A Comparative Study of AI-assisted Assessment and Teacher-Assessment in an EFL Writing Course

Alina Fazilova

A thesis submitted to the Faculty of Education and Humanities

in partial fulfillment of the requirements for the degree of

MASTER OF ARTS

In Teaching English as a Foreign Language

SDU University

Department of Language Teacher Education

June, 2025

Thesis Advisor:

Associate Professor, PhD, Kymbat Smakova

SDU University

Faculty of Education and Humanities
Department of Language Teacher Education

This is to certify that the Master's Thesis of

Alina Fazilova

has met the thesis requirements of
SDU University

Kaskelen, 2025

Approved by:

PhD, Associate Professor, Kymbat Smakova
Professor, Thesis supervisor

PhD., Doctor Gulzhaina KASSYMOVA
Committee Chair

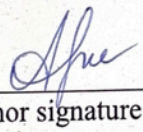
PhD., assistant professor
Akmarzhan NOGAIBAYEVA
Department Chair

PhD., associate professor Zhainagul DUISEBEKOVA
Dean, Faculty of Education and Humanities

AUTHOR AGREEMENT

The Author, I, Alina Fazilova, hereby grants SDU University (SDU) full and exclusive rights to reproduce the manuscript, make revisions, and reproduction. SDU rights include but are not limited to the following: (1) to reproduce, publish, sell, and distribute copies of the thesis, selections of the thesis, and translations and other derivative works based upon the thesis, in print, audio-visual, electronic, or by any and all media now or hereafter known or devised; (2) to license reprints of the thesis to third persons for educational photocopying; (3) to license others to create abstracts of the thesis and to index the thesis; (4) to license secondary publishers to reproduce the thesis in print or digital form, including electronic on-line databases; and (5) to license the thesis for document delivery. These exclusive rights run the full term of the copyright, and all renewals and extensions thereof.

I hereby accept the terms of the above Author Agreement[1] and I sign.



Author signature

20.06.25

Date

[1] *This page is to be separately printed, then signed and dated by the author, after that scanned and finally copy-pasted in this page.*

DECLARATION

I the undersigned, hereby declare that this submission is entirely my own work, in my own words, and that all sources used in researching it are fully acknowledged and all quotations properly identified to my best extent. It has not been submitted, in whole or in part, by me or another person, for the purpose of obtaining any other credit, except where due acknowledgement is made in the thesis. I understand the ethical implications of my research, and this work meets the requirements of the Faculty of Education and Humanities Research Ethics Policy.

Alina Fazilova

Name

Signature

20.06.25

Date

ACKNOWLEDGMENT

First of all, I want to thank God for giving me so many life opportunities and guiding me throughout my whole life, showing me the light and enlightenment, caressing me when I need it the most, tempering me like steel in various challenges, and sending me the best people for guidance. As I am now standing tall in the shadow of a large tree, it shelters me from the heat of the excruciatingly hot sun and protects me from the most violent storms.

Secondly, I want to thank my lovely mother for supporting me in this difficult journey; she was my inspiration, my lighthouse in an hour of darkness, and the harshest critic (in the best possible way). I could not have done it without her immeasurable contribution, resilience in the face of her illness, and constant warm presence in my heart.

Thirdly, I want to thank my friends who patted me on the back, screamed the ever-needed words of appreciation, whispered soothing words of ever-loving consolation, cheered the loudest for me, and laughed the loudest with me.

Lastly, I want to thank my vivid, full of life and energy students, motivating me to become better. I want to see vibrant smiles on their faces, to hear their unbelievable achievements, and experience sincere joy in my heart for every little or big step they make in their lives. You all are truly wonderful.

Table of Contents

List of Figures

List of Tables

A Comparative Study of AI-assisted Assessment and Teacher-Assessment in an EFL Writing Course

Abstract

With the rapid development of artificial intelligence in education, automated writing evaluation (AWE) tools such as ChatGPT are becoming increasingly popular for providing feedback on students' written works. This study explores the perceptions of first-year EFL students about AI-assisted assessment compared to teacher assessment in an EFL writing course at a private Kazakhstani university. Over the period of one semester, 33 participants have written 4 essays and received both the teacher and ChatGPT feedback on each essay. Their perceptions have been compared and analyzed with a quantitative research design with elements of qualitative analysis. A questionnaire with Likert-scale closed-ended and open-ended questions was used. The results revealed that while students recognize the importance of AI-assisted assessment for surface-level corrections (grammar, vocabulary, structure), the majority of students prefer teacher feedback for its clarity, personalized support, and depth. Additionally, most students viewed the ideal approach as a combination of two types of feedback: AI for quick technical feedback and teachers for more complex aspects like structure, argumentation, and tone. The study concludes that although AWE tools have potential as supplementary support in EFL writing instruction, they cannot replace the human connection and pedagogical insight offered by teachers. Implications for integrating AI tools into classroom practice and teacher training are also discussed, along with recommendations for future research in this evolving field.

Keywords: AI in education, ChatGPT, EFL writing, teacher feedback, automated writing evaluation, feedback literacy, Kazakhstan

Ағылшын тілі шетел тілі ретіндегі жазу курсында жасанды интеллект (ЖИ) көмегімен бағалау мен оқытушының бағалауын салыстырмалы зерттеу

Аңдатпа

Білім беру саласында жасанды интеллекттің қарқынды дамуына байланысты ChatGPT сияқты автоматтандырылған жазбаша жұмыстарды бағалау (АЖБ) құралдары студенттердің жазбаша жұмыстарына кері байланыс беруде кеңінен қолданылуда. Бұл зерттеу Қазақстандағы университеттің академиялық жазылым курсына қатысатын ағылшын тілін шет тілі ретінде (EFL) үйренуші студенттердің дәстүрлі мұғалім беретін кері байланыс пен ЖИ-дің кері байланысына деген көзқарастарын салыстырға қарастырады. Аралас әдіснамалық тәсіл қолданыла отырып, бір семестр бойы 33 студент төрт эссе жазып, әр эссеге мұғалімнен де, ChatGPT-ден де кері байланыс берілді. Студенттердің көзқарастары сандық зерттеу дизайны негізінде, сапалық талдау элементтерімен салыстырылып, талданды. Деректер жабық (Ликерт шкаласы негізінде) және ашық сұрақтардан тұратын сауалнама арқылы жиналды. Зерттеу нәтижелері көрсеткендей, студенттер жасанды интеллекттің грамматикалық, лексикалық және құрылымдық қателерді түзетуге көмектесетінін мойындағанымен, олардың көпшілігі мұғалімнің кері байланысын анағұрлым түсінікті, жеке бағытталған және терең деп қабылдайтынын көрсетті. Сондай-ақ, көптеген студенттер екі түрлі кері байланысты біріктіріп қолдануды ең тиімді тәсіл деп санады: жасанды интеллект техникалық кемшіліктерді тез түзету үшін, мұғалімдер құрылым, дәлел келтіру, стиль секілді күрделі аспектілер бойынша кері байланыс беру үшін. Зерттеу нәтижесі бойынша, АЖБ құралдары EFL жазылымды оқытуда қосымша қолдау құралы ретінде әлеуетке ие болғанымен, олар мұғалім ұсынатын адами өзара әрекеттестік пен педагогикалық түсініктің орнын толтыра алмайды. Сонымен қатар, бұл жұмыста ЖИ құралдарын оқу үдерісіне және мұғалімдерді даярлауға енгізудің мүмкіндіктері мен осы салада болашақ зерттеулерге ұсыныстар қарастырылады.

Түйінді сөздер: білім берудегі ЖИ, ChatGPT, ағылшын тіліндегі жазылым, мұғалімнің кері байланысы, автоматтандырылған жазбаша бағалау, кері байланыс сауаттылығы, Қазақстан

Сравнительное исследование оценивания с помощью ИИ и оценивания учителей на занятиях письменности английского языка как иностранного

Аннотация

С развитием искусственного интеллекта в сфере образования системы автоматизированного оценивания письменных работ (АОПР), такие как ChatGPT, становятся всё более популярными для предоставления обратной связи на письменные работы студентов. Настоящее исследование направлено на изучение восприятия первокурсников, изучающих английский язык как иностранный, относительно оценки, осуществляемой с помощью ИИ, по сравнению с оценкой, предоставляемой преподавателем, в рамках курса академического письма в одном из частных университетов Казахстана. В течение одного семестра 33 участника написали 4 эссе и получили обратную связь как от преподавателя, так и от ChatGPT по каждому из них. Их восприятие было проанализировано в рамках количественного подхода с элементами качественного анализа. В качестве инструмента исследования использовалась анкета, включающая как закрытые вопросы с шкалой Лайкерта, так и открытые вопросы. Результаты показали, что, несмотря на признание пользы ИИ-оценки для поверхностных исправлений (грамматика, лексика, структура), большинство студентов предпочли обратную связь от преподавателя за её ясность, персонализированную поддержку и глубину. Кроме того, большинство респондентов посчитали идеальным вариантом комбинирование двух типов обратной связи: ИИ для оперативных технических исправлений, преподаватель для более сложных аспектов, таких как структура, аргументация и тон. Исследование подтверждает, что, несмотря на потенциал АОПР как вспомогательного инструмента в обучении письму на английском языке, они не могут заменить человеческое взаимодействие и педагогическую проницательность преподавателя. В работе также рассматриваются возможные пути интеграции ИИ-инструментов в образовательную практику и подготовку преподавателей, а также даются рекомендации для будущих исследований в данной развивающейся области.

