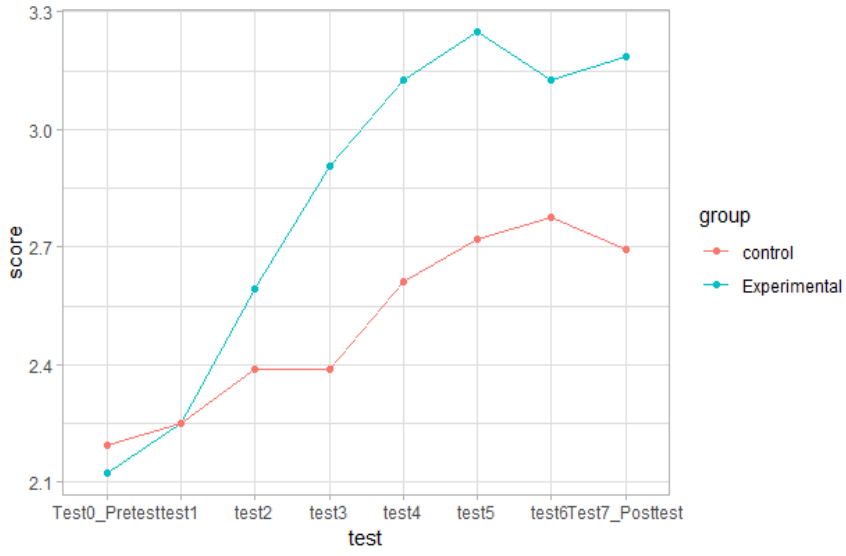


Figure 8. *Comparison of Grammar Development in the two Groups*

Based on the figure, both groups exhibited an upward trend in terms of grammar performance. Specifically, the experimental group experienced an increase in grammar level of about 1 point, from 2.1 to 3.1. Similarly, the control group increased its scores in this area, from approximately 2.1 to 2.7, a bit lower than the experimental group. Moreover, there are some fluctuations in the trend of grammar development in both groups.

4.4.4 Lexis

The fourth key aspect of writing rubrics was lexis which referred to the amount of vocabulary used in the essays such as the words, phrases and expressions based on their level and the context of the topic. In this regard, lexis was scored from 0 to 4 in each test, and figure 9 presents the comparison of means related to lexis points for both groups.

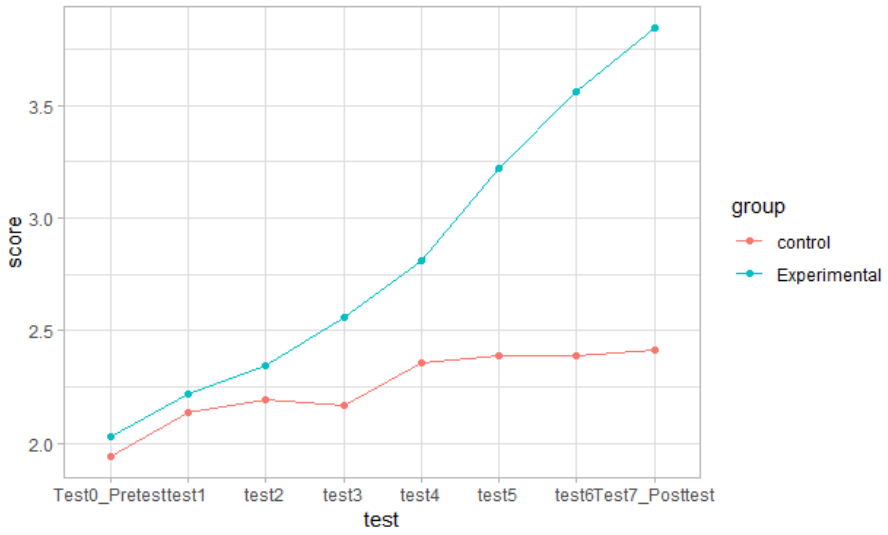


Figure 9. *Comparison of Lexis Development in the two Groups*

We can see from the figure that the experimental group developed significantly throughout the study, showing a sharp upward trend from approximately 2 to 3.8. In fact, the experimental group has experienced a constant rise in lexis development. In contrast, the control group showed only moderate growth in this area from roughly 2 to 2.4 over time. In fact, the amount of vocabulary remained stable in some while from test 4 to the posttest for the control group, as can be seen in the relatively flat line in the figure during this time frame.

4.4.5 Mechanics

As discussed in chapter 3, the fifth key aspect of scoring writing was mechanics which referred to spelling, capitalization, and punctuation. The mechanics aspect was also scored from 0 to 4, and figure10 depicts the comparison of means of scores related to the mechanics of the tests for both the experimental and the control group.

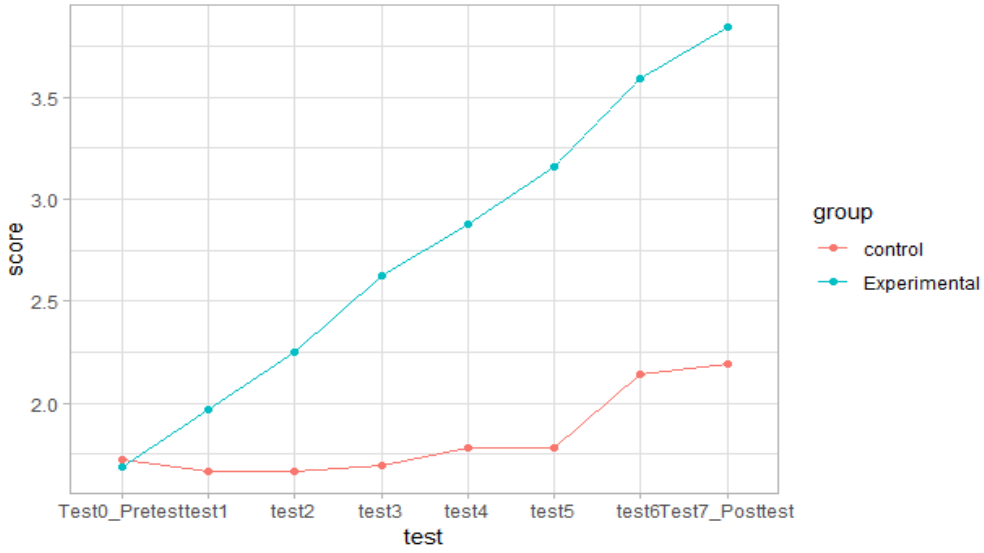


Figure 10. *Comparison of Mechanics Development in the two Groups*

According to the figure, starting from approximately the same level, the experimental group has experienced a tremendous increase in terms of mechanics improvement over time from test 0 to test 7, while the control group has remained stable from the pretest to test 5. Although the control group showed a slight increase from test 5 to posttest, this is so much lower than the sharp growth in the experimental group. To be more specific, from the pretest to the posttest, the means of mechanics increased impressively from 1.6 to 3.8 in the experimental group, while in the control group it increased only slightly from 1.7 to 2.2.

4.4.6 Improvement from the Pretest to the Posttest for the Individual Rubrics

To elaborate more on the five key aspects of writings-TF, L, M, G, O- in the two groups, for each participant and each aspect of the rubric, the differences between the pretest scores and the posttest scores were calculated and named “improvement scores”. Figure 11 depicts the overview of the improvement of the key aspects in the two groups.

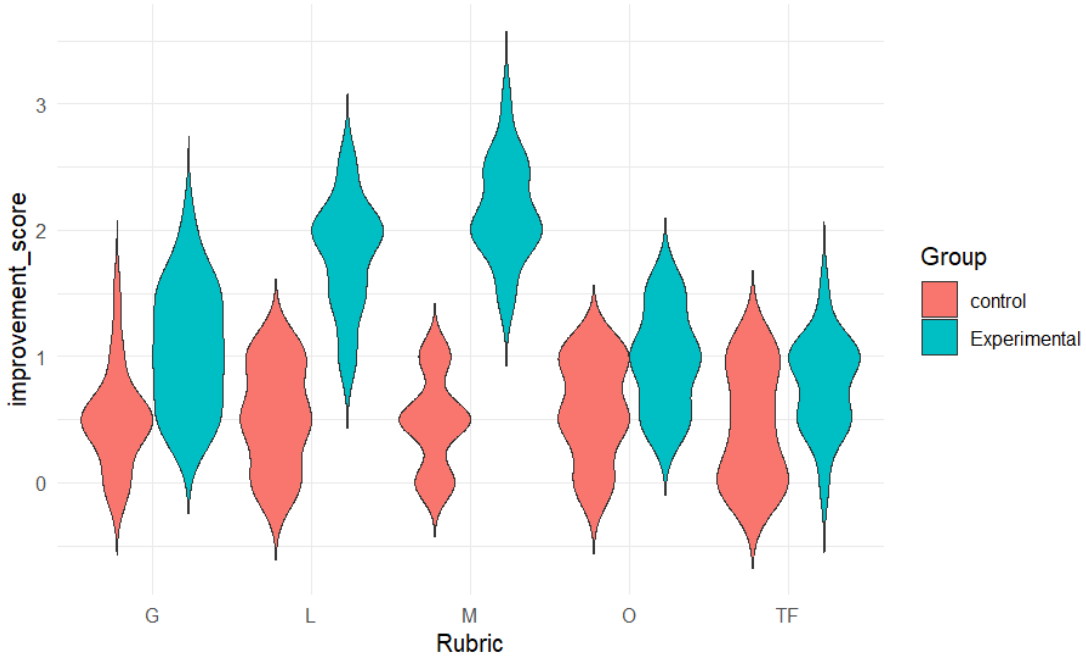


Figure 11. *Improvement of the Rubrics*

In the figure, x-axis refers to the five key aspects of rubric and y-axis refers to the value of improvement. Based on the figure, generally it seems that improvement has happened for all of the participants to some extent. In the experimental group three aspects, mechanics, lexis, and grammar have improved significantly, while there seem to be less significant improvement in task fulfillment and organization of the writings. Regarding the control group, the value of improvement is 0.5 to 1 score for most of the key aspects, which is not significant in comparison to the experimental group.

In order to see whether the improvement of the key aspects in the two groups was significant, a linear mixed-effects model was run, and table 5 provides the details.

***Table 5**. Linear Mixed-Effects Model*

```
Linear mixed model fit by REML ['lmerMod']

Formula: improvement_score ~ Rubric * Group + (1 | participants)
   Data: data

REML criterion at convergence: 173.7


Scaled residuals:
     Min       1Q   Median       3Q      Max
-1.87561 -0.67668  0.01123  0.62677  2.26909


Random effects:

Groups       Name         Variance   Std.Dev.
 participants (Intercept)  0.05035   0.2244
 Residual                  0.11524   0.3395

Number of obs: 170, groups:  participants, 34


Fixed effects:
                           Estimate Std. Error t value
(Intercept)                 0.50000    0.09591   5.213
RubricL                     0.02778    0.11316   0.245
RubricM                    -0.02778    0.11316  -0.245
RubricO                     0.08333    0.11316   0.736
RubricTF                   -0.05556    0.11316  -0.491
GroupExperimental           0.56250    0.13982   4.023
RubricL:GroupExperimental   0.72222    0.16495   4.378
RubricM:GroupExperimental   1.12153    0.16495   6.799
RubricO:GroupExperimental  -0.17708    0.16495  -1.074
RubricTF:GroupExperimental -0.22569    0.16495  -1.368


Correlation of Fixed Effects:
            (Intr) RubrcL RubrcM RubrcO RbrcTF GrpExp RbL:GE RbM:GE RbO:GE
RubricL     -0.590
RubricM     -0.590  0.500
RubricO     -0.590  0.500  0.500
RubricTF    -0.590  0.500  0.500  0.500
GrpExprmntl -0.686  0.405  0.405  0.405  0.405
RbrcL:GrpEx  0.405 -0.686 -0.343 -0.343 -0.343 -0.590
RbrcM:GrpEx  0.405 -0.343 -0.686 -0.343 -0.343 -0.590  0.500
RbrcO:GrpEx  0.405 -0.343 -0.343 -0.686 -0.343 -0.590  0.500  0.500
RbrcTF:GrpE  0.405 -0.343 -0.343 -0.343 -0.686 -0.590  0.500  0.500  0.500
```

Based on the table, the experimental group gained improvement scores 0.56 units higher than the control group which is statistically significant. The experimental group improved in terms of mechanics, lexis and grammar significantly. In addition, negative estimated coefficient in

the rubric for task fulfillment and organization indicates that the experimental group improved less than what expected.