Ключевые слова: ИИ в образовании, ChatGPT, EFL письмо, обратная связь учителя, автоматизированное оценивание письменных работ, грамотность в работе с обратной связью, Казахстан

CHAPTER 1

Introduction

With the emergence of new contemporary technology, new gates of novel education methods are opening now, especially in the field of foreign language learning and teaching. After the unwavering advancement of globalization and the establishment of English as a lingua franca, it became an utmost necessity for EFL teachers to satisfy the growing demand for quality education in the field (Kirkpatrick & Schaller-Schwaner, 2022). This demand revealed the struggles of teaching one of the most complex language skills, the skill that requires learners to produce language and indicates overall fluency in a language - writing. This has resulted in a heavier workload for educators, as evaluating writing is both intricate and time-intensive (Aydin, 2012; Al-Gharabally, 2015; Çankaya, 2018). Some of the previous attempts to implement different learning technologies showed auspicious results with high hopes for improvement in the future (Haddad & Draxler, 2002; Cloete, 2017; Lai & Bower, 2019). The multipurpose use of digital devices is becoming more and more prevalent among the growing generation of children immersed in the digital world. Apart from entertainment, easy access to education and other sources of information make learning more attainable than ever. Moreover, with the advancement of technology, studying traditionally with a pen and paper becomes less and less effective as new teaching and learning methods adapt to a more innovative system that involves technology that is not just a provider of information but also a guide, tutor, assessor, and creator (Haleem et al., 2022). This synergy with technologies helps students and educators create an effective and interesting learning environment corresponding to contemporary needs (Qureshi et al., 2022). The recent emergence of generative AI chatbots (ChatGPT and similar) that generate

personalized human-like texts and base their responses on large corpora of data has created a myriad of opportunities for education stakeholders to streamline their work (Lo, 2023).

Although rapid advancement of technologies denotes optimization of the working process for teachers, they still experience many difficulties. Apart from the concerns related to personal/family issues, professional development, working conditions, curriculum design, social engagement, teacher-parent communication, school/college administration, and the physical condition of a teacher, Aydin (2012) also notes that the other part of the workload falls onto the shoulders of teachers for assessing writing. It is one of the main causes of the intensive workload for teachers who conduct writing courses. It appears from the amount of time educators devote to assessing and giving extensive feedback on students' writing. All these factors have influenced their psychological and physical well-being, revealing the tendency among almost half of the novice EFL teachers to quit their jobs in the first 5 years of their teaching career (Fathi et al., 2021).

However, some of the studies show that this particular workload caused by peculiarities of writing assessment could be alleviated with innovative methods of assessment based on the idea of integrating Computer-assisted language learning (CALL) into the learning and teaching processes. Computer-assisted language learning (CALL) is a field in language education that concerns the use of computer technology in language learning and the diverse pedagogical principles underlying that. It has a long history of development and revisions in terms of its definition. At first, CALL was defined by Levy (1997, p.1) as the "study of the computer applications in language teaching and learning." Later, the concept of CALL drastically widened as technology became an integral part of everyday life and required a more rigid definition.

Thus, considering its ever-changing nature, Beatty (2010, p.7) defined CALL as “any process in which a learner uses a computer and, as a result, improves his or her language.”

A promising tool that aids in reducing the workload of foreign language teachers while checking written works is Automated Writing Evaluation (AWE). Automated Writing Evaluation systems (AWE) could be a useful tool in facilitating an efficient and objective assessment of students’ writing that is automated and instant. At its core, AWE is an AI-based writing evaluation tool based on 1) Natural Language Processing (NLP), which allows them easily analyse and interpret different text features, such as cohesion, lexical aspects, syntactic complexity, and discourse; 2) and machine learning, allowing them to evaluate the written pieces (assign a certain score) and give feedback based on the analyzed text features. Moreover, recently-developed AWEs are equipped with more advanced functions, if they are combined with other technology-assisted learning methods and teaching approaches (Learning management systems, peer feedback, ethical guidelines), they may not only evaluate and give feedback but also potentially improve the learning-teaching experience for educators and students (Balfour, 2013). The reliability of AWEs and their accuracy were tested in a series of studies reflected in the study review conducted by Wilson and Roscoe (2019).

Despite the advantages that AWEs offer, they might pose some challenges and concerns that researchers must consider. Although generative AIs may provide a very comfortable environment for autonomous learning, this contradicts the core principle of constructivism. The first is the question of whether these systems are truly capable of autonomous work without the intervention of the teacher. Secondly, not all educators have enough competency and technical skills to work with these tools. Thirdly, the ethical aspect of the use of different AWEs arises as they are the property of third parties, and all the information and texts uploaded to these tools

might not be confidential. Although many companies owning these tools claim all the user data uploaded into their applications is kept confidential, it is almost impossible to know whether the data you upload stays 100% confidential. Lastly, although the reliability and accuracy of the AWEs were proven to be high by some of the previous research, they still lack human factor due to the rigor and objectivity of the assessment they produce; they may not be as accurate as human teachers in identifying some of the mistakes that require socio-emotional understanding that only human raters possess (Fu et al., 2024).

With these systems in use, it may be possible to decrease the burden of assessment on the EFL writing course teachers. However, we should not neglect the drawbacks of these systems as they can play a substantial role in the overall accuracy of the assessments given and students' grades subsequently.

Thus, this study aims to compare the assessments given to first-year EFL students' essay writing with two different assessment methods (AI-assisted assessment (AWE) and instructor's assessment) and to identify which method is preferable by students in providing detailed feedback. The first assessment method is done with the help of the Automated Writing Evaluation system, and the second is performed by the actual instructors of the EFL writing course. This aim could be achieved by answering the following research questions:

1) In what ways do the assessments done by teachers differ from the assessments provided by the AWE according to students?
2) How do these differences influence students' preferences and writing performance?

The obtained results might unfold the possible application of AWE in higher education in Kazakhstan. These valuable insights from the learners' perspective might improve the learning and teaching experience by integrating contemporary technology into this process.

CHAPTER 2

Literature Review

2.1 Short overview of Computer-Assisted Language Learning

The theoretical basis for all the approaches in teaching involving partial or full integration of computer-based technology is built on the concept of computer-assisted language learning. The CALL has been developing from the 1960s to the present day. In the early years, it was used to practice traditional foreign language teaching styles in line with drilling and simple translation practices. In recent times, technology has significantly advanced. It is now able to provide a range of types of learning, methods, and techniques that integrate computers and their capabilities, starting from distance learning to the most recent use of multimedia applications (Levy, 1997).

As CALL technologies were developing, new types of CALL emerged, and Warschauer and Healey (1998) developed a classification of CALL technologies based on the pedagogy they used:

- Behavioristic CALL: a collective term for all CALL technology developed in the period between the 1950s and 1970s. The primary focus of the programs conceived at that time was in accordance with the behaviorist theory. Language learners were taught through the means of drills and controlled practice that was then checked with the help of CALL via text.

- Communicative CALL: the second phase of CALL development that occurred from the 1970s to the 1980s. Here, taking into consideration the communicative approach, the focus shifted to the enhancement of language use and fluency development. The main innovation of this stage was the provision of the context to increase the authenticity of the English language practice. However, the computer was still seen as a tutor who could not fully provide the growth of all four language skills.
- Integrative CALL was the apogee of the technological advancements that embraced the appearance of the World Wide Web and its multimedia features from the 1990s up until now. From this period, education became accessible outside of the classroom. The internet gave its start to the prosperity of online education and more independent programs, applications, and games that learners could access without the dire need to come to classes.
- Intelligent CALL or ICALL is the most recent branch of CALL that is focused on the integration of AI technology into language learning (Davies, 2002). In particular, Artificial Intelligence (AI) is capable of processing large amounts of linguistic data through Natural Language Processing, which is a term to describe the synthesis of speech and text. Another characteristic of AI is machine learning. Simply put, it allows AI systems to generalize and perform tasks without particular instructions.