Moreover, table 6 presents the proportion of the participants who improved and gained the improvement scores above 0, 0.5, and 1 in the rubrics.

Table 6*. Improvement in the Rubrics*

```
Tapply(data$BinAbove0, list(data$Group, data$Rubric), mean)

                  G         L         M         O         TF
control      0.7777778 0.7222222 0.7222222 0.7777778 0.5555556
Experimental 1.0000000 1.0000000 1.0000000 1.0000000 0.9375000


> tapply(data$BinAbove0.5, list(data$Group, data$Rubric), mean)

                  G         L         M         O         TF
control      0.1666667 0.3333333 0.2222222 0.3888889 0.3333333
Experimental 0.6875000 1.0000000 1.0000000 0.6875000 0.5625000


> tapply(data$BinAbove1, list(data$Group, data$Rubric), mean)

                  G        L      M     O      TF
control      0.05555556   0.000   0   0.00   0.0000
Experimental 0.37500000   0.875   1   0.25   0.0625
```

Based on the table, all the participants in the experimental group improved above 0 across the five aspects, while in the control group the proportions vary, and they are lower than the experimental group. Regarding the improvement scores above 0.5, there are higher proportions of the participants in the experimental group in comparison to the control group, especially in terms of mechanics, organization, lexis, and task fulfillment. About improvement scores above 1, again the experimental group shows a greater proportion of the participants compared to the control group who could only improve in grammar.

4.5 The Result of Time Spent for Home Assignment

As mentioned in chapter 3, participants in both groups were asked to record the time spent on doing home assignments to know whether the development simply refers to the time spent over writing or not. In fact, both groups were assigned a topic each session to write an essay about 140 to 190 words and record the time before delivering. The following figure presents the comparison of means of average time spent for doing per assignment for the two groups.

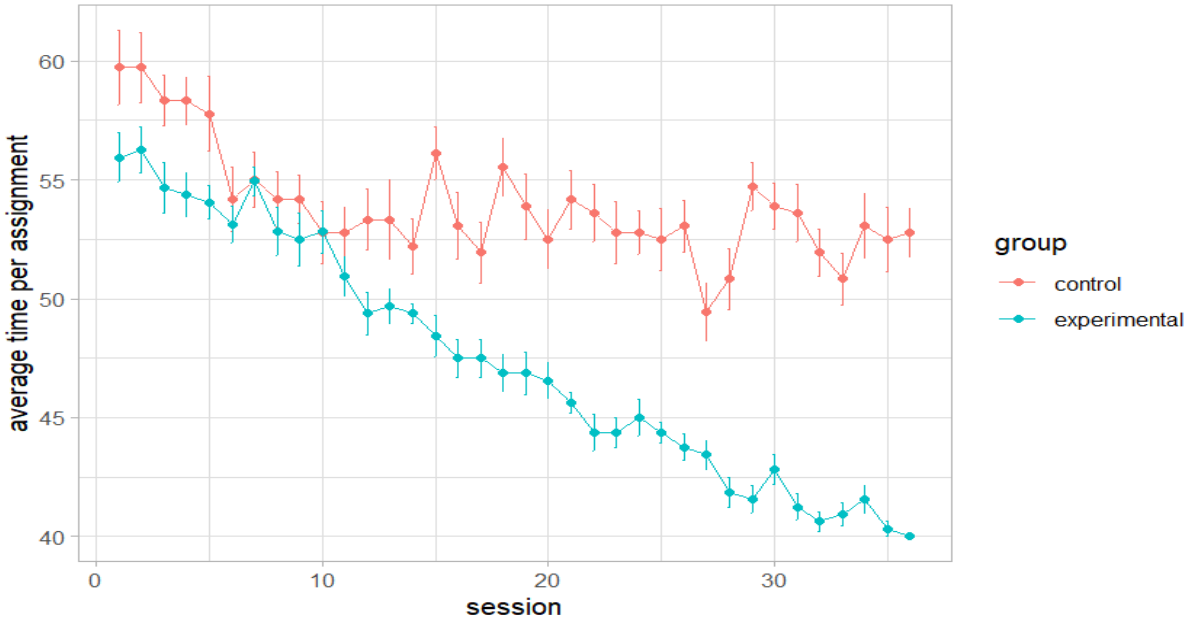


Figure 12. *Average Time Spent per Assignment*

Based on figure 12, x-axis refers to the number of sessions which was 36 over this study, and y-axis shows the average time spent per assignment for each group. As recorded by the participants, time spent per assignments ranged from 40 to 60 minutes in the experimental group, and from 45 to 70 minutes in the control group. Accordingly, the control group, using traditional pen-and-paper method for writing assignments spent a bit more time than the experimental group who used writing applications to generate essays. In fact, both groups started with spending relatively the same amount of time, but after about ten sessions the participants in the experimental group did the assignments a bit faster. It must be emphasized that the topics and the range of words (140-190) for the assignments were the same for both groups.

4.6 The Students' Attitude

In order to know the students' attitude toward integration of technology specially AIEd into language learning, the students in the experimental group were asked to fill out a questionnaire (see appendix 2). According to the survey, all the students agreed that using the applications was helpful in improving general English and specifically in developing their writing skill. Based on the results gained from the questionnaire, all the students agreed that the applications were helpful in terms of the aspects mentioned below, which is presented in figure 13 as well:

a) By providing feedback on my grammar errors
b) Punctuation, capitalization, and spelling checkers
c) Sentence suggestions
d) Faster writings
e) By helping me practice my writing skills

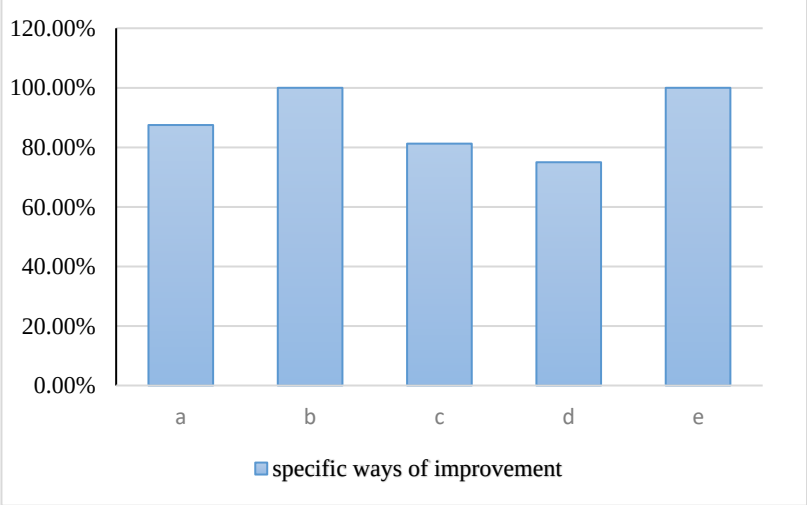


***Figure 13**. Ways of Improvement*

Based on the graph, all the students found AI-base writing tools helpful in providing opportunity to practice their writing skills, as well as providing instant correction for spelling, capitalization and punctuation. In addition, about 90% of the students believed that AI-base writing tools were helpful by providing feedback on grammar errors. Moreover, about 80% of the students found the feature of sentence suggestion very practical and interesting. Plus, 70%

of the students found writing faster when using AI-base writing tools and were satisfied with the high writing pace.

About the liked features, among the suggestions, the students mostly agreed that the best feature of the applications is providing comprehensive written corrective feedback which includes accurate grammar and spell checking. The following features were listed in the survey and figure 14, depicts the results as well:

a) User-friendly Interface
b) Helpful writing suggestions
c) Accurate grammar and spelling checks
d) Comprehensive feedback

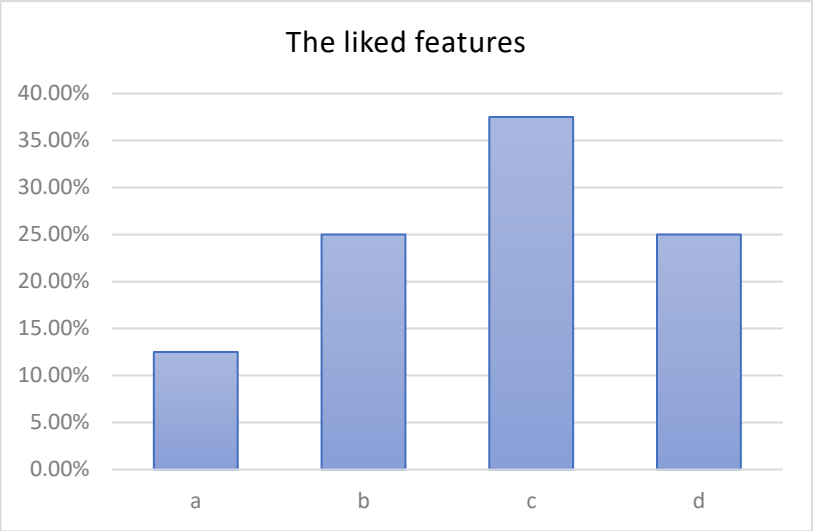


***Figure 14**. Pros of Using Writing Applications*

Based on the bar chart, the best feature of the writing applications is providing accurate grammar and spell checking. Plus, providing comprehensive feedback and helpful writing suggestions are two features that 25% of the students liked. In addition, approximately 15% of the students believed that the apps were user friendly.

Furthermore, students faced some challenges when using the mentioned writing applications in this project. One of the most frequent challenges that the students faced was that *"*Wordtune*"* wasn't available 24/7 for free, if so, they had to pay for full version. Another challenge was that they sometimes found some technical bugs and did not know how to solve them. However, the students expressed that they were generally satisfied with using such applications and would recommend it to other peers as well.

4.7 The Teacher's Attitude

Based on the questionnaire related to the teacher's attitude toward implementing AI-based writing applications in the lesson (see appendix 3), the teacher in charge of the experimental classroom had a positive attitude and expressed her overall satisfaction. As a result of applying this method, the teacher found AI-based applications helpful for the learners and also for herself. In fact, it was found out that the students were more motivated and engaged in the classroom than her previous experiences of teaching. In addition, the teacher found the instant corrective feedback, provided by the writing applications, so effective for increasing awareness about the recurrent errors and for development of the students' writing skills. Moreover, implementing writing applications was beneficial for the teacher in terms of saving the time that was spent over correction of all recurrent errors and giving corrective feedback individually. Since the correction part of the assignments was done by AI and students had received automatic corrective feedback beforehand, the teacher had to only revise them. In addition, lots of common errors previously made by students were automatically corrected and the students had improved and tried not to repeat them in test papers.

However, some concerns were expressed by the teacher about writing applications such as lack of access to free versions, possible technical bugs, and the students' dependence on them. Finally, she expressed that further training for both students and teachers is necessary to get the most of integration of technology into language learning.

4.8 Chapter Summary

The main objective of this chapter was to examine whether AI-based writing applications had any impact on the EFL learners' writing development, and whether the experimental group and the control group had any significant differences in terms of writing development. In addition, it was aimed to see if there were significant differences between both groups, which key aspect

of writing had been affected more. Furthermore, the comparison of time spent per assignment, recorded by both groups, were compared to see if there were significantly different. Finally, the results regarding the attitudes of the students and the teacher toward AIEd especially in language learning were presented.

The result of some statistical analysis such as linear mixed-effects models between the scores gained from the pretest, the 6 periodic tests, and the posttest revealed that the experimental group outperformed the control group in terms of writing development. Furthermore, the comparison of five key aspects including task fulfillment, organization, grammar, lexis, and mechanics showed that the mechanics and lexis improved more than other aspects, followed by grammar and organization which improved similarly. Finally, the task fulfillment aspect of writing was affected less than others. In fact, in comparison to the pretest, the experimental group could improve in terms of error corrections related to punctuation, spelling, and capitalization, and they could also use more diverse vocabulary with highly accurate structure leading to coherent and cohesive texts.

The result of the comparison between the two groups in terms of time spent per assignment showed that they started with the same rate while the experimental group spent a bit less time for later sessions after session 10.

The results of the survey about the students' attitude revealed that they were satisfied with having a new experience and recommend it to the other students. In fact, the most important feature for the students was providing error corrections related to spelling, capitalization, and punctuation. Similarly, the teacher in charge of the study expressed her positive attitude and stated that the study increased the student's motivation and engagement, and also, the applications acted like an assistant correcting recurrent mistakes and giving her more time to devote for other tasks. Finally, some concerns were expressed such technical bugs, lack of full access to the applications, and need for further training.

5 Discussion

The focus of this chapter is to discuss the research questions and the predictions raised in chapter 3 as well as the results presented in chapter 4. As previously mentioned, the present study addresses the following research questions:

5. What is the impact of using AI-based writing applications on the EFL learners' writing skills?
6. Which aspect of writing among task fulfillment, organization, grammar, lexis, and mechanics is mostly influenced due to using AI-based writing tools?
7. Does using AI-based writing applications contribute to time efficiency in the writing process?
8. What are the learners' and the teacher's attitudes toward AIEd specially in the process of language learning?

Accordingly, the following assumptions were formed as well:

1. Based on the model for AIEd proposed by Chiu et al., (2023) and the previous research findings, it is predicted that using AI-assisted feedback tools improves overall writing skills in EFL context.
2. Based on Ellis (2008) strategy of direct CF, and also based on CGF and AWCF, it is predicted that as a result of observing errors corrected automatically, grammar, lexis and mechanics of writing are affected and improved more than other aspects.
3. Using computer generated sentences and benefiting AWCF, it is predicted that the students in the experimental group spend less time on generating texts.
4. Based on the previous research, it is predicted that students experience a new learning environment through which they have more confidence and motivation toward learning a foreign language. In addition, based on the previous research, EFL teachers suffer from a great burden of giving feedback and error corrections. It is predicted that they benefit in several ways as well.

Considering the results presented in chapter 4, the research questions and predictions are discussed one by one in the following sections.

5.1 The Role of AI-Assisted Writing Tools in the EFL Learners' Writing Skills

Regarding the first research question, it was predicted that the implementation of AI-assisted writing tools would have a positive impact on the development of the EFL learners' writing skills. As presented in section 4.3, the result of a linear mixed-effects model, comparing the scores collected from the pretest, the 6 periodic tests, and the posttest, revealed that the participants in the experimental group outperformed the control group in overall writing scores. In addition, the picture of writing development trend for both groups is presented in figure 3. the participants started the study with almost the same level of writing proficiency while ultimately the experimental group developed in overall writing skills far more significantly than the control group as a result of utilizing AI-assisted writing tools. This is in line with the findings of Choo and Li (2017), Ghufron and Rosyida (2018), Ghufron (2019), Barrot (2020), Rad et al., (2023), and Al Mahmud (2023) about the use of AI-assisted writing tools for enhancing writing skills.

To elaborate more, Al Mahmud (2023) and Rad et al. (2023) conducted an investigation about the impact of an AI-assisted writing tools such as "Wordtune" on the learners' writing skills which resulted in the experimental group's significant improvement. This may be due to several reasons the primary one being the AWCF that the learners received in the experimental group which was missed in the control group. As a result of AIEd, and due to working with AI-assisted feedback tools, the experimental group benefited from AWCF while the control group received teacher CF which is mostly limited and delayed. As previously mentioned, students learning a foreign language need more exposure as well as more personalized feedback which AI-assisted feedback tools are able to provide. Particularly in the EFL contexts where competent and native teachers are not available, and where teachers are highly occupied with a lot of responsibilities and mostly have insufficient time, the fact that such tools seem to fulfill these needs is a highly promising result. The students can benefit from the advantage of AI feedback tools that have infinite time and patience to give detailed personalized feedback which is not found in teacher-generated feedback, especially for a large number of students. This shows that language institutes and educational organizations should invest in AIEd and AI-assisted language learning tools on behalf of the students.

The findings are further supported by Ghufron and Rosyida (2018) who found that due to working with Grammarly software several times, the learners were able to select not only the

correct form of mechanics but also the correct grammar and lexis, which significantly improved their writing skills. In line with their findings, the students in the experimental group who received AWCF had a significant reduction on their errors and could produce error free texts to a great extent. This may be due to increasing self-corrections, developing self-regulations, categorizing the errors, and personalized feedback provided by AI-assisted writing tools (O'Neill & Russell, 2020). Similarly, the findings are in line with much research about the impact of AWCF on the development of writing skills specially in an EFL context, through which the learners can recognize their errors and improve them (Daniels & Leslie, 2013; Saadi & Saadat, 2015; Qassemzadeh & Soleimani, 2016). Observing their errors corrected immediately and automatically increases the students' awareness and makes them think critically about language outcomes even in the absence of the tools. As reported by the students in the experimental group, when they were writing in the absence of the tools, they would continuously revise the sentences and do self-corrections which was not so before.

Language practitioners might worry that students may mostly rely on the tools and may not adequately produce texts in the absence of these artificial agents. The findings of the study shows that this is not true and the students through using applications developed the quality of their writings and decreased their errors to a great extent compared to the control group who had no access to such resources. In fact, the implementation of technology in the curriculum should be purposeful and used with disciplines so that students do not get distracted from the main aims of the learning, and also, they should become aware of the reasons of using such tools.

Furthermore, having anxiety and feeling embarrassed over mistakes is mentioned as common challenges for learning a foreign language (Hsu, 2015). Yet another obstacle for language learning in EFL contexts is anxiety and fear of making mistake. Sometimes fear of getting explicitly and directly corrected, and feeling embarrassed in the classroom makes the learners inactive and unwilling to participate in the activities. Saadi & Saadat (2015) found out that AWCF creates a less intimidating environment for the learners to revise their own work with less anxiety which contributes to being able to learn from their mistakes.

In sum, the findings indicate that, EFL learners can benefit from AI-assisted writing tools such as "Wordtune" and "Insta text" to improve their writing skills. In fact, the provided AWCF

gives the learners the opportunity to have more exposure to the language, to have immediate personalized feedback, to observe their errors as they occur, to identify their weak points, to learn from the mistakes, and to improve their writing skills in a less intimidating atmosphere. Therefore, the findings support the prediction 1.

5.2 The Impact of AI-Assisted Tools on Different Aspects of Writing Skills

Regarding the second research question, it was predicted that utilizing AI-assisted writing tools improves the experimental group's writing skills mostly in terms of mechanics, lexis, and grammar compared to the other aspects. Based on the results of the mixed-effects linear model presented in table 5 and also based on the figure 11, the experimental group improved in terms of mechanics, lexis and grammar more significantly than organization and task fulfillment. In fact, there is an interaction between being in the experimental group and improving in the rubrics. In addition, table 6 presents the proportion of the participants who improved above 0, 0.5, and 1, where the experimental group outperformed the control group in terms of the five aspects of writing. The control group on the other hand experienced slight improvement, not significant, in terms of grammar.

The findings are in line with Al Mahmud (2023), who found out that the students benefiting Wordtune-assisted writing improved significantly not only in terms of lexis and mechanics but also in terms of grammar, which means only these three aspects were affected significantly compared to other aspects. Similarly, based on the study by Ghufron and Rosyida (2018) about the impact of Grammarly on EFL learners' writing development, it was revealed that three aspects of writing such as grammar, mechanics, and diction improved significantly. Additionally, in a study by Lin and Yang (2011), due to the use of Wiki, the learners significantly improved in writing skills in terms of vocabulary, spelling, and grammar. Discussions regarding different aspects and how they responded to AI-based feedback tools are presented below.

5.2.1 Lexis

In terms of lexical development, the findings are in line with Dizon and Gayed (2021) who found out that the use of AI-based writing tools increases lexical diversity. In addition, in line with the findings of Aljohani (2021) and Zhao (2022), the learners in the experimental group could create texts with lexical resourcefulness, and used more contextually appropriate nouns,

adjectives, and verbs. Moreover, because of using “Wordtune”, the experimental group could make more compound and complex sentences in the writings supporting the findings of Rad et al. (2023) and Coenen et al. (2021). Benefiting from instant corrections, alternative words and sentence suggestions, the next sentence predictions, and autocompletion of the sentences provided by AI-based writing tools such as “Wordtune” and “Insta text” may be influential in lexis development in the experimental group. Compared to AI-generated feedback, teacher’s direct CF is usually restricted to the error corrections, and they rarely tend to offer open ended alternatives. In most of the cases when the students’ choice of the lexis is right, teachers do not offer alternative words or phrases. However, the writing feedback tools that automatically generate alternatives for the students will end up exposing broader lexical choices that the students might not come up with themselves. This can lead to the expanding of the vocabulary domains of the experimental students and is thus a source of lexical improvement.

Another point to consider is the kind of feedback tools, that whether they work as a leader in text generation or as collaborator, whether they assistant during the process of writing or work as a text evaluator after the text was generated by the learners, since different tools have divergent functions, strengths and weaknesses. It seems that richness of the lexis in the experimental group emerged from “Wordtune”, which works as collaborator in text generation and assist during the process of writing by offering alternative sentences based on the context set by the topic and the first lines. In addition, “Insta text” worked as complementary aid in final evaluation of the text to decrease the errors and possible bugs as much as possible. It can be inferred that generative AI tools have a special role in lexical development compared to purely evaluators of the text. Further research is necessary to see how different tools work for developing lexical sources, and whether the lexical development will disappear in the absence of generative AI tools.

5.2.2 Mechanics

The findings in terms of mechanics development due to utilizing AI-assisted feedback tools are in line with findings of Al Mahmud (2023) and Ghufron and Rosyida (2018). As recorded in chapter 4, table 4, the participants in both groups gain the lowest points in terms of mechanics in the pretest which showed that the learners had the most errors in spelling, punctuation, and capitalizations. However, by using AI-assisted writing tools such as “Wordtune” and “Insta text” the experimental group improved significantly in the posttest which may be due to

benefiting from AWCF and having more exposure to authentic language examples. Receiving automatic corrective feedback and observing how the error is corrected can be the source of improvements in the spelling, capitalization, and punctuations. Errors in mechanics are the most recurrent ones in EFL contexts and teachers find it demanding to correct all errors for all students, and to do so takes a lot of time. In addition, if the teachers devote time and correct the errors individually, it is impossible to provide the corrections in real time. Teachers' CF is mostly delayed and postponed to the next sessions, after the text is generated and the assignment is handed in. It is understandable that this kind of immediate error corrections, as the student is writing, has different psychological effects on the learning process compared to the delayed teachers' CF. The advantage of AI-feedback tools in providing immediate and personalized correction feedback as the errors occur, should be considered for the students' improvement in the errors related to mechanics of writings. Referring to the challenges for EFL learners such as lack of exposure, which can be the source of errors to some extent, AI tools can come to aid by providing authentic language examples. We know that having more exposure to a language can decrease the possible errors. Further research is needed to evaluate varied materials for increasing exposure to authentic language resources to see which is more effective in decreasing errors related to mechanics. Having the potential of providing alternative context-based sentences along with real time error corrections are among the advantages of AI-based writing tools that might cause improvements in the experimental group's writing skills.