2.1.1 Advantages and disadvantages of ICALL

It is important to explore the positive and negative implications of AI integration into foreign language learning and teaching before applying it in a real-life context. So far, the scholarly studies conducted to identify the advantages and disadvantages of ICALL in FLT present different stances on the way we should approach it. De la Vall and Araya (2023) have

distinguished some benefits of using AI and juxtaposed them against some of the drawbacks. The main positive characteristics of ICALL were mentioned to be time efficiency as they provide immediate automated feedback that is objective and accurate; personalization of learning as AI algorithms learn from the input and preferences of the user and adjust accordingly; accessibility as any learner can access this technology with the only requirement of having a device with internet access; and lastly, cost-effectiveness as the majority of AI-powered tools are free to access or have a low cost. On the other hand, Kannan and Munday (2018) also described some of the issues with the use of language learning tools that were powered by AI specifically. Firstly, dehumanization may pose a great threat to developing and establishing normal human communication between teachers and students if technology is the main mediator between them. Secondly, many AI-powered tools lack originality in generated works, which stems from the problem of recognizing more advanced linguistic features of the text. Thirdly, AI language models may produce bias and inaccuracies in many spheres, especially in recognizing and generating culture-specific content. The list of drawbacks is even longer if the ethical considerations are also taken into account. With that being said, Mason (2017) analyzed four main ethical concerns regarding the use of technology (PAPA):

- Privacy - the boundaries of what information one should post on the web and what parts of it should stay confidential are blurred. Under these circumstances, the author highlighted two main threats to our privacy - “one is the growth of information technology” and “the increased value of information in decision-making” (p. 5).
- Accuracy - Since the value of the information grew, the accuracy of the information declined drastically. Misinformation is the bane of this age as it poisons the minds of

susceptible people and takes advantage of vulnerable groups (children, elderly, mentally impaired).

- Property - Although many organizations, leagues, and companies attempt to fight the theft of intellectual property, the problem is still to be extirpated. Our data remains unprotected from the all-seeing gaze of third parties.
- Access - due to the low cost of access to the compilation of large amounts of data on the net, it is essential that all users know how to navigate this data safely and effectively, without falling into the traps of mala fide people. The author suggests that digital literacy and critical thinking should go hand in hand to avoid these traps.

After learning about the possible implications of using CALL, ICALL, and AI technology in education, it is essential to approach this research carefully and with all considerations in mind. The advantages seem solid; however, disadvantages and various ethical concerns may outweigh them if not approached carefully. Although ICALL explores the close interconnection of relationship between language learning and modern technology, it is primarily imbedded in technological and pedagogical frameworks. Thus, understanding how students perceive and respond to different assessment sources (teacher and AI-generated assessment), this study adopts Feedback Literacy Framework (Carless & Boud, 2018) as its theoretical basis.

2.2 Implications for AI in writing assessment

The integration of Automated Writing Evaluation (AWE) systems into educational settings has sparked interest among many educators and scholars, as it is essential to understand how AI-powered assessments compare to traditional teacher assessments. This section explores feedback as an integral part of assessment and the differences between AWE and teacher

assessment regarding students' and teachers' perceptions in terms of types, timeliness, and personalization of feedback.

2.2.1 Assessment and feedback: types and focus. This study is guided by the Feedback Literacy Framework (Carless & Boud, 2018). It conceptualizes feedback as a student-centered process that involves affective, behavioral, and cognitive domains. According to this framework, students are not viewed as passive recipients of feedback, but rather have to actively participate in it by comprehending, reflecting, and acting upon the given feedback to improve. Carless and Boud (2018) defined four main dimensions of feedback to consider:

1. *Appreciation of feedback*. The learning process is recognized as valuable and contains a meaningful purpose (feedback for learning).
2. *Making judgments*. Developing the ability of assessing the quality of student's own work and picture the idea of what a good work should look like.
3. *Managing affective reactions*. Dealing with negative emotions (anxiety, writer's block, frustration, etc.) and developing motivation to grow.
4. *Taking action*. Transferring feedback suggestions into meaningful revisions and future improvements.

In English as a Foreign Language, assessment and feedback are distinct yet interrelated. Assessment is defined as "systematic collection of information about learners' knowledge, skills, abilities, and progress" (Black & William, 2009, p. 10). It is categorized into two types - formative and summative. While the former focuses on fostering supportive learning during the instructional process, the latter evaluates this progress at the end of the instructional period, assigning a certain value relative to the learner's endeavour. Feedback, on the other hand, is the information provided to learners about their current progress and performance on a task in

alignment with the learning goals; it is considered to be an essential part of the ongoing formative assessment (Hyland & Hyland, 2006). Overall, effective assessment always relies on timely and meaningful feedback, and meaningful feedback is grounded in assessment data. For instance, the comments given by a teacher on a written task act both as an assessment of a student's performance and a source of information (feedback) for him to follow and improve.

The discussions about the influence of assessment and feedback on learning are vast and always relevant, especially with new advancements in education and technology. Combining assessment and feedback and, moreover, balancing them can be a challenge for any educator. As Watling and Jinsburg (2019) put it, "whereas feedback stresses development and learning, assessment stresses judgement and decision making" (p.77). Finding a silver lining while combining both effectively is of the highest importance in the field of education. It is often difficult for teachers to provide effective and fulfilling written feedback as they struggle with the proper use of rating scales and put uneven focus on a certain area for improvement, neglecting other areas (emphasizing linguistic features over content) (Kong et al., 2022). Feedback is considered to be effective when students notice it, take action, and improve their future work according to the recommendations and guidance provided by a teacher. According to Gibbs and Simpson (2004), you should make sure that your written feedback is 1) timely, sufficient, and responsive to the needs of your student; 2) follows the task criteria and purpose; 3) clear and uses simple language; and 4) emphasizes learning rather than simple marking. Although its importance is out of question, traditional feedback provided on students' written works frequently focuses on identifying weaknesses without providing any corrective feedback or explanations (Glover & Brown, 2015). In the same study, after reviewing more than 4400 written feedback from a number of assignments in two universities, scholars analyzed the depth of this

written feedback, concluding that oftentimes, assessors prioritized marking over feedback “for learning”.

2.2.2 AWE systems used in assessment of writing. The standard method of essay assessment has proved to be tedious, time-consuming, and overwhelming for educators to use (Gao & Ma, 2020; Kong et al., 2022). To lower this intensifying workload and provide objective, immediate, and valid feedback, Automated Writing Evaluation (AWE) systems were introduced to address these problems. AWE systems excel in providing detailed and immediate feedback on technical aspects of writing, such as grammar, syntax, and vocabulary. In contrast, teachers also focus on content, creativity, and structure, offering a more holistic review of student writing (Sari & Han, 2024; Cheng & Zhang, 2024). Additionally, many generally accessible GPT-powered AI models (ChatGPT, GrokAI, DeepSeek) can work as essay raters given a proper prompt with a rubric and assessment criteria. The validity, reliability, and validation of a range of AWE systems were proven to be high (Clarkson, 2023). While teacher assessment lacks consistency and it is difficult to establish high inter-rater reliability, AWE offers consistent and objective feedback, reducing variability (Liu & Liu, 2023). This makes the use of AWE an easy assessment tool for educators when the specific focus of the evaluation is accuracy and vocabulary, as these programs can easily indicate the problems in the linguistic structure of an essay. However, the authors also noticed one main concern with all AWE platforms. It is a question of ethics whether computer-assisted evaluation systems can evaluate the subjective content of the essays.

ChatGPT-4o was used in this study as the automated writing evaluation (AWE) system to provide students with AI-generated feedback on their essays. The reason why the preference was given to ChatGPT over other more traditional AWE systems like Grammarly, Criterion, and others was based on several pedagogical and functional considerations. Most of the traditional

AWE systems are very strict rule-based and accuracy-driven models that focus on checking and improving grammatical features of writing. Unlike these systems, ChatGPT is capable of generating context-based, personalized feedback that tries to simulate human generated text. So, its ability to process and analyze natural language inputs makes it more flexible for giving structural advice, detailed suggestions, and clear explanations that could reflect teacher-like feedback to some degree (Zhai, 2022). Although ChatGPT has significant advantages, it is important to recognize that it also has some limitations that might affect the generated output. Occasionally, it can produce overgeneralized suggestions and might have lack of awareness about macro-level aspects of writing (cultural, emotional, or contextual nuances) that a human teacher can provide (Dwivedi et al., 2023).

Wang (2020) reviewed the particular use of AWE systems in an EFL reading and writing course in China. The participants of this study were university students of different years and majors. During the whole course, students were asked to write four essays, and afterward, their essays were assessed by both the AWE system and teachers. The author searched to determine the comparison between the two ratings. The study findings revealed that students perceived AWE raters more positively than human raters. However, despite their capabilities to provide detailed feedback on grammar, syntactic, and language mechanics, Wang concludes that “human raters are indispensable where there is no consistent standard, such as logic consistency, organization, and development.” (p. 14). Overall, the author advises the mindful use of AWE raters in cooperation with human raters to make up for the disadvantages of both. Similar findings were discovered by Gayed et al. (2022) in exploring the impact of AI-powered writing assistant AI KAKU (based on GPT-2 architecture) on EFL students’ writing skills. They explored the specific features of this tool in assessing and providing feedback on students’

written pieces. The researchers noted that this tool performs positively in recognizing mistakes in syntactic and linguistic mechanics; however, it still lacks a more conscious understanding of organizational and contextual errors made by students.

On the one hand, the effectiveness of AWE raters and AI-powered tools has proved to be valid in terms of the efficiency and timeliness of the feedback they provide. On the other hand, it cannot be completely deprived of the teacher's supervision since AWE systems may overlook the organizational and subjective features of the written pieces. There is a comprehensive list of concerns about the use of AI in education posed by Swiecki et al. (2022):

1. **Responsibility shift**. Although we assume that AI is a perfectly objective system deprived of bias, it is a wrong notion. The idea of using AI for grading sounds appealing to school administrators, teachers, and even some students, as we shift the responsibility to AI and the people behind its development. So the question is, who is responsible if it is wrong or unfair?
2. **Personalization issues**. AI assessment might lack the personalization of feedback that is usually used by teachers in order to motivate or punish students by addressing more contextual needs and resorting to the pedagogical understanding of human nature. This concern also touches on the topic of the complete or partial replacement of educators by AI and its possible outcomes.
3. **Imposing certain values**. Many AI models may favor a certain type of thinking and behaving that reflects the common individualistic, white, male, Westernized values. This approach might ignore “other” non-conventional values, cultures, and ways of thinking that stand out and do not fit the “standardized” point of view.

4. **Shallow assessments**. AI systems are well-trained for recognizing texts and meaning, but they are not able to understand subtleties like humor, sarcasm, and irony. It is difficult for them to assess subtle aspects like ethics, originality, or even improvisation, which are vital for real learning.
5. **Surveillance pedagogy**. AI systems always collect data and track students' progress. While it may sound like a good feature for overall progression, it might feel like constant surveillance and cause problems with trust. As more and more AI services grow, it is hard to ensure full confidentiality as many third parties get involved.

2.3 Perceptions of AWE

According to Wilson et al. (2021), elementary school teachers perceive AWE as both a supportive and challenging tool. While it offers many new opportunities for assessment, it also presents challenges in the way teachers must learn how to navigate in their instructional practices. This requires additional teacher training and professional development in order to smoothly incorporate AWE into the classroom. Another study by Zhai and Ma (2021) explored the perceptions of 53 English major students enrolled in English writing class. The course was taught with the help of the AWE program Criterion. After 18 weeks of studying, collecting, and revising drafts, results showed that most of the students positively perceived the use of Criterion as it helped in the technical aspects of the language. However, many of them also expressed a preference for using it in combination with teacher assessment, as it was the most effective way to enhance their writing skills. Although students were satisfied with the timely feedback and error correction, many organizational improvements in writing were attributed not to the Criterion but to the teacher's feedback. The overall acceptance of AWE by students depends on several factors, according to the Technology Acceptance Model:

1) How useful they think it is (perceived usefulness);
2) How easy it is to use (perceived ease of use);
3) What others think (like teachers and peers);
4) Their trust and confidence in using the system;
5) How well the system provides meaningful feedback;
6) The system's design and features.

After reviewing studies on AWE and its perceived usefulness, Fu et al. (2022) found that the majority of them were quantitative, short-term (no longer than 10 weeks), and small-scaled (low number of participants), which proved the need for a more multifaceted and longitudinal approach to studying perceptions of AWE systems among stakeholders. Their findings revealed that most of the studies reported that the majority of students were positive about AWE feedback, finding it useful and motivating. Although AWE feedback was regarded as timely, useful, and motivating, it often was inferior to teachers' feedback in accuracy, clear explanations, and personalization. It was often regarded as too formulaic and generic, which supports the previous findings (Zhai and Ma, 2021; Wilson et al., 2021; Wang, 2020; Swiecki et al., 2022; Gayed et al., 2022).

2.4 Use of AI in Kazakhstani Education

The rapid growth of AI systems and their global utilization leads to a logical question of whether these systems can be ubiquitously integrated into different education systems. Several Kazakhstani scholars have noted possible prospects of how AI could be integrated into the pedagogical process of our educational institutions; however, many of them base these ideas on Western experience and lack information about the actual practices of employing AI technologies. For example, in the study by Kaldygozova et al. (2025), they reviewed 25

Kazakhstani articles about artificial intelligence in the field of monitoring and quality assurance of education. They generally describe possible applications of AI and introduce several useful platforms, applications, and programs that might be suitable for certain teaching practices, such as AI scoring, partial substitution of teachers, analysis of academic activity, and proctoring. They mentioned the development of an original to Kazakhstan AI chatbot “Talapker”, that will serve as a guide for secondary school students who prepare to pass the unified national testing (UNT). Despite providing a comprehensive introduction to AI systems, they have not provided any authentic teaching practices involving them. Another similar study by Dussembinova and Ramazanova (2024) provides a short review of international practices in integrating AI in higher education, its benefits and drawbacks, and a short discussion of its application in the Kazakhstani context. According to their review, international practices show mostly positive results of employing AI programs for proctoring, creating adaptive learning environments, and assessment. The reviews suggest that both articles have scarce information on the authentic implementation of AI in teaching and learning. It becomes apparent that there is a huge gap in knowledge about the practical application of AI in Kazakhstani education. As a result, this field requires a more in-depth understanding of AI in our region.

CHAPTER 3

Methodology

3.1 Research Design

This study employs a mixed research design with a quasi-experimental method to address the following research questions:

1) In what ways do the assessments done by teacher differ from the assessments provided by the AWE, according to students?
2) What reasons do students give for preferring teacher assessment or AWE assessment on their writing?

In order to state clear expectations from the results, the study also employs two hypotheses to see if the expectations about students' responses and preferences will be met or not. As follows:

Null hypothesis (H^0): There is no significant difference in the perceived effectiveness of teacher feedback and AWE feedback in improving students' performance in writing.

Alternative hypothesis (H^1): AWE feedback is perceived to be more effective than teacher feedback in improving students' performance in writing.

3.2 Participants

The initial number of participants in the study was 39. The convenience sampling method was used as students enrolled in the course at the researcher's convenience. However, one participant dropped out in the middle of the course due to a high absence index. Overall, 38 students participated in the study, submitting their essay drafts and receiving two types of feedback, and only 33 of them filled in the questionnaire.

3.3 Procedures

The study lasted 15 weeks (one semester) and was integrated into the university's Reading and Writing course for first-year students of two English-related disciplines. To keep the clarity of academic conduct, Turnitin was used to ensure compliance with academic integrity, so all the drafts violating it were rejected. The number of students who received feedback on each essay was inconsistent, as some students could not participate in the draft-writing lessons, plagiarized, or had their essays generated by AI. Throughout the course, students were assigned to write 4 different essays (persuasive, opinion, research-based, and descriptive-narrative essays) concerning the topic of the corresponding coursebook unit. Before proceeding to write their first draft, the teacher instructed them on the peculiarities and distinctive features of each essay. Students were given more than one hour to write a full draft (introduction, body, conclusion) in class. Those students who were not able to participate in the lesson were provided with the opportunity to submit an online draft written at home. After collecting all the drafts, the

instructor transferred all the physical writing to an electronic format in Google Docs with the help of ChatGPT and its text recognition. Although ChatGPT's optical character recognition (OCR) technology is very advanced, it still confused some students' handwriting. So, to ensure the full authenticity of written works, the teacher revised the transferred text for typos and errors. Each student was given feedback individually. In order to avoid bias and ensure assessment rigor, both the teacher and ChatGPT used the standard IELTS Writing Task 2 rubric to calibrate feedback. Another strategy to mitigate bias was to follow a fixed order of feedback provision: the teacher evaluated the work first, then uploaded the essay to ChatGPT for AI evaluation. This could help the human assessor avoid the influence of ChatGPT's scoring on independent teacher scoring. As soon as both types of feedback were ready, the students were sent a document where they could see teacher and AI feedback on their draft and compare for themselves. The feedback did not affect students' grades, but was a gentle guidance for them to improve. Later, they could improve their drafts and submit the revised version for grading.

After the stage of essay review, students were sent a questionnaire to share their preferences about feedback and express their perception of the two different types of feedback they received. Upon collecting all the data, it was carefully analyzed and interpreted.