5.2.3 Grammar

In the present study, in line with the above-mentioned studies, grammar improved significantly in the experimental group. Improvement in terms of grammar can be attributed to several resources. Similar to lexis and mechanics, students in the experimental group benefited from AWCF and improved in terms of grammar. In addition, as mentioned above, along with immediate correction of the errors, these tools provide several authentic alternatives with highly accurate structures that can be effective in improving grammar. Having exposure to grammatically accurate resources like AI text generators is another factor to be considered here.

A question one might come up with is why grammar improved less than mechanics and lexis, although it improved significantly in the posttest compared to the pretest. This may be due to mismatches between the grammar taught in the lessons and the text generated by the applications which was not focused specifically on the grammar taught in the classroom.

Although the text generators are creating sentences with accurate grammar, the point is that the textbooks taught in EFL classrooms are categorized based on the level of proficiency and it is expected to follow the taught material in the test papers which has an impact on the scoring as well. For example, when the focus of the lesson is "the present perfect tense", it is expected that the student produce an essay focusing on that tense to get a higher score. However, this can be controlled in the collaborative AI text generators by navigation of the text flow, and also by the related topic which is prompted first in the applications. Indeed, it takes sufficient training to be able to direct the text toward the goal of the task, and the students need instructions in advance. In addition, topics of the assignments should be related to the target grammatical point, for example topics like "a future invention" requires the students to write about the future and use the future tense. Further research comparing result with different proficiency levels and different subject matters would be interesting here.

Another question one might ask is why grammar improved in the control group as well, although not to the same degree as the experimental group. It can be argued that the development of grammar may be due to other factors besides AI-assisted tools, since it happened to some extent in the control group as well. This may be due to the time spent per assignment which is discussed in the next section. In fact, the control group spent slightly more time on doing assignment which may result in their grammar improvement as well. In addition, this may be due to the grammar exercises provided in the student book as well as the exercise book which were practiced in both classes. Therefore, the control group had access to grammar resources, although not the ones provided by the AI-assisted tools.

5.2.4 Organization and Task Fulfillment

Regarding the other key aspects of writing, organization and task fulfillment, the result showed that in the experimental group, they were placed as the least improved features in the hierarchy. Such findings are further supported by Ghufron and Rosyida (2018) and Al Mahmud (2023) who found that implementation of "Grammarly" and "Wordtune" fail to significantly affect organization and content of writings. This may be due to the fact that AWCF was not enough to focus on such aspects of writing. In other words, due to some research findings, key aspects of writing such as organization, argumentation strength, and coherence can be listed among the areas that AWCF may fail to help, which is attributed primarily to the complexity of human language and its pragmatic features (Huang & Wilson, 2021; Grimes & Warschauer, 2010).

This could be why organization and task fulfillment aspects encompassing cohesion, coherence, content, length and format improved least in the hierarchy.

However, it might be argued that how the organization aspects of a text can be improved even to some extent through collaborative AI text generators. The sentences generated by "Wordtune" are based on the topic prompted first which sets the context for the whole text. Therefore, the following suggestions and predictions are related to the topic and the context, leading toward a more coherent and cohesive text. In addition, one can navigate the flow of the text by prompting linking words and setting the direction of the next sentence. For example, while creating the text, one can type "in addition, or next" to direct the application to suggest a new idea, or it is possible to type "finally, or in conclusion" to conclude the text. However, again students need to be trained to control the text flow to produce a more organized text. It must be noted that such text navigation is possible with tools like "Wordtune" as a text generator, but not in "Insta text" which works as a text evaluator. Therefore, different tools should be investigated to see how such aspects can be affected more. In the study, although the students had instructions prior the experimental phase, the desired outcome was not met in terms of the organization and task fulfillment. Perhaps the proficiency level of the students was not high enough to expect such developments in the writing skills. Further investigations would be desired to compare the outcome at higher proficiency levels. In addition, task difficulty and the kind of textbook as well as the topics for the assignments should be considered in the future studies to see how the outcomes will differ.

In summary, the findings in this section indicate that, EFL learners benefiting from immediate error corrections, personalized corrective feedback, more exposure to authentic sentences, and alternative sentence suggestions provided by AI-assisted writing tools such "Wordtune" and "Insta text" could improve in writing skills in terms of mechanics, lexis, and grammar more than organization, and task fulfillment. Therefore, such findings support the prediction 2.

5.3 The Role of AI-Assisted Tools in Time Spent for Text Generation

The third research question concerns about the impact of time spent per assignment. The effect of time spent per assignment is examined to see whether the improvement of writing through AI-assisted tools in the experimental group is due to the amount of time spent generating a text or other factors are effective. If they spend more time doing the home assignments than the

control group, the writing development could be attributed in part to the amount of time spent. Therefore, the participants in both groups were required to record the time spent on doing the home assignments and creating texts. As the figure 12 shows the two groups initially spent the same amount of time, since training how to work with the tools took time. However, for the later assignments the experimental group spent slightly less time than the control group. This is in line with Koltovskaia (2020) who found out that newly invented writing assistant tools, which are able to provide the AWE and AWCF features in one integrated application, offer adaptable and time-saving enhancements to the writing curriculum. In fact, utilizing AI-assisted text generators and feedback tools accelerates the process of producing a text. This can be due to the fact that AI-assisted writing tools immediately correct the errors and provide predictions and sentence completions which contribute to the faster flow of the text generation leading to time efficiency. This is further supported by Dale and Viethen (2021), who stated that automatically completing sentences and phrases as well as suggesting alternative wording features are among the greatest developments that AI has brought to the writing domain. Therefore, the hypothesis that the writing improvement in the experimental group may be attributed to the time spent on writing is rejected.

As shown, the control group spent slightly more time producing texts through the pen and paper technique. While it is undeniable that spending time has a positive correlation with improving learning, but the study shows that having helpful language learning tools are more important. Particularly in EFL contexts having efficient materials and language resources are necessary to overcome the challenges mentioned before. It would be interesting to do further research to see whether the same outcome is observed in ESL contexts. It can be concluded that having an effective writing tool, gaining AWCF, benefiting alternative sentence suggestions, and having more exposure to authentic language examples are more effective than the traditional pen-and-paper teacher corrected approaches to writing even if the latter spend more time producing a text. Thus, the prediction three is supported, although the difference in the average time spent is slight.

5.4 The Teacher's and the Students' Attitude Toward AIEd

Regarding the last research question, the related data about attitudes of the teacher in charge of the study as well as the students in the experimental group toward integration of technology

into language learning, was collected by two questionnaires presented in Appendix 2 and 3. Additionally, the result is demonstrated on chapter 4, section 4.6 and 4.7.

In relation to the students' attitude toward AIEd, it was predicted that they would experience a new learning environment which will increase their motivation and engagement. Accordingly, it was discovered that the students found the AI-assisted writing tools extremely helpful, mostly in terms of AWCF related to the errors of mechanics. In addition, providing accurate grammar and spell checking, comprehensive feedback, and alternative sentence suggestions are among the favorable features of AI-assisted writing tools that allow the learners to practice their writing skills.

As discussed before, EFL learners facing challenges such as lack of competent teachers, lack of availability and time of the teachers practicing in the classroom and using insufficient teaching methods which are in line with the findings of Kurniawan and Radia (2017). Such challenges may be overcome as a result of AIEd, such as implementing AI-assisted tools. In line with the findings of Drayton, et.al, (2010) integration of technology into language learning directs the learners toward more self-regulated learning. Similarly, Brown (2002) states that as a result of incorporation of technology into curriculum, learners take responsibility for most of the work that was previously assigned to be done in the classroom. Furthermore, in line with the findings of Qassemzadeh and Soleimani (2016), such features of writing tools give the students a sense of being an autonomous learner and relatively independent when they can evaluate their work with the help of AI-assisted feedback tools. Moreover, it was found out that there were some challenges and concerns using the writing tools such as a few technical bugs appeared sometimes in "Wordtune". Such challenges need further research to know how to assign more effective tools and how to fix the bugs. In addition, it was revealed that the students should be given more instructions to learn how to direct the flow of the text to fulfill the task requirements.

Regarding the teacher's attitude toward AIEd, it was predicted that the teacher would experience a new teaching environment. The result of the questionnaire showed that the teacher found AIEd effective from several aspects which are discussed here. In line with the findings of (Mørch et al., 2005), the teacher found CGF as supplementary feedback to improve the learners' writing skills more significantly. In addition, CGF could save time and the teacher

could have more comprehensive lesson plans devoting time to other skills as well. In addition, the teacher could devote time in the class to discuss the errors the students mostly made and the kinds of feedback they received through AI, and also they could share their experiences.

As elaborated before, EFL contains its own distinct set of difficulties related to possible gaps in syntax, lexis, grammar, pragmatic understanding which make it relatively challenging. In addition, giving effective CF is very demanding for EFL teachers, and researchers debate how to provide learners with efficient CF (Bitchener & Ferris, 2012). Teachers often find it very time-consuming to provide CF on students' writings individually, particularly the large classes with lower levels that are mostly teacher centered. According to studies, when compared to human raters and instructor-provided CF, AWE tools can sometimes offer more detailed and consistent CF (Grimes & Warschauer, 2010; Hussein et.al, 2019). This may be due to the fact that in EFL classes the responsibility of teacher is greater and involves a huge amount of work. Teachers mostly do not devote time to correct all the errors in the writings specially the recurrent errors.

Moreover, the teacher expressed that the learners were more engaged in the lessons and this experience was different from the previous teaching experiences. In line with the findings of some researchers about the integration of technology into education, AIEd resulted in higher motivation and engagement of the learners (Godzicki, 2013; Baytak, et al., 2011; Hennessy, 2005; Arifah, 2014); more learners' responsibilities (Drayton et al., 2010); increased exposure to English in a meaningful authentic context (Warschauer, 2000; Parvin & Salam, 2015; Zhao, 2013).

Furthermore, the teacher expressed some concerns and stated that it is necessary for the teachers to design tasks that implement automatic text generators along with students' effort which is in line with the findings of Godwin-Jones (2021). In other words, students should not rely only on such tools and get distracted from the main objectives of learning. In fact, the use of AI-based writing tools should be integrated within a comprehensive language curriculum with the value of communication at the core (Link et, al. 2020).

5.5 Chapter Summary

This chapter discusses the findings with referring to the research questions, the predictions, and the results. For the first question, it was found out that AI-assisted writing tools have a positive

effect on the EFL learners' writing skills in general. This may be attributed to several reasons mainly because of AWCF that gives the learners opportunity to have more exposure to the language, to observe their errors as they occur, to notice their weak points, to learn from them, and ultimately to improve their writing skills.

In addition, in terms of the second question, it was discovered that although all of the five key aspects of writing improved to some extents in the experimental group, three of them- mechanics, lexis, and grammar- were impacted more than the others. This may be due to AWCF, alternative sentence suggestions, and more exposure to authentic language examples. In addition, organization and task fulfillment were least affected features, which may be due to complexity of human language.

Regarding the third question, it was found out that the factor of time here, in comparison to the use of AI-assisted writing tools has no significant impact on the learners' writing skills development. In other words, the hypothesis that the writing development in the experimental group can be attributed to spending more time on generating text is rejected which confirms that other factors like benefiting from AWCF, having more exposure to English, and using alternative sentences were more effective.

Finally, regarding the last questions, both the teacher and the students had a positive attitude toward AIEd. However, there are some concerns and challenges such as getting dependence on the tools and facing technical bugs.