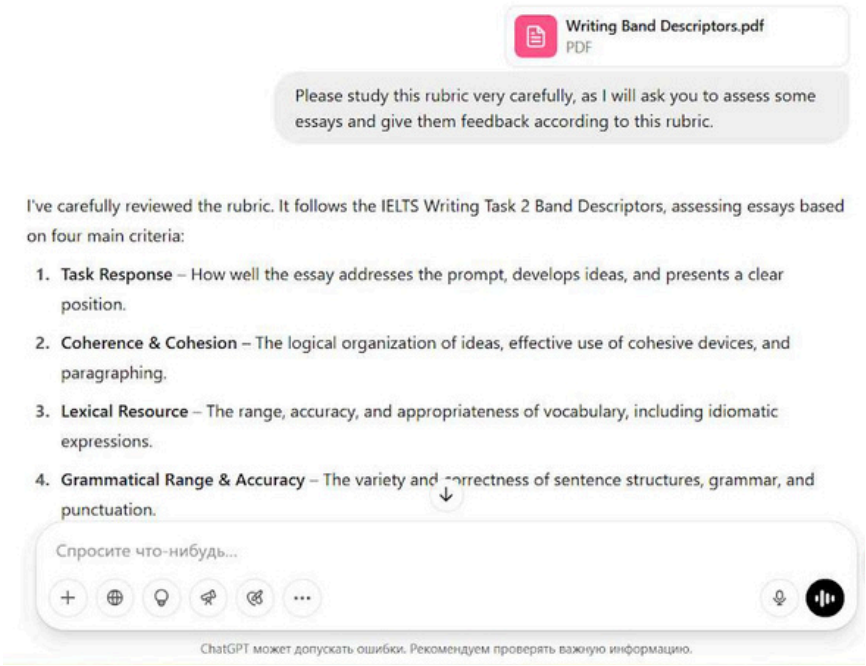


Figure 1. A ChatGPT prompt for essay evaluation.

3.4 The Data Collection Tools

The process of data collection included the provision of feedback during the course and a questionnaire to assess students' perception of both types of feedback. For this reason, the questionnaire consisted of three main parts. The first part, *General Feedback Preferences*, was adapted from the Feedback Orientation Scale (FOS) developed by Linderbaum and Levy (2010). It was modified to fit the context of EFL students' perceptions. It aims to measure the general preferences of students in feedback given to their written works. The second part of the questionnaire, *Comparative evaluation of Teacher and AI feedback*, was adapted from the System Usability Scale by Brooke (1996). This well-designed tool was modified to evaluate the perceived usability of both AI-generated and teacher feedback in a writing course context. The third part is a direct comparison of the two feedback types across different areas (clarity of feedback, details, helpfulness, ease of application). All three parts use a 5-point Likert scale (1 =

Strongly Disagree, 5 = Strongly Agree) of measurement to assess students' attitudes towards feedback. The last section included 4 open-ended questions to gain more in-depth insights from students:

(1) What aspects of the teacher's feedback did you find most helpful for improving your essay?

(2) What aspects of the AI system's feedback did you find most helpful for improving your essay?

(3) Were there any areas where you felt the feedback from either the teacher or the AI system was lacking? Please explain.

(4) How do you think combining teacher and AI feedback could improve the feedback process for students?

(5) Please explain why you prefer either teacher feedback or AI feedback when working on your writing.

In total, the questionnaire has 28 questions.

3.5 Data Analysis

For analyzing the quantitative part of this study, Jamovi statistical software was used (The Jamovi Project, 2024). Since the major section of the questionnaire consists of Likert-scale items, it is important to calculate the internal consistency of the questionnaire sections (*General Feedback Orientation, AI Feedback Perceptions, and Teacher Feedback Perceptions*). Cronbach's alpha was calculated for each section to assess this reliability.

3.5.1 General feedback orientation. Six questions were dedicated to measuring students' perception of feedback on written works and understanding how important feedback is for

students in the EFL context. It used five-point Likert scale questions with a good Cronbach's α of 0.81.

3.5.2 AI feedback perceptions. This section had six questions inquiring about students' perceptions of the effectiveness of the AI feedback they received. Its reliability was calculated with Cronbach's α and had a satisfactory value of 0.77.

3.5.3 Teacher feedback perceptions. To juxtapose two pieces of feedback, this section also uses six questions measuring the same parameters of feedback provided by the teacher. The Cronbach's α reliability score of this section was 0.87.

Relatively lower reliability scores for the first two sections could be attributed to a rather small number of items.

Descriptive statistics (means, SD, and frequency distributions) were calculated to summarize and interpret students' perceptions and preferences. It indicates general tendencies in their answers. Additionally, to address research question 1, a series of paired t-tests were used to compare the ratings of AI-generated feedback and teacher feedback on matched items (clarity, details, helpfulness, application). Paired t-tests evaluate whether the same participants perceived significant differences between the two types of feedback across multiple aspects. Finally, to examine overall trends for categorical preference data (preferred feedback source), frequency tables were generated.

As for the qualitative part of this study, all open-ended question entries were analyzed according to the inductive thematic analysis introduced by Braun and Clarke (2006). It follows a six-step procedure of:

1) Familiarizing with the data collected;
2) Assigning initial codes;

3) Identification of themes;
4) Themes' revision;
5) Establishing and naming themes;
6) Reporting your findings.

The responses were manually coded. Major themes and subthemes were derived and interpreted by the researcher. Next, students were coded (S1, S2, … S33) to follow a neat order and organize their responses. The results of this analysis provide additional evidence about the outcomes of this study. They might provide some suggestions for improving methodology in the future.

3.6 Ethical Considerations

At the beginning of the course, all enrolled students received an informed consent letter with information about the research and its procedures, including ethical considerations, benefits, and possible risks, where they could approve their participation or refuse it. Participation of underage students was consented to via a parental consent form sent to their parents or guardians. Participation in this study is entirely voluntary. Additionally, to uphold privacy, no personal data was collected, ensuring the confidentiality and anonymity of participants throughout the research process. This ethical approach aims to prioritize participant welfare and rights.

CHAPTER 4

Results

4.1 Descriptive Statistics Analysis

4.1.1 Analysis of general feedback preferences. In order to see whether students perceive feedback as a valuable tool for improving their writing skills, descriptive statistics were used in this study. The values of mean, SD, min/max, and skewness were calculated to gain more detailed information.

Table 1

General Feedback Preferences

Item	Mean	Median	SD	Min	Max	Skewness
I actively seek detailed feedback to improve my writing skills.	4.18	4	0.950	2	5	-0.851
I find feedback valuable for identifying my strengths and weaknesses in writing.	4.45	5	0.938	1	5	-2.28
I prefer feedback that provides specific examples or corrections.	4.70	5	0.529	3	5	-1.55
Constructive criticism motivates me to work harder on improving my essays.	4.12	4	0.927	1	5	-1.25
I am more likely to use feedback when it is easy to understand and practical.	4.48	5	0.619	3	5	-0.783
I believe feedback should focus on both content (ideas, arguments) and mechanics (grammar, style).	4.70	5	0.585	3	5	-1.84

Note. SD = Standard Deviation

According to the results of descriptive statistics, participants rated feedback as an important aspect in their learning process, as the mean values range from 4.18 to 4.70. It indicates a generally positive attitude towards feedback across all items. The highest mean scores are for Q3 “I prefer feedback that provides specific examples or corrections” (M=4.70) and Q4 “Feedback should focus on both content and mechanics” (M=4.70). Both of these have medians of 5 which suggests that students have a strong consensus that a specific and comprehensive feedback is highly valued. The range for SD (0.529-0.950) demonstrates that there was some variation in agreement across items. The lowest SD (0.529) indicates a high consistency in responses for specific examples and corrections. The highest SD (0.950) is for actively seeking feedback showing more variability in the way students seek feedback, where some are less proactive than others. According to the distribution (skewness), all of the values are negative, which emphasizes that participants rated the questions on the higher end of the scale from 4 to 5. The most negatively skewed items are Q2 “Feedback valuable for identifying strengths and weaknesses” (Skewness = -2.28) and Q6 “Feedback should focus on both content and mechanics” (Skewness = -1.84). This strong negative skewness confirms that students strongly agreed with the statements and very few selected lower ratings. The values of minimum (ranging from 1 to 3) show that despite the majority of students rating feedback preferences positively, a few of them gave lower ratings, particularly for the value of feedback in identifying strengths and weaknesses (min = 1) and constructive criticism as motivation (min = 1).

Overall, participants express a very strong positive orientation towards feedback, highlighting that detailed and actionable feedback that addresses both content and mechanics helps them achieve the most results. This information provides valuable insights into students’ perception of feedback and its importance from their perspective.

4.1.2 Decriptive analysis of AI feedback. This part of the questionnaire measured whether feedback provided by ChatGPT was effective across aspects like clarity of explanations, detailed corrections, comprehensiveness, meeting expectations, ease of application, and timeliness. According to the mean range (3.61-4.00), it is apparent that students perceived AI feedback as somewhat neutral to positive. None of the items reached the mean of 4.5 and higher, demonstrating a neutral and slightly positive perception. The highest mean score (4.00) was scored for promptness and convenience, emphasizing students' favor of AI feedback in terms of its timeliness. The lowest mean score (3.61) was noted in the statement "AI feedback helped improve the essay more than expected", suggesting that students still hold skepticism about AI's effectiveness in enhancing their essay writing quality. The variability (Standard Deviation) values range from 0.742 to 0.917 indicating a moderate variability in responses. The lowest (SD = 0.742) was recorded for clarity of explanations and ease of understanding presenting rather consistent responses. The highest (SD = 0.917) is for detailed explanations, demonstrating more varied opinions among participants on how thorough AI feedback was. In terms of skewness, they range from -0.942 to 0.230 indicating mostly aligned or slightly negatively skewed distributions, except for clarity and helpfulness, which are slightly positive. The most negatively skewed item (Skewness = -0.942) shows that most participants rated it the highest agreeing about AI feedback comprehensiveness. The range of answers is shown through Minimum = 2 and Maximum = 5 scores. It suggests that although there was some variability in answers, the Median = 4 is the same across all statements; hence, most students generally agreed that AI feedback was acceptable but not exceptional.