6 Conclusion

The present study investigated the impact of AIEd, here two AI-based writing tools on EFL learners' writing skills. In this regard, two AI-based writing tools, a collaborative text generator "Wordtune", and an automatic writing evaluator "Insta text" were implemented for the experimental group's home assignments, while the control group followed the traditional handwritten home assignments with teacher corrective feedback. The study lasted for three months, and the participants in both groups took part in the essay writing pretest, 6 periodical tests, and the posttest. In addition, the scores collected from the writing tests were collected and assessed based on the rubrics, considering five key aspects of writings: task fulfillment, organization, grammar, lexis, and mechanics.

As a result of linear mixed-effects model for the total scores, it was found out that the experimental group significantly improved in overall writing skills over the study. Moreover, regarding the five key aspects of writing, it was found out that, primarily the mechanics, which was the weak point of the students, followed by lexis and grammar, noticeably enhanced. Subsequently, organization and task fulfillment slightly improved, although not as significant as the other three aspects.

In addition, the factor of time was examined to see whether the writing's improvement in the experimental group was affected by the amount of time spent per assignment. It was discovered that although both groups started with spending relatively the same amount of time, the experimental group spent slightly less time toward the end of the study. This rejected the effect of time and confirmed that utilizing an effective writing tool, benefiting AWCF, gaining alternative sentence suggestions, and having more exposure to authentic language examples are more effective than the traditional pen-and-paper techniques and teacher corrected approaches to writing, even if they spend more time on producing a piece of text.

Finally, the attitudes of both the teacher in charge of the study and the experimental students toward AIEd were collected through some questionnaires. As a result, they both had a positive attitude and would continue this method. Accordingly, the students mostly benefited from AWCF in terms of mechanics and grammar errors as well as sentence suggestions with a variety of vocabulary. In addition, the teacher benefited AWCF as a complementary assistant to her

own which helped to save the time to devote to other skills as well. There were also some concerns which is elaborated in the next section.

In sum it is concluded that AI-assisted writing tools are beneficial in writing development through the following ways:

- Providing AWCF relating to grammar, spelling, punctuation, and capitalization
- Providing immediate and personalized error corrections
- Enhancing vocabulary by suggesting alternative words and sentences
- Providing suggestions based on the context and topic
- Generating brainstorming ideas by predicting the next sentence to overcome lack of ideas

6.1 Research Implications

The present study is an attempt toward AIEd leading to more self-directed learning specially in the EFL learning context. Therefore, the findings may have several implications for the EFL learners, EFL teachers, language instructors and researchers.

The findings are firstly significant for the EFL learners to benefit AI-assisted tools and become more independent in the process of language learning specially in terms of writing skills. Through applying AI in the process of language learning, especially for writing skills, the learners can accelerate the process by having more exposure to authentic language examples, AWCF, alternative sentence suggestions, and faster pace writing.

In addition, EFL teachers and language instructors can benefit from implementing AI-based language tools in the curriculum and lead the classroom toward more learner centered approaches. By explaining to the learners and training them how to take advantage of technology for the purpose of learning, instructors can contribute to more effective and self-paced learning.

Furthermore, researchers who are interested in AIEd and its role in language learning specially in EFL learners' writing skills can benefit from the findings. Additionally, researchers may get motivated to do further studies related to the integration of technology into language learning.

6.2 Limitations and Future Research Possibilities

The present study has limitations from several aspects that should be considered in future research possibilities. For example, in terms of sample size, this study is limited to 34 learners, 16 in the experimental and 18 in the control group. It is valuable to conduct such study in larger samples to see how the results differ. In addition, the proficiency level of the present sample was intermediate, because there were not enough participants in advanced levels. Therefore, it is worth conducting AIEd in advanced language classes to examine the result.

Furthermore, due to limitations in access in Iran, the present study is limited to two specific AI-assisted writing tools “Wordtune” and “Insta text”. Therefore, the results may vary with greater access to various writing tools. It must be noted that the results may be also influenced by the extent to which the participants are familiar with the technology and how much they are technologically proficient.

Another factor that must be considered is the duration of the present study which is limited to two terms in the language academy which equals three months. Future studies with longer durations are recommended to see the result over the longer span of time, how much the improvement in writing may vary over time.

Additionally, the present study is limited to the investigation of EFL learners’ writing development, which can be expanded to other skills as well. Future studies are recommended to investigate how the other skills like listening, reading, and speaking may be affected due to AI-assisted tools. Finally, it will be interesting if the future studies consider how ESL learners differ from EFL learners in terms of integration of technology into education specially in language learning process.

References

Agélii Genlott, A., & Grönlund, Å. (2013). Improving literacy skills through learning by writing: The iWTR method presented and tested. *Computers & Education*, 67, 8–104.

Ahmadi, M. R. (2018). The use of technology in English language learning: A literature review. *International Journal of Research in English Education (IJREE)*, 3(2), 115-125.

Akbari, Z. (2015). Current challenges in teaching/learning English for EFL learners: The case of junior high school and high school. *Procedia - Social and Behavioral Sciences*, 199, 394-401.

Al Mahmud, F. (2023). Investigating EFL Students' Writing Skills Through Artificial Intelligence: Wordtune Application as a Tool. *Journal of Language Teaching and Research*, 14(5), 1395-1404, https://doi.org/10.17507/jltr.1405.28

Aldeman, N. L. S., Aita, K., Machado, V. P., da Mata Sousa, L. C. D., Coelho, A. G. B., da Silva, A. S., Mendes, A. P. D., Neres, F. J. D., & do Monte, S. J. H. (2021). Smart path (k): A platform for teaching glomerulopathies using machine learning. *BMC Medical Education*, 21(1). https://doi.org/10.1186/s12909-021-02680-1

Alharbi, W. (2023). AI in the Foreign Language Classroom: A Pedagogical Overview of Automated Writing Assistance Tools. *Hindawi Education Research International*, Article ID 4253331. https://doi.org/10.1155/2023/4253331

Aljohani, R. A. (2021). Teachers and students' perceptions on the impact of artificial intelligence on English language learning in Saudi Arabia. *Journal of Applied Linguistics and Language Research*, 8(1), 36–47. https://www.jallr.com/index.php/JALLR/article/download/1156/pdf1156

Alsaleem, B. I. A. (2014). The effect of "WhatsApp" electronic dialogue journaling on improving writing vocabulary word choice and voice of EFL undergraduate Saudi Students. *Harvard: 21st century academic forum conference proceedings*. http://www.readwritethink.org/lesson_images/lesson782/Rubric.pdf

Alsied, S. M., & Ibrahim, N. W. (2017). Exploring challenges encountered by EFL Libyan learners in research teaching and writing. *IAFOR Journal of Language Learning*, 3(2), 143–158. https://doi.org/10.22492/ijll.3.2.06

Annamalai, N., Eltahir, M. E., Zyoud, S. H., Soundrarajan, D., Zakarneh, B., & Al Salhi, N. R. (2023). Exploring English language learning via Chabot: A case study from a self-determination theory perspective. *Computers and Education: Artificial Intelligence*, 5(100148). https://doi.org/10.1016/j.caeai.2023.100148

Annamalai, N., Rashid, R. A., Hashmi, U. M., Muhamed, M., Alqaryuoti, M, H., & Sadeq, A. E. (2023). Using chatbots for English language learning in higher education. *Computers and Education: Artificial Intelligence*, 5(100153). https://doi.org/10.1016/j.caeai.2023.100153

Arifah, A. (2014). *Study on the use of technology in ELT classroom: Teachers' perspective.* [M.A. Thesis, Department of English and Humanities, BRAC University].

Attali, Y. (2013). *Validity and reliability of automated essay scoring*. In M. D. Shermis, & J. Burstein (Eds.), Handbook on automated essay evaluation: Current applications and new directions (pp. 181–198). Routledge.

Attali, Y. & Burstein, J. (2006). Automated Essay Scoring With e-rater. *Journal of Technology, Learning, and Assessment*, 4(3). http://www.jtla.org

Ayedoun, E., Hayashi, Y., & Seta, K. (2019). Adding communicative and affective strategies to an embodied conversational agent to enhance second language learners' willingness to communicate. *International Journal of Artificial Intelligence in Education*, 29, 29–57.

Azah, D. N. (2019). *The effectiveness of Grammarly checker toward student's writing quality of the fourth semester of English Department at IAIN Tulungagung* [Unpublished undergraduate thesis]. IAIN Tulungagung.

Bates, D., Maechler, M., Bolker, B., and Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67, 1–48.

Barani, G. (2011). The relationship between Computer Assisted Language Learning (CALL) and listening skill of Iranian EFL learners. *Procedia - Social and Behavioral Sciences*, 15, 4059–4063. https://doi. org/10.1016/j.sbspro.2011.04.414.

Barrot, J. S. (2022). Integrating technology into ESL/EFL writing through Grammarly. *RELC Journal*. 53(3), 764–768. https://doi.org/10.1177/0033688220966632

Baytak, A., Tarman, B., & Ayas, C. (2011). Experiencing technology integration in education: children's perceptions. *International Electronic Journal of Elementary Education*, 3(2), 139-151. https://www.iejee.com/index.php/IEJEE/article/view/233

Bellod, H. C., Ramon, V. B., Fernandez, E. C., & Lujan, J. F. G. (2021). Analysis of stress and academic-sports commitment through self-organizing artificial neural networks. *Challenges*, 42, 136–144. https://doi.org/10.47197/RETOS.V42I0.86983

Biggs, J. (1996). Enhancing teaching through constructive alignment. *Higher Education*, 32, 347- 364.

Bitchener, J., & Ferris, D.R. (2012). *Written Corrective Feedback in Second Language Acquisition and Writing*. Routledge.

Bitchener, J., & Basturkmen, H. (2006). Perceptions of the difficulties of postgraduate L2 thesis students writing the discussion section. *Journal of English for Academic Purposes*, 5(1), 4–18. https://doi.org/10.1016/j.jeap.2005.10.002

Black, P., & Wiliam, D. (2009). Developing the theory of formative assessment. *Educational Assessment, Evaluation and Accountability*, 21(1), 5–31. doi: http://dx.doi.org/10.1007/s11092-008-9068-5

Boninger, F., Molnar, A., & Saldaña, C. (2020). Big Claims, Little Evidence, Lots of Money: The Reality Behind the Summit Learning Program and the Push to Adopt Digital Personalized Learning Platforms. *Boulder, CO: National Education Policy Center*. http://nepc.colorado.edu/publication/summit-2020

Brown, H. (2002). *Strategies for Success. A practical guide to learning English*. New York: Addison Wesley Longman Inc.

Burstein, J., Tetreault, J., & Madnani, N. (2013). *The e-rater® automated essay scoring system*. In M. D. Shermis, & J. Burstein (Eds.), Handbook of Automated Essay Evaluation (pp. 55–67). Routledge. https://doi.org/10.4324/9780203122761

Casal, J. E., & Kessler, M. (2023). Can linguists distinguish between ChatGPT/AI and human writing?: A study of research ethics and academic publishing. *Research Methods in Applied Linguistics*, 2 (3). https://doi.org/10.1016/j.rmal.2023.100068

Chamot, A.U., (2005). Language learning strategy instruction: Current issues and research. *Annual Review of Applied Linguistics*.25,112-30. DOI: 10.1017/S0267190505000061

Chappelle, C. A. (2001). *Computer applications in second language acquisition: Foundations for teaching, testing and research*. Cambridge University Press.

Chiu, T. K. F. (2021). A holistic approach to Artificial Intelligence (AI) curriculum for K-12 schools. *Tech Trends*, 65, 796–807. https://doi.org/10.1007/s11528-021-00637-1

Chiu, T. K. F., Meng, H., Chai, C. S., King, I., Wong, S., & Yeung, Y. (2022). Creation and evaluation of a pre-tertiary Artificial Intelligence (AI) curriculum. *IEEE Transactions on Education*, 65(1), 30–39. https://doi.org/10.1109/TE.2021.3085878

Chiu T. K. F., Xia, Q., Zhou, X., Chai, C. S., & Cheng, M. (2023). Systematic literature review on opportunities, challenges, and future research recommendations of artificial intelligence in education. *Computers and Education: Artificial Intelligence*, 4, 100118. https://doi.org/10.1016/j.caeai.2022.100118.