Table 2

AI Feedback

Item	Mean	Median	SD	Min	Max	Skewness
AI feedback was clear and easy to understand	3.64	4	0.742	2	5	0.230
AI provided detailed explanations for corrections	3.82	4	0.917	2	5	0.124
AI feedback focused on both grammar and content	3.94	4	0.827	2	5	-0.942
AI feedback helped improve the essay more than expected	3.61	4	0.899	2	5	0.0733
Felt confident applying AI feedback	3.70	4	0.883	2	5	-0.211
AI feedback was delivered promptly and conveniently	4.00	4	0.791	2	5	-0.404

In summary, students appreciated AI feedback in its promptness and convenience and its focus on both content and mechanics. However, the lower means for areas like detailed explanations and effectiveness in suggestions on how to improve an essay could be perceived as lacking depth.

4.1.3 Descriptive analysis of teacher feedback. This section of the questionnaire is designed to measure students' perceptions of their teacher feedback in comparison to AI-generated feedback. Overall trends according to the range of mean values (3.79 - 4.55) indicate a general positive perception of teacher feedback across all aspects. The highest mean score (4.55) is attributed to focusing on both grammar and content, demonstrating that students strongly agree that teacher feedback extensively addressed both areas. The lowest mean score (3.79) is for addressing areas not covered by AI, showing that teacher feedback did not always fill gaps left by the AI-generated feedback according to participants. Additionally, SD ranges from 0.699 to 0.992 meaning moderate variability in students' responses. The items with lower scores for detailed explanations (SD = 0.699), clarity (SD = 0.754), and comprehensiveness (SD = 0.794) suggest

more consistent perception of teacher feedback as informative and thorough. The highest SD score for addressing areas not covered by AI, indicating that participants had mixed perceptions of how effectively the teacher addressed the gaps not covered by AI. The negative skewness (skewness = -0.770 to -1.47) highlights that participants rated most of the items highly from 4 to 5. The most negatively skewed item (-1.75) demonstrated that they strongly agreed that feedback provided by their teacher could address both content and grammar areas. On the contrary, the item on addressing areas not covered by AI (skewness = -0.770) is less negatively skewed, indicating more balanced responses, with some students feeling that teacher feedback did not sufficiently cover areas omitted by AI. The range of Minimum/Maximum values is from 1 to 5, the lowest being on the last question about addressing areas skipped by AI. Other items have the minimum score of 2. It emphasizes the other parameters where participants believe that teacher feedback might have not been sufficient enough for addressing areas missed by ChatGPT.

Table 3

Teacher Feedback

Item	Mean	Median	SD	Min	Max	Skewness
Teacher feedback was clear and easy to understand	4.45	5	0.754	2	5	-1.47
Teacher provided detailed explanations	4.36	4	0.699	2	5	-1.23
Teacher feedback focused on both grammar and content	4.55	5	0.794	2	5	-1.75
Teacher feedback helped improve essay more than expected	4.30	4	0.810	2	5	-1.00
Confident applying teacher feedback	4.30	5	0.883	2	5	-1.24
Teacher addressed areas not covered by AI	3.79	4	0.992	1	5	-0.770

Generally, feedback provided by the teacher was perceived more positively and was considered more effective across the same aspects in comparison to AI-generated feedback. Despite a more positive attitude toward instructor feedback, a few participant expressed slight uncertainty with it covering more improvement areas than AI.

4.1.4 Descriptive analysis of comparative perceptions. The main purpose of this section was to compare the two types of feedback according to aspects like usefulness, objectivity, and students' trust. As it is apparent from the statistics, the highest mean scores of 4.30 were achieved on finding that students consider teacher's feedback more useful and more trustworthy along with negative skewness indicating that these two items were consistently rated high (min. = 2; max. = 5). On the contrary, The other two items had significantly lower means (2.48 and 2.85), showcasing less trust and reliability for AI feedback considering it less consistent in comparison to teacher feedback. Moreover, Standard Deviation is high (0.972) for AI being more objective and demonstrates mixed perceptions of AI's objectivity among students. The lower SD (0.795) was for trust in AI's consistency, where students rated teacher feedback as relatively more consistent than AI-generated feedback. These two items had positive skewness of 0.312 and 0.0526 respectively, which shows that students were divided in their responses. While some viewed AI as more consistent, the others did not.

Table 4

Comparative Perceptions

Item	Mean	Median	SD	Min	Max	Skewness
Teacher's feedback more useful than AI	4.30	4	0.810	2	5	-1.38
AI feedback more objective than teacher	2.85	3	0.972	1	5	0.321

Trust in teacher's expertise over AI	4.30	4	0.810	2	5	-1.00
Trust in AI's consistency over teacher	2.48	2	0.795	1	4	0.0526

In brief, participants showed a strong preference for teacher feedback over AI-generated feedback in terms of perceived usefulness and trust to expertise. It was evidenced from higher means and negative skewness. In contrast, they consider AI feedback as less consistent and useful, although positive skewness highlighted some variability in their perceptions of objectivity. Generally, ChatGPT feedback was found to be perceived somewhat objective, conceding to teacher feedback in trustworthiness and practicality.

4.2 Comparative Analysis

4.2.1 Frequency table. This question was analyzed with frequency table analysis to show students' preferences between AI-generated and teacher feedback based on the question:

Table 5

Frequencies of "If I had to choose one feedback method, I would prefer":

If I had to choose one feedback method, I would prefer:	Counts	% of Total	Cumulative %
AI Feedback	3	9.1%	9.1%
Teacher Feedback	30	90.9%	100.0%

An overwhelming number of students 90.9% (30 out of 33) demonstrated a very strong preference for teacher feedback. Conversely, only 3 (9.1%) participants expressed a preference for AI feedback.

4.2.2 Paired t-test analysis. Paired T-test analysis (displayed in Table 6) demonstrates results of comparing participants' perceptions of teacher feedback versus AI-generated feedback across different categories like clarity, detailed explanations, grammar and content focus, and perceived helpfulness in enhancing writing.

1) **Clarity**: with the negative t-value of -4.16 and p-value < .001, it is evident that there is a significant difference in perceived clarity between AI and teacher feedback. Negative t-value indicates teacher feedback being more highly rated in clarity and ease of understanding than AI feedback.
2) **Detailed explanations**: this aspect has a slightly higher t-value of -2.73 and p-value of .010 which is significant. It shows a notable difference in the perceived accuracy of explanations. According to the results, students consider teacher feedback more thorough and explanatory rather than AI-generated feedback.
3) **Grammar and content focus**: t-value = -3.29 and significant p-value = .002 suggest that students found teacher feedback more comprehensive, covering both content and mechanics while checking their written drafts, whereas AI feedback might not have been as effective in this regard.
4) **Impact on writing improvement**: a similar tendency is seen in the results for this comparison with t-value of -3.30 and significant p-value of .002. It implies that students consider teacher to be more helpful in improving their writing skills with feedback provison than ChatGPT.

Table 6

Paired t-test analysis

Comparison	t-value	df	p-value	Interpretation
Clarity: AI vs. Teacher	-4.16	32	< .001	Significant
Detailed explanations: AI vs. Teacher	-2.73	32	0.010	Significant
Focus on grammar & content: AI vs. Teacher	-3.29	32	0.002	Significant
Impact on writing improvement: AI vs. Teacher	-3.30	32	0.002	Significant

Note: H_a μ Measure 1 - Measure 2 ≠ 0

For all four aspects, participants preferred teacher feedback. These findings align with the results of descriptive statistics which also indicates a very strong preference for teacher feedback in terms of clarity, detailed explanations, comprehensiveness, and impact on writing improvement.

4.3 Thematic Analysis

In order to explore reasons behind students' preference for either teacher or AI-generated feedback, thematic analysis was used. Responses were collected and analyzed via an open-ended section of the questionnaire. According to the inductive approach by Braun and Clarke (2006), all responses were interpreted through several stages, including initial coding, deriving themes, revising themes to ensure alignment with the corresponding research question. The data was collected from 33 participants who shared their experiences, reflecting on the perceived usefulness, drawbacks, and value of the feedback they received. All respondents were assigned a number to track their responses and code their personalities. The coded names range from S1, S2, S3, …, to S33. The emerged themes are described in the tables below, supported by direct quotes from the participants.

4.3.1 Comparison of Q1 and Q2.

4.3.1.1 Theme 1. Perceived usefulness of each feedback. The first question "What aspects of the teacher's feedback did you find most helpful for improving your essay?" measured perceived usefulness of teacher feedback from students' perspective. All 33 entries were analyzed and grouped into codes and emerged into themes and subthemes. Subthemes are identified as feedback areas.

Table 7

Perceived usefulness of teacher feedback

Feedback area	Summary	N
Grammar and mechanics	Helped improving grammar, punctuation, and spelling	15
Detailed and specific explanations	Provided constructive and in-depth information with examples	16
Argumentation/structure/thesis development	Strengthened reasoning and expanded arguments	10
Clear and structured feedback	Well-organized and easy-to-understand feedback	10
Clear criteria (IELTS rubric)	Adherence to the assessment criteria of the IELTS Writing rubric	5
Balanced feedback (constructive + positive)	Sandwich technique feedback - encouragement and praise followed by constructive points	4
Promotes independence	Development of autonomous learning in students	3
Accessible and comfortable	Encouraging, warm tone feedback	3

The majority of participants noted teacher feedback usefulness in providing detailed and specific feedback (n=16) that also focused on grammar and mechanics (n=15). Apart from grammar and specificity, students noted that it was clear, structured and helped them improve their essay structure, thesis, and argumentation (n=10). Another frequently mentioned thing is its strict adherence to the IELTS Writing assessment rubric (n=5). The same number of students (n=4) mentioned that teacher feedback was both balanced (sandwich technique - first praise then point out areas for improvement) and effective in language mechanics. The rest highlighted that teacher feedback promotes independence (n=3), creates a more human connection (n=3), is clear to understand (n=2), and provides detailed examples (n=1).

The second question “What aspects of the AI system’s feedback did you find most helpful for improving your essay?”, on the other hand, measured the perceived usefulness of AI-generated feedback according to students.

Table 8

Perceived Usefulness of AI Feedback

Feedback area	Summary	N
Grammar and mechanics	Fixed grammar, punctuation, and spelling	22
Word choice and vocabulary	Suggested better paraphrasing and vocabulary to enhance lexical variety	17
Sentence structure and clarity	Improved sentence flow, clarity, and conciseness	12
Detailed and actionable	Effective revisions based on detailed examples and explanations	9
Objective and accurate	Fair, unbiased, precise	5
Accessible and timely feedback	Fast and easy to use	5
Argument/idea development	Clarifying ideas and providing support in argument development	4
Comfortable	Less intimidating than human feedback	2

From the responses to the second question, the majority of students mention AI’s precision in pointing out grammatical and mechanical errors (n=22). The next most frequently mentioned point is word choice and vocabulary suggestions (n=17). Apart from that, they emphasize improvement of sentence structure (word order, transitions) and clarity of their writing (n=12). Next, the less popular responses consider that AI provides detailed and actionable feedback (n=9) that is also quite unbiased and actionable (n=5). The rest believe it is

very accessible, fast, and timely (n=5), comfortable for shy students (n=2), and helps with argument and idea development (n=4).

Table 9

Comparison of Most Recurring Subthemes

Feedback area	Teacher Feedback	AI Feedback	Frequency
Grammar and mechanics	Emphasized but less frequently. Viewed as accurate and helpful, but not the sole focus.	The most dominant theme, praised repeatedly. It is strong for identifying grammar, punctuation and minor errors.	Shared; AI dominant
Word choice and vocabulary	Helps refine wording in meaningful context; stronger, more accurate words.	Synonym and variety suggestions, but some noted relevance issues.	Shared
Sentence structure and clarity	Provided more nuanced advice on sentence structure and clarity, connected to meaning and purpose.	Praised for improving clarity and sentence conciseness.	Shared
Feedback detail and examples	Personalized and detailed feedback that is supported by clear examples.	Clear and structured, praised for an ability of further prompting.	Shared; teacher dominant
Objective and accurate	Seen as mostly fair. Some students expressed trust for teacher expertise over AI.	Objective and unbiased, but prone to generate false information.	Shared
Accessible and timely feedback	Slow, but more thoughtful and individualized	Immediate and timely, praised by shy students for low-pressure	Shared; AI dominant
Argument/idea development	Helps in clarifying ideas, developing arguments, and restructuring essays.	Rarely mentioned.	Shared; teacher dominant

After analyzing responses from Q1 and Q2, the most recurring themes in both feedback areas emerged. The response trend shows that most of the feedback focus in both feedback types

is mostly shared, but some are stronger in a certain type of feedback. Generally, both teacher and AI-generated feedback are considered to help students equally well across all areas, but grammar and mechanics along with accessibility and timeliness are stronger in AI. On the other hand, teacher feedback dominates in detailed explanations with examples and helpfulness in argument and idea development.

4.3.1.2 Theme 2: Drawbacks of both assessment types. In order to gain deeper insights about some lacking areas of both assessment types, students answered the question "Were there any areas where you felt the feedback from either the teacher or the AI system was lacking? Please explain." Here, they were able to identify the main weaknesses of AI-generated and teacher assessment from their personal experiences. Their responses were analyzed and interpreted in the following table (see Table 10).

Table 10

Comparison of Feedback Drawbacks

Drawback	Teacher Feedback	AI Feedback	Frequency (T/AI)
Timeliness and speed	Feedback was deemed as delayed, not fast enough	Fast and timely feedback but at the cost of quality	3/2
Lack of depth	Some comments were considered: "too general", "not specific", "lack of examples"	Feedback was provided on a surface level and lacks depth in argument logic, essay purpose: "surface understanding", "cannot assess depth"	3/10
Clarity and consistency	It could be occasionally vague or stylistic: "hard to understand", "stylistic focus", "lacked detail"	Inconsistent, unclear, or contradictory: "inconsistent answers", "short answers", "different answers"	4/9

Argument and idea development	Rare comments on originality/creativity improvement: “lacked suggestions for engagement”, “rare suggestions on originality”	Provided little help on deepening and developing ideas: “no idea development”, “focus on grammar, not ideas”	2/8
Contextual understanding	Sometimes missed tone or originality: “didn’t find my tone”, “disencourage complex phrasing”	Lacks emotional capacity, nuanced understanding of writing purpose: “lacks context”, “didn’t get deeper meaning”	2/5
Mechanics and technical accuracy	Overlooked minor punctuation and grammar mistakes: “overlooked mechanical parts”	Sometimes gives inaccurate or incorrect corrections: “wrong answers”, “contradicts itself”, “unreliable”	1/5

According to the results, students were more dissatisfied with AI feedback. Although teacher feedback was praised for clarity and detailed explanations, some students expressed dissatisfaction with these feedback areas. Other weaknesses of teacher feedback from students’ perspective include tardiness (n=3), insufficient help with argument and idea development (n=2), lack of contextual understanding (n=2) and overlooking minor mistakes (n=1). Dissatisfaction is more prevalent for AI-generated feedback for the same categories but slightly different reasons. Many students highlighted the shallow level feedback (n=10), unclear and inconsistent answers (n=9), not enough help with argument and idea development (n=8), lack of contextual understanding and writing intent (n=5), inaccurate or wrong answers (n=5).

4.3.1.3 Theme 3: Prospects of combining both assessment types. The fourth question “How do you think combining teacher and AI feedback could improve the feedback process for students?” was designed to see how students perceive the combination of two feedbacks for learning and assessment purposes. Out of 33 entries, 30 of them positively evaluated the combination of two types of feedback, suggesting that combining these two would create a balanced and

comprehensive feedback system that compliments weaknesses of one another. For example, several quotes that support this idea – S1: “Combining these two methods will give full, well-structured, clear and individual feedback that helps further in each student's academic journey.”, S2: “Yes, combining teacher and AI feedback can make the feedback process more effective for students.”, S4: “I think that combining feedback from the teacher and AI can improve the learning process well…”, S5: “The combination of feedback from teachers and AI allows you to develop as efficiently as possible.”

In particular, students mentioned different feedback areas that teacher and AI assessment can cover while combined. Feedback areas are the subthemes.

Table 11

Teacher Feedback Area Focus

Feature	N	Sample quotes
Content level feedback	17	S2: “... teachers can focus more on giving feedback about content, creativity—areas where human judgment is important.”
Human connection, personalized support	10	S1: “... personally for me teacher's feedback gives more sense of peace.”, S21: “... the teacher gives from heart to heart.”
Critical thinking development	5	S5: “teacher gives a deep assessment of the content side and develops critical thinking.”, S7: “teacher could focus on content, critical thinking, and higher-level writing skills.”, S10: “teachers can focus on the content, critical thinking, and creativity.”
Teacher as main assessor	4	S19: “Additionally, the teacher pointed out arguments and mistakes that AI might not notice.”, S20: “teachers will effectively use AI for perfecting their essay-teaching.”
Motivational aspect	3	S3: “...my teacher motivates me to write essays.”

Holistic feedback and irreplaceable position	3	S3: “Help of a teacher can never be replaced.”

Mainly, teacher feedback is seen as a content and argument idea developer that motivates students to revise their essays and improve writing skills. Many emphasized the importance of human connection and holistic understanding that a human assessor has.