Choo, Y. B., & Li, K. L. (2017). Digital writing in English language writing instruction. *ARIEL-An International Research Journal of English Language and Literature*, 28, 1–16. https://sujo.usindh.edu.pk/index.php/ARIEL/article/view/19/16

Coenen, A., Davis, L., Ippolito, D., Reif, E., & Yuan, A. (2021). Wordcraft: A human-AI collaborative editor for story writing. *arXiv*. https://doi.org/10.48550/arXiv.2107.07430

Cotos, E. (2014). Genre-based automated writing evaluation for L2 research writing: From design to evaluation and enhancement. *Palgrave Macmillan*. DOI: 10.1057/9781137333377

Dale. R., & Viethen, J. (2021). The automated writing assistance landscape in 2021. *Natural Language Engineering*, 27(4), 511–518.

Daniels, P., & Leslie, D. (2013). Grammar software ready for EFL writers? *On Cue Journal*, 9(4), 391–401.

Dickson, B. (2017, November 20). How artificial intelligence is shaping the future of education. *PC Magazine*. https://www.pcmag.com/news/how-artificial-intelligence-is-shaping-the-future-of-education

Dikli, S. (2006). An overview of automated scoring of essays. *The Journal of Technology, Learning and Assessment*, 5(1). https://ejournals.bc.edu/index.php/jtla/article/view/1640

Dignen, B. (2014). Five reasons why feedback may be the most important skills. Available at: http://www.cambridge.org/elt/blog/2014/03/fivereasons-feedback-may-important-skill/

Dizon, G., & Gayed, J. M. (2021). Examining the impact of Grammarly on the quality of mobile L2 writing. *The JALT CALL Journal*, 17(2), 74–92.

Dodigovic, M., & Tovmasyan, A. (2021). Automated writing evaluation: the accuracy of Grammarly's feedback on form. *International Journal of TESOL Studies*, 3(2), 71–88.

Dong, Y., & Shi, L. (2021). Using Grammarly to support students' source-based writing practices. *Assessing Writing*, 50 (100564).

Drayton, B., Falk, J. K., Stroud, R., Hobbs, K., & Hammerman, J. (2010). After installation: Ubiquitous computing and high school science in three experienced, high-technology schools. *Journal of Technology, Learning, and Assessment*, 9(3), 1-57. https://eric.ed.gov/?id=EJ873677

Edwards-Groves, C, J. (2011). The multimodal writing process: changing practices in contemporary classrooms. *Language and Education*, 25(1), 49–64. DOI: 10.1080/09500782.2010.523468

Ellis, R. (1994). *The study of second language acquisition*. Oxford University Press.

Ellis, R. (2008). Corrective feedback and teacher development. *L2 Journal*, 1, 3-18.

Engeness, I. Mørch, A. (2016). Developing Writing Skills in English Using Content-Specific Computer-Generated Feedback with EssayCritic. *Nordic Journal of Digital Literacy*, 11, (2), 118–135. DOI: 10.18261/issn.1891-943x-2016-02-03

Fageeh, A. I. (2011). EFL learners' use of blogging for developing writing skills and enhancing attitudes towards English learning: An exploratory study. *Journal of Language and Literature*, 2(1), 31–48.

Ferreira, A. (2006). Estrategias efectivas defeedback positivo y correctivo en el español como lengua extranjera [Effective positive and corrective feedback strategies in Spanish as a foreign language]. *Signos*, *39*(62), 379-406.

Finn, H. B. (2018). Articulating struggle: ESL students' perceived obstacles to success in a community college writing class. *Journal of Second Language Writing*, 42, 101–106. https://doi.org/10.1016/j.jslw.2018.09.001

Floridi, L. & Chiriatti, M. (2020). GPT-3: its nature, scope, limits, and consequences. *Minds and Machines*, 30, 681– 694.

Flowerdew, J. (2019). The linguistic disadvantage of scholars who write in English as an additional language: Myth or reality. *Language Teaching*, 52(2), 249–260. https://doi.org/10.1017/S0261444819000041

Gan, Z., Humphreys, G., & Hamp-Lyons. (2004). Understanding successful and unsuccessful EFL Students in Chinese universities. *The Modern Language Journal*, 88, 229-44.

Gerami, M. H., & Baighlou, S. M. G. (2011). Language learning strategies used by successful and unsuccessful Iranian EFL students. *Procedia-Social and Behavioral Sciences*, 29, 1567-1576. DOI: 10.1016/j.sbspro.2011.11.399

Ghufron, M. A. (2019). Exploring an automated feedback program 'Grammarly' and teacher corrective feedback in EFL writing assessment: modern vs. traditional assessment. The

3rd English Language and Literature International Conference, *ELLiC*, Semarang, Indonesia.

Ghufron, M. A., & Rosyida, F. (2018). The role of Grammarly in assessing English as a Foreign Language (EFL) writing. *Lingua Cultura*, 12(4), 395-403. https://doi.org/10.21512/lc.v12i4.4582

Gilakjani, A. P., & Sabouri, N. B. (2016). Learners' listening comprehension difficulties in English language learning: A literature review. *English Language Teaching*, 9(6), 123-133.

Goda, Y., Yamada, M., Matsukawa, H., Hata, K., & Yasunami, S. (2014). Conversation with a chatbot before an online EFL group discussion and the effects on critical thinking. *Journal of Information Systems Education*, 13(1), 1–7.

Godwin-Jones, R. (2021). Big data and language learning: opportunities and challenges, *Language Learning & Technology*, 25(1), 4–19.

Godzicki, L., Godzicki, N., Krofel, M., & Michaels, R. (2013). *Increasing motivation and engagement in elementary and middle school students through technology-supported learning environment*. [M.A thesis, Saint Xavier University]. http://www.eric.ed.gov.ezproxy.cu

Goldin, I. M., & Ashley, K. D. (2012). Eliciting formative assessment in peer review. *Journal of Writing Research*, 4(2), 203–237.

Gonzalez-Calatayud, ´ V., Prendes-Espinosa, P., & Roig-Vila, R. (2021). Artificial intelligence for student assessment: A systematic review. *Applied Sciences*, 11(12), 5467. https://doi.org/10.3390/app11125467

Graves, D. H. (1982). *A Case Study Observing the Development of Primary Children's Composing, Spelling, and Motor Behaviors during the Writing Process*. Writing Process Laboratory, Department of Education, University of New Hampshire. http://files.eric.ed.gov/fulltext/ ED218653.pdf

Grimes, D., & Warschauer, M. (2010). Utility in a fallible tool: a multi-site case study of automated writing evaluation. *Journal of Technology, Language, and Assessment*, 8 (6), 1–43.

Gupta, K. P., & Bhaskar, P. (2020). Inhibiting and motivating factors influencing teachers' adoption of AI-based teaching and learning solutions: Prioritization using analytic hierarchy process. *Journal of Information Technology Education: Research,* 19, 693–723. https://doi.org/10.28945/4640

Hamzaoui, C. (2021). Scrutinizing Algerian EFL students' challenges in research teaching and writing. *Hungarian Educational Research Journal*, 11(4), 478–488. https://doi.org/10.1556/063.2021.00047

Hanauer, D. I., Sheridan, C. L., & Englander, K. (2019). Linguistic injustice in the writing of research articles in English as a second language: Data from Taiwanese and Mexican researchers. *Written Communication*, 36(1), 136–154. https://doi.org/10.1177/0741088318804821

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research,* 77(1), 81–112. Doi: http://dx.doi.org/10.3102/003465430298487.

Hennessy, S. (2005). *Emerging teacher strategies for supporting*. University of Cambridge.

Holmes, W., Hui, Z., Miao, F., & Ronghuai, H. (2021). *AI and education: A guidance for policymakers*. UNESCO Publishing.

Hsu, W.-H. (2015). Transitioning to a communication-oriented pedagogy: Taiwanese university freshmen's views on class participation. *System*, 49, 61-72. https://doi.org/10.1016/j.system.2014.12.002

Huang, Y., & Wilson, J. (2021). Using automated feedback to develop writing proficiency. *Computers and Composition*, 62, 102675.

Huang, J., Shen, G., & Ren, X. P. (2021). Connotation analysis and paradigm shift of teaching design under artificial intelligence technology. *International Journal of Emerging Technologies in Learning*, 16(5), 73–86. https://doi.org/10.3991/ijet.

Hussein, M. A., Hassan. H, and Nassef, M. (2019). Automated language essay scoring systems: a literature review. *PeerJ Computer Science*, 5, ID e208.

Jarke, J., & Macgilchrist, F. (2021). Dashboard stories: How narratives told by predictive analytics reconfigure roles, risk and sociality in education. *Big Data and Society*, 8(1). https://doi.org/10.1177/20539517211025561

Jiang, L., Yu, S., & Wang, C. (2020). Second language writing instructors' feedback practice in response to automated writing evaluation: a sociocultural perspective. *System*, 93, 102302.

Kabilan, M. K., Ahmad, N., & Abidin, M. J. Z. (2010). Facebook: An online environment for learning of English in institutions of higher education?. *Internet and Higher Education*, 13(4), 179–187. https://doi. org/10.1016/j.iheduc.2010.07.003.

Kachru, B. B. (1992). Teaching world Englishes. *The other tongue: English across cultures*, 2, 355-366.

Kangasharju, A., Ilomäki, L., Lakkala, M., Toom, A. (2022). Lower secondary students' poetry writing with the AI-based Poetry Machine. *Computers and Education: Artificial Intelligence*, 3, 100048.

Karim, K., & Nassaji, H. (2020). The revision and transfer effects of direct and indirect comprehensive corrective feedback on ESL students' writing. *Language Teaching Research*, 24(4), 519–539. https://doi.org/10.1177/1362168818802469

Karyuatry, L. (2018). Grammarly as a tool to improve students' writing quality: Free online-proof reader across the boundaries. *JSSH (Jurnal Sains Sosial dan Humaniora)*, 2(1), 83–89. https://doi.org/10.30595/jssh.v2i1.2297

Kasirzadeh & Gabriel, I. (2022). In conversation with Artificial Intelligence: aligning language models with human values. *Philos. Technol*, 36(27). https://doi.org/10.1007/s13347-023-00606-x

Kim, N.-Y. (2018). Chatbots and Korean EFL students' English vocabulary learning. *Journal of Digital Convergence*, 16(2), 1–7. https://doi.org/10.14400/

Kirkpatrick, A. (2014). World Englishes. *In The Routledge companion to English studies* (pp. 63-75). Routledge.

Koltovskaia, S. (2020). Student engagement with automated written corrective feedback (AWCF) provided by Grammarly: A multiple case study. *Assessing Writing*, 44, 100450. 10.1016/j.asw.2020.100450

Komba, S. C. (2015). Challenges of writing theses and dissertations among postgraduate students in Tanzanian higher learning institutions. *International Journal of Research Studies in Education*, 5(3), 71–80. https://doi.org/10.5861/ijrse.2015.1280

Krashen, S.D. (1985). *The input hypothesis: issues and implications*[M]. Longman.

Kress, G. (2010). *Multimodality: A social semiotic approach to contemporary communication*. Routledge.

Kukich, K. (2000). Beyond automated essay scoring. *IEEE intelligent systems and their applications*, 15(5), 22–27. http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=889104

Kurniawan, M., & Radia, E. H. (2017). *A Situational Analysis of English Language Learning among Eastern Indonesian Students*. Advances in Social Science, Education, and Humanities Research (ASSEHR) (pp. 1-6). Atlantis Press. DOI: 10.2991/yicemap-17.2017.1

Kuteeva, M. (2011). Wikis and academic writing: Changing the writer-reader relationship. *English for Specific Purposes*, 30(1), 44–57. https://doi.org/10.1016/j. esp.2010.04.007

Laflen, A. (2020). Using Eli review as a strategy for feedback in online courses. *Assessing Writing*, 46, Article 100486.

Lahuerta, A. (2017). Analysis of accuracy in the writing of EFL Students enrolled on CLIL and Non-CLIL programmes: The impact of grade and gender. *The Language Learning Journal*, 48(2), 121–132. https://doi.org/10.1080/09571736.2017.1303745

Lan, R. & Oxford, R.L. (2003). Language learning strategy profiles of elementary school learners in Taiwan. *International Review of Applied Linguistics in Language Teaching*, 41, 339-379.

Lee, I. (2007). Assessment for learning: Integrating assessment, teaching, and learning in the ESL/EFL writing classroom. *Canadian Modern Language Review/La Revue canadienne des langues vivantes*, 64(1), 199–213. DOI: http://dx.doi.org/10.3138/cmlr.64.1.199

Lee, C., Cheung, W. K. W., Wong, K. C. K., & Lee, F. S. L. (2013). Immediate web-based essay critiquing system feedback and teacher follow-up feedback on young second language learners' writings: an experimental study in a Hong Kong secondary school. *Computer Assisted Language Learning*, 26(1), 39–60. DOI: http://dx.doi.org/10.1080/09588221.2011.630672

Li, A. W. (2023). Using Peerceptiv to support AI-based online writing assessment across the disciplines. *Assessing Writing*, 57 (100746).

Liberg, C. (2007). Språk och kommunikation. In A. Ewald, & B. Garme (Ed.), *Attläsa och skriva – forskning och beprövad erfarenhet* (p. 7–23). Skolverket.

Lin, M. P. C., & Chang, D. (2020). Enhancing post-secondary writers' writing skills with a chatbot. *Journal of Educational Technology & Society*, 23(1), 78–92.

Lin, W., & Yang, S. (2011). Exploring students' perceptions of integrating Wiki technology and peer feedback into English writing courses. *English Teaching: Practice and Critique*, 10(2), 88-103. https://eric.ed.gov/?id=EJ944900

Lin, L. H. F., & Morrison, B. (2021). Challenges in academic writing: Perspectives of engineering faculty and L2 postgraduate research students. *English for Specific Purposes*, 63, 59–70. https://doi.org/10.1016/j.esp.2021.03.004

Link, S., Mehrzad, M., and Rahimi, M. (2022). Impact of automated writing evaluation on teacher feedback, student revision, and writing improvement. *Computer Assisted Language Learning*, 35 (4), 605–634.

Liu, D., Ahn, G.-S., Baek, K.-S., & Han, N.-O. (2004). South Korean high school English teachers' code switching: Questions and challenges in the drive for maximal use of English in teaching. *TESOL Quarterly*, 38(4), 605–638. https://doi.org/10.2307/3588282

Luckin, R. (2017). Toward artificial intelligence-based assessment systems. *Nature Human Behavior*, 1(3), 0028. DOI:10.1038/s41562-016-0028

Mahammoda, S. A. (2016). Factors affecting the quality of undergraduate research work in Bahir Dar University, Ethiopia. *International Journal of Innovative Research and Development*, 5(12), 23–27. https://www.internationaljournalcorner.com/index.php/ijird_ojs/article/view/136671/95794

McCarthy, J., Minsky, M. L., Rochester, N., & Shannon, C. E. (2006). A proposal for the Dartmouth Summer Research Project on artificial intelligence, August 31, 1955. *AI Magazine*, 27(4), 12–14. https://doi.org/10.1609/aimag.v27i4.1904

McDonough, K., & Crawford, W. J. (2018). Identifying effective writing tasks for use in EFL write-to-learn language contexts. *The Language Learning Journal*, 48(4), 469–480. https://doi.org/10.1080/09571736.2018.1465990

Melor M. Y., & Salehi, H. (2012). The effectiveness of Facebook groups on teaching and improving writing: Students' perceptions. *International Journal of Educational and Information Technologies*, 6(1), 87–96. https://doi.org/10.5539/elt.v6n7p1

Mercer, S., & Ryan, S. (2010). A mindset for EFL: Learners' beliefs about the role of natural talent. *ELT Journal*, 64(4), 436–444. https://doi.org/10.1093/elt/ccp083

Mizumoto, A. & Eguchi, M. (2023). Exploring the potential of using an AI language model for automated essay scoring. *Research Methods in Applied Linguistics*, 2 (2),100050.

Mouza, C. (2008). Learning with laptops: Implementation and outcomes in an urban, underprivileged school. *Journal of Research on Technology in Education*, 40(4), 447 - 472. https://eric.ed.gov/?id=EJ826086

Mørch, A., Cheung, W., Wong, K., Liu, J., Lee, C., Lam, M., & Tang, J. (2005). Grounding Collaborative Knowledge Building in Semantics-Based Critiquing. In R. H. Lau, Q. Li, R. Cheung, & W. Liu (Eds.), *Advances in Web-Based Learning – ICWL*. (Vol. 3583, pp. 244–255): Springer Berlin Heidelberg.

Mørch, A., Irina, E. (2016). Developing Writing Skills in English Using Content-Specific Computer-Generated Feedback with EssayCritic. *Nordic Journal of Digital Literacy*, 11(2), 118–135. DOI: 10.18261/issn.1891-943x2016-02

Muñoz, C. (2014). Contrasting effects of starting age and input on the oral performance of foreign language learners. *Applied Linguistics*, 35(4), 463–482.

https://doi.org/10.1093/applin/amu024

Murray, D. (1972). Teach writing as a process not product. *The Leaflet*, 71(3), 11–14. http://www.larue.k12.ky.us/userfiles/1085/

Naba'h, A. A., Hussain, J., Al-omari, A., & Shdeifat, S. (2009). The effect of computer assisted language learning in teaching english grammar on the achievement of secondary students in Jordan. *The International Arab Journal of Information Technology*, 6(4), 431–439.

Nobles, S.L., & Paganucci, L. (2015). Do Digital Writing Tools Deliver?. *Computers and Composition*, 38, A, 16-31. https://doi.org/10.1016/j.compcom.2015.09.001

Nova, M. (2018). "Utilizing Grammarly in evaluating academic writing: narrative research on EFL students' experience. *Premise: Journal of English Education and Applied Linguistics*, 7(1), 80–96.

O'Neill, R., & Russell, A. (2019). Stop! Grammar time: University students' perceptions of the automated feedback program Grammarly. *Australasian Journal of Educational Technology*, 35(1), 42–56.

O'Neill, R., & Russell, A. M.T. (2020). Grammarly: help or hindrance? Academic learning advisors' perceptions of an online grammar checker. *Journal of Academic Language and Learning*, 13 (1), A88–A107.

Ouyang, F., & Jiao, P. (2021). Artificial intelligence in education: The three paradigms. *Computers & Education: Artificial Intelligence*, 2, 100020. https://doi.org/10.1016/j.caeai.2021.100020

Page, E. B. (1967). The imminence of grading essays by computer. *Phi Delta Kappan*, 48, 238–243.

Parvin, R. H., & Salam, S. F. (2015). The effectiveness of using technology in English language classrooms in government primary schools in Bangladesh. *FIRE: Forum for International Research in Education*, 2(1), 47 - 59. http://preserve.lehigh.edu/fire/vol2/iss1/5

Phuong, W. T. N. (2021). Difficulties in studying writing of English-majored sophomores at a university in Vietnam. *European Journal of Education Studies*, 8(10), 313–330. https://doi.org/10.46827/ejes.v8i10.3962

Qassemzadeh, A., & Soleimani, H. (2016). The impact of feedback provision by grammarly software and teachers on learning passive structures by Iranian EFL learners. *Theory and Practice in Language Studies*, 6(9), 1884-1894. https://doi.org/10.17507/tpls.0609.23

Qinghua, Y., & Satar, M. (2020). English as a foreign language learner interaction with chatbots: Negotiation for meaning. *International Online Journal of Education and Teaching (IOJET)*, 7(2), 390–410. http://iojet.org/index.php/IOJET/article/vie w/707

Rad, H. S., Alipour, R., & Jafarpour, A. (2023). Using artificial intelligence to foster students' writing feedback literacy, engagement, and outcome: A case of Wordtune application. *Interactive Learning Environments*, 1–21. https://doi.org/10.1080/10494820.2023.2208170

Ranalli, J. (2013). Online strategy instruction for integrating dictionary skills and language awareness. *Language Learning & Technology*, 17(2), 75–99.

https://www.lltjournal.org/item/10125-44325/

Rapanta, C., & Walton, D. (2016). The use of argument maps as an assessment tool in higher education. *International Journal of Educational Research*, 79, 211–221. https://doi.org/10.1016/j.ijer.2016.03.002

Razak, N. A., Saeed, M., & Ahmad, Z. (2013). Adopting social networking sites (SNSs) as interactive communities among English foreign language (EFL) learners in writing: Opportunities and challenges. *English Language Teaching*, 6(11), 187–198. https://doi.org/10.5539/elt.v6n11p187

Richards, J., C., & Schmidt, R. (2010). *Longman dictionary language teaching & applied linguistics*. Great Britain: Pearson education limited.

Rodinadze, S., & Zarbazoia, K. (2012). The advantages of information technology in teaching English language. *Frontiers of Language and Teaching*, 3(5), 271-275.

Rosário, P., Högemann, J., Núñez, J. C., Vallejo, G., Cunha, J., Rodríguez, C., & Fuentes, S. (2019). The impact of three types of writing intervention on students' writing quality. *PLoS One*, 14(7), e0218099. https://doi.org/10.1371/journal.pone.0218099

Ruan, S., Willis, A., Xu, Q., Davis, G. M., Jiang, L., Brunskill, E., & Landay, J. A. (2019). BookBuddy: Turning digital materials into interactive foreign language lessons through a voice chatbot. *Proceedings of the Sixth (2019) ACM Conference on Learning Scale* (pp. 1-4). USA. https://doi.org/10.1145/3330430.3333643

Ruscetti, T., Krueger, K., & Sabatier, C. (2018). Improving quantitative writing one sentence at a time. *PLoS One*, 13(9), e0203109. https://doi.org/10.1371/journal.pone.0203109

Sabzian, F., Pourhossein Gilakjani, A., & Sodouri, S. (2013). Use of technology in classroom for professional development. *Journal of Language Teaching and Research*, 4(4), 684-692. DOI:10.4304/jltr.4.4.684-692

Saadi, Z. K., & Saadat, M. (2015). EFL learners' writing accuracy: Effects of direct and metalinguistic electronic feedback. *Theory and Practice in Language Studies*, 5(10), 2053–2063. DOI:10.17507/tpls.0510.11

Singh, M. K. M. (2017). International EFL/ESL master students' adaptation strategies for academic writing practices at tertiary level. *Journal of International Students*, 7(3), 620–643. https://doi.org/10.5281/zenodo.570025

Sireci, S. G., & Rizavi, S. (2000). *Comparing Computerized and Human Scoring of Students' Essays*. http://files.eric.ed.gov/fulltext/ED463324.pdf

Soori, A., Kafipour, R., & Soury, M. (2011). Effectiveness of different types of direct corrective feedback on correct use of English articles among the Iranian EFL students. *European Journal of Social Sciences*, 26(4), 494-501.

Stern, H. H. (1983). *Fundamental Concepts of Language Teaching*. Oxford University Press.