Table 12

AI Feedback Area Focus

Feature	N	Sample quotes
Grammar, technical error correction	20	S2: “AI can help by quickly identifying grammar [mistakes]”, S3: “[AI] will help correct grammar.”, S5: “AI quickly helps to correct technical errors.”
Instant feedback	13	S7: “The AI could quickly catch technical issues…”, S4: “AI compliments it with instant analysis…”
Structure development	8	S3: “[AI can] help with the structure.”, S6: “AI could take care of technical aspects like grammar, spelling, and sentence structure”, S12: “AI can quickly check things like grammar, structure, and facts.”
Unbiased, objective nature	3	S1: “AI (ChatGPT) provides more objective responses.”, S13: “AI can help reduce human errors and bias, offering objective feedback.”
Lowers scores/unnecessary advice	1	S25: “AI… often lowers scores and gives unnecessary advice.”
AI overlooks some aspects	1	S19: “Teacher pointed out arguments and mistakes that AI might not notice.”
Robotic	1	S21: “AI gives feedback like a robot…”

According to students, AI could enhance the feedback process for students by focusing on mechanical aspects of the language like grammar, punctuation, spelling, and sentence structure. They regard it as an unbiased and objective machine that provides instant feedback.

However, some of them expressed dissatisfaction with its unreasonable assessment and suggestions. Apart from that, students consider AI's negligence and its robotic nature as drawbacks.

Altogether, their combination is highly positively perceived. Only 2 students expressed uncertainty about using both types of feedback. S25: "I partially agree", S32: "Maybe yes."

4.3.1.4 Theme 4: Main reasons of feedback preference. In order to understand the main reasons why students favor one feedback over another, the last question "Please explain why you prefer either teacher feedback or AI feedback when working on your writing." was addressed. This question was added to answer the second research question. Since this question was added later to the questionnaire, only 10 responses were collected.

Table 13

Feedback Preference Reasons

Feedback	N	Reasons
Teacher	8	Expertise, personal connection, useful and personalized comments, detailed explanation of mistakes, precise instructions, quick, models answers and examples, clear, effective, aware of individual writing style, trustworthy, knowledgeable
AI-generated	2	Good for fast grammar fixes, precise identification of areas for improvement, various phrasing suggestions, patient, accessible, no limit to prompt

The majority of students (n=8) prefer teacher feedback that supports the results of the frequency analysis. One of the main reasons for favoring teacher feedback over AI-generated is teacher's expertise and personal connection. One of the students mentioned:

> They [teachers] are more familiar with it [writing style] and know what I have to do to improve it. Their comments on what is good and what needs work assist me with my writing. Their comments are more useful than those of AI since they are personal.

They believe that teachers have more personal connection with their students and understand their teaching style better upon working with students for a longer time. Due to the fact that AI might not know individual styles of student writing, it might not understand some mistakes that need to be meaningfully interpreted before correcting. A similar opinion is expressed by another student:

> I like teacher feedback more when I work on my writing. My teacher knows me and my writing style. She can tell me what I do well and what I need to improve. Also, teachers can give clear advice and examples. I sometimes use AI feedback too. It is good for checking small mistakes like grammar or spelling. It is fast and easy to use. But I think teacher feedback is better because it is more personal and helpful for learning.

Although this student sometimes resorts to the help of AI for fixing mechanical problems like grammar or spelling, they still prefer teacher feedback for its personalized approach and clarity of explanations. The rest of the responses express similar ideas. Some of the students also mentioned trustworthiness of teacher feedback: “Also teacher's feedback [is] more … trustworthy than AI feedback.” and “... teacher knows better where to improve or clarify our ideas.”

Preference for AI feedback is justified by its precision, ease of use, limitless prompting, and usefulness for fixing grammatical errors:

> I like AI feedback since AI can pinpoint exactly what sentence or word is out of place and suggest numerous methods of paraphrasing or changing the sentence. In addition, AI is patient, so you can ask as many questions as you want.

Overall, thematic analysis supports the quantitative findings in feedback preferences (significant advantage of teacher feedback) and suggests that teacher and AI feedback types highly differ in the areas of writing they cover. Distinctive features of human feedback are clear, balanced assessment that not only covers mechanical aspects of writing like grammar, punctuation, and spelling but also develops students' argumentation, idea elaboration in detail. It is also personalized and encouraging. However, it is not ideal and might be too slow and inconsistent, not deprived of bias. AI-generated feedback, on the other hand, is objective, accurate, fast, mechanical, somewhat detailed, and well-structured. The drawbacks of AI feedback include surface leveled feedback, lack of depth, inconsistency in answers, and lack of contextual understanding.

CHAPTER 5

Discussion

This chapter discusses the findings of the study in light of the existing literature and the theoretical framework adopted, namely, the Feedback Literacy Framework by Carless and Boud (2018). The discussion will follow the main results obtained through both quantitative and qualitative data analysis, linking them with prior studies reviewed in Chapter 2. It will also reflect on students' preferences and perceived usefulness of teacher and AI-generated feedback, and will consider the implications for EFL teaching and learning, particularly in the context of Kazakhstani higher education.

5.1 Students' Perceptions of Feedback in EFL Writing

One of the first findings from the study was that students hold a generally positive attitude toward receiving feedback on their writing. This is supported by the results of the General Feedback Preferences section of the questionnaire, where high mean scores were reported for statements emphasizing the usefulness of feedback in identifying strengths and weaknesses, providing corrections, and focusing on both content and mechanics. This outcome aligns well with Carless and Boud's (2018) Feedback Literacy Framework, particularly with the dimension of "appreciation of feedback." Students appeared to recognize the importance of feedback in learning and saw it not just as a mark or correction but as a valuable tool for their development.

The high levels of agreement with statements about clarity, usefulness, and actionability also suggest that students are actively engaging in the feedback process. This reflects the second and fourth dimensions of the framework: "making judgments" and "taking action", as students not only accepted feedback but showed signs of being willing to apply it to improve their future

writing. The idea that effective feedback helps learners to develop these skills is strongly supported by previous research (Hyland & Hyland, 2006; Gibbs & Simpson, 2005), and the results of this study confirm that this holds in the EFL context as well.

5.2 Comparative Perceptions of Teacher and AI Feedback

The central aim of the study was to compare how students perceive feedback from their teacher versus that provided by an AWE system, specifically, ChatGPT. Results from both the quantitative and qualitative sections show a clear preference for teacher feedback. The frequency table analysis showed that 90.9% of participants (30 out of 33) preferred teacher feedback when asked to choose just one source.

This is consistent with the findings of Zhai and Ma (2021), who concluded that while students valued the efficiency of AWE systems, they still preferred teacher feedback for its depth, personalization, and emotional connection. In this study, teacher feedback was described by participants as more detailed, supportive, and easier to understand. Paired t-test analysis confirmed statistically significant differences in clarity, comprehensiveness, content and grammar focus, and helpfulness for improving writing, all favoring teacher feedback. These results support earlier findings by Cheng and Zhang (2024) and Sari and Han (2024), who also noted that teacher feedback was more influential in students' writing development, especially when it comes to higher-order writing skills like argumentation, structure, and thesis clarity.

Interestingly, students acknowledged that while AI feedback had certain benefits, it lacked the nuance and human touch of their instructor's comments. Some participants mentioned that AI-generated comments felt generic or lacked a deep understanding of their writing intent. This supports the concerns raised by Swiecki et al. (2022) and Fu et al. (2024) regarding the limitations of AI in recognizing tone, context, and emotional content in writing. These aspects

are especially important in argumentative or opinion-based essays where logic and persuasion matter as much as grammatical accuracy.

5.3 Strengths and Weaknesses of Teacher and AI Feedback

When looking at the open-ended answers in the questionnaire, some clear patterns began to show up in how students described the strengths and weaknesses of both teacher and AI feedback. When it comes to teacher feedback, a lot of students said they liked how it was balanced, teachers gave them both praise and criticism in a way that felt fair and motivating. A few mentioned that their teachers followed the IELTS criteria clearly, which helped them understand what was expected. But the biggest thing students appreciated was that the teacher noticed not only small grammar mistakes but also deeper things like structure, argument strength, and how ideas were developed throughout the essay.

This kind of detailed and encouraging feedback seems to make a big difference in how students feel about their writing. Some said that their teacher's comments gave them confidence or made them want to improve more, which reflects the emotional side of learning that's often hard to measure. In terms of the feedback literacy model, this would connect with "managing affective reactions," where learners not only read the feedback but also respond to it emotionally and use it as motivation. AI, at least for now, seems to lack this emotional connection.

On the other hand, many students still found value in the AI-generated feedback. A lot of them said it helped fix basic errors quickly, things like grammar, punctuation, and word choice. Some students even said they liked using ChatGPT as a sort of "first checker," just to clean up small issues before working on bigger revisions. In that sense, the AI tool worked kind of like a proofreading assistant, which