Tegos, S., Demetriadis, S., & Tsiatsos, T. (2014). A configurable conversational agent to trigger students' productive dialogue: A pilot study in the call domain. *International Journal of Artificial Intelligence in Education*, 24, 62–91. https://link.springer.com/article/10.1007/s40593-013-0007-3?no-access=true

Thompson, I. (2013). The mediation of learning in the zone of proximal development through a co-constructed writing activity. *Research in the Teaching of English*, 47(3), 247–276. http://www.ncte.org/library/nctefiles/resources/journals/rte/0473feb2013/rte0473mediation.pdf

Thompson, A., Gallacher, A., & Howarth, M. (2018). Stimulating Task Interest: Human Partners or AI? *Research presentation at EUROCALL International Convention*, Jyväskylä, Finland. DOI:10.14705/rpnet.2018.26.854

Ur, P. (2006). *A course in language teaching: Practice and theory*. Cambridge University Press.

Vinall, K., & Hellmich, E. A. (2022). Do you speak translate? Reflections on the nature and role of translation. *L2 Journal*, 14(1), 4–25.

Warschauer, M. (2000). The death of cyberspace and the rebirth of CALL. *English Teachers' Journal*, 53, 61 –67. http://www.gse.uci.edu/markw/cyberspace.html

Wegerif, R. (2004). The role of educational software as a support for teaching and learning conversations. *Computers & Education*, 43(1–2), 179–191. https://doi.org/10.1016/j.compedu.2003.12.012

Weragama, D., & Reye, J. (2014). Analysing student programs in the PHP intelligent tutoring system. *International Journal of Artificial Intelligence in Education*, 24(2), 162–188. https://doi.org/10.1007/s40593-014-0014-z

Wibowo, S., & Zhai, C. (2023). A systematic review on artificial intelligence dialogue systems for enhancing English as foreign language students' interactional competence in the university. *Computers and Education: Artificial Intelligence*, 4, 100134.

Williamson, B., & Eynon, R. (2020). Historical threads, missing links, and future directions in AI in education. *Learning, Media and Technology*, 45(3), 223–235. https://doi.org/10.1080/17439884.2020.1798995

Winerip, M. (2012). Facing a robo-grader? just keep obfuscating mellifluously. *New York Times*, 22. http://thelawsofrobotics2013.iankerr.ca/files/2013/09/49- Facing-a-RoboGrader.pdf

Xia, Q., Chiu, T. K. F, Lee, M., Temitayo I., Dai, Y., & Chai, C.S. (2022). A Self-determination theory design approach for inclusive and diverse Artificial Intelligence (AI) K-12 education. *Computers & Education*, 189, 104582. DOI: 10.1016/j.compedu.2022.104582

Xu, L., Sanders, L., Li, K., & Chow, J. C. (2021). Chatbot for health care and oncology applications using artificial intelligence and machine learning: Systematic review. *JMIR cancer*, 7(4), e27850. https://doi.org/10.2196/27850

Xu, J., & Zhang, S. (2021). Understanding AWE feedback and English writing of learners with different proficiency levels in an EFL classroom: A sociocultural perspective. *The Asia-Pacific Education Researcher*, 31, 357-367. https://doi.org/10.1007/s40299-021-00577-7

Yunus, M., Salehi, H., Sun, C. H. U. I., Yong, J., Yen, P., Kwan, L., & Li, S. U. (2011). Using Facebook groups in teaching ESL writing. *Recent Researches in Chemistry, Biology, Environment and Culture*, 75–80.

Yunus, M. M., Salehi, H., & Chenzi, C. (2012). Integrating social networking tools into ESL writing classroom: Strengths and weaknesses. *English Language Teaching*, 5(8), 42–48. 10.5539/elt.v5n8p42

Zhai, N., & Ma, X. (2021). Automated writing evaluation (AWE) feedback: A systematic investigation of college students' acceptance. *Computer Assisted Language Learning*. 35(4):1-26. DOI: 10.1080/09588221.2021.1897019

Zhang, T. (2018). *The Effect of Focused Versus Unfocused Written Corrective Feedback on the Development of University-Level Learners' Explicit and Implicit Knowledge in an EFL Context*. [Unpublished doctoral thesis, The University of Sydney].

Zhang, Z., & Zhang, Y. (2018). Automated writing evaluation system: Tapping its potential for learner engagement. *IEEE Engineering Management Review*, 46(3), 29–33. https://doi.org/ 10.1109/emr.2018.2866150

Zhao, Z. (2001). Reflections on the Crux of English Teaching in China[J]. *Foreign Language and Their Teaching*, 10, 38-39.

Zhao, Y. (2013). Recent developments in technology and language learning: Literature review and meta-analysis. *CALICO Journal*, 21(1), 7 -27. https://eric.ed.gov/?id=EJ674877

Zhao, X. (2022). Leveraging artificial intelligence (AI) technology for English writing: Introducing Wordtune as a digital writing assistant for EFL writers. *RELC Journal*. https://doi.org/10.1177/00336882221094089

Zhongde, Zhao. (2001). Reflections on the Crux of English Teaching in China[J]. *Foreign Language and Their Teaching*, 10, 38-39.

Åkerfeldt, A. (2014). Re-shaping of writing in the digital age: A study of pupils' writing with different resources. Universitets forlaget. *Nordic Journal of Digital Literacy*, 9, 172–193.

Appendices

Appendix 1: Letter of Information and Consent Form

Are you interested in taking part in the research project

"The effect of AI-based applications on the EFL (English as a Foreign Language) learners' writing skill"?

Dear participants

Please read the information provided below about the experiment.

You are invited to participate in a research project where the main purpose is to investigate the effect of AI-based applications on the EFL students' writing skills. In this study the main objective is to investigate how two AI-based tools ("Wordtune" and "Insta text") can be used in teaching and learning English as a second language.

The study will be published as a master's thesis from the Arctic University of Norway, UiT, and all data will be handled by the university of Tromsø. We ask you to participate in the study since you are a second learner of English who take part in a language course.

If you chose to take part in the project, this will involve that you write essays about the topics written in the course book. It will take approx. 40 minutes every session. Your writings will be stored and graded for the purpose of this project.

Participation in the project is voluntary. If you chose to participate, you can withdraw your consent at any time without giving a reason. All information about you will then be made anonymous. There will be no negative consequences for you if you chose not to participate or later decide to withdraw.

It will not affect your grades in the academy. In fact, this project is just an effort towards improving your language skills, if you withdraw at any point there will not be any negative consequences.

We will only use your personal data for the purpose(s) specified here and we will process your personal data in accordance with data protection legislation (the GDPR).

- Only the first author of the study, Fatemah Etaat, will have access to your personal results of the study.

- Your name and contact details will be replaced with a code. The list of names, contact details and respective codes will be stored separately from the rest of the collected data, and will be stored the data on a hard disk, locked away.

The participants will not be recognized in any publications.

The planned end date of the project is 1st September. Any personal data, including names and email addresses if the participants will be deleted and all the writings will be anonymised at the end of the project.

So long as you can be identified in the collected data, you have the right to:

- access the personal data that is being processed about you

- request that your personal data is deleted

- request that incorrect personal data about you is corrected/rectified

- receive a copy of your personal data (data portability), and

- send a complaint to the Norwegian Data Protection Authority regarding the processing of your personal data

We will process your personal data based on your consent.

Based on an agreement with UiT, The Data Protection Services of Sikt – Norwegian Agency for Shared Services in Education and Research has assessed that the processing of personal data in this project meets requirements in data protection legislation.

If you have questions about the project, or want to exercise your rights, contact:

- UiT via Fatemeh Etaat/Björn Lundquist (bjorn.lundquist@uit.n).

- Our Data Protection Officer at the University of Norway: Joakim Bakkevold (joakim.bakkevold@uit.no)

If you have questions about how data protection has been assessed in this project by Sikt, contact:

- email: (personverntjenester@sikt.no) or by telephone: +47 73 98 40 40.

Yours sincerely,

Fatemah Etaat

Consent form

I have received and understood information about the project [**The effect of AI-based applications on the EFL learners' writing skill**] and have been given the opportunity to ask questions. I give consent:

- ☐ to participate in the study
- ☐ for Fatemeh Etaat to give information about me to this project

I give consent for my personal data to be processed until the end of the project.

(Signed by participant, date)

Appendix 2: Students' Attitude Questionnaire

Name: ……..., ***Age***: ……, ***English level***…….

A. Using AI-powered learning tools

2) Before this course, did you use any AI-powered tools to assist you in the process of English learning, especially for writing skills?

a) Yes

b) No

B. Any particular tool

3) If yes, which AI-powered tool did you use?

a) Insta text

b) Wordtune

c) ChatGPT

d) Grammarly

e) Other (please specify)

C. Frequency of usage

4) How often did you use it?

a) Daily

b) Weekly

c) Monthly

d) Rarely

D. Improvement of English skills

5) Did it help you to develop your English skills?

a) Yes

b) No

c) Not sure

E. Particular ways of improvement

6) If yes, in what ways did it benefit you?

a) By providing automated written feedback on my grammar errors

b) By providing automated written feedback on punctuation, capitalization, and spelling errors

c) By suggesting alternative sentences

d) By making the text production faster

e) By encouraging me to write more and practice my writing skills

f) By being available at all time

g) Other (please specify)

F. Preferred characteristics

7) What impressed you the most about the applications?

a) Having straight-forward design

b) helpful writing recommendations

c) Accurate grammar, spelling, capitalization, and punctuation checks

d) detailed feedback

e) Other (please specify)

G. Unliked characteristics

8) Which aspect of the applications did you find challenging?

a) Limited features

b) Wrong suggestions

c) Technical bugs

d) Complicated design

e) Other (please specify)

H. Suggestions for improvement

9) How may the writing applications be enhanced to better support English learners?

a) Be more individualized

b) Be capable of generating more context-based suggestions

c) Be able to suggest more accurate feedback

d) Other (please specify)

I. Satisfaction level

10) How satisfied are you with the experience of learning through AI-based writing applications? (1=very dissatisfied, 5=very satisfied)

a) 1

b) 2

c) 3

d) 4

e) 5

J. Recommendation

11) Would you recommend learning through AI-powered applications to your peers?

a) Yes

b) No

c) Not sure

Appendix 3: Teacher's Attitude Questionnaire

Name…… teaching experiences (years)……. Level……..

1. Which AI-based writing applications have you used in the classroom?
   f) Insta text
   g) Wordtune
   h) ChatGPT
   i) Other (please specify)
2. How often did you use it?
   a. Daily
   b. Weekly
   c. Monthly
   d. Rarely
3. How satisfied are you with this experience? (1=very dissatisfied, 5=very satisfied)
   a. 1
   b. 2
   c. 3
   d. 4
   e. 5
4. Would you recommend it to other EFL teachers?
   a. Yes
   b. No
   c. Not sure
5. How much AI-based writing applications could assist the EFL learners with the aspects mentioned below? Please rate (1=strongly disagree, 5= strongly agree)
   a. Improving students' overall writing skill ( )
   b. Providing instant corrective feedback ( )
   c. Decreasing mechanical errors ( )
   d. Increasing students' motivation and engagement ( )

6. Was implementing AI-based writing applications in your lesson plans beneficial to you?

    d) Yes

    e) No

    f) Not sure

7. If yes, how did was it beneficial to you?

    a. By providing automated written feedback on repetitive errors

    b. By increasing students' engagement and motivation

    c. By helping me devote time to other tasks

    d. Other (please specify)

8. What are the possible obstacles and concerns you foresee for the future of implementing AI-powered writing tools in your lesson plans?

    a. Students' dependence on them
    b. Inaccuracy of the feedback
    c. Technical issues and bugs
    d. lack of access to the applications

9. Do you believe that EFL teachers need further training to know how to effectively integrate technology into language learning?

    a. strongly disagree
    b. disagree
    c. neutral
    d. agree
    e. strongly agree

10. If you have any ideas, experiences and insights about implementing AI-powered writing tools in your lesson plans please share them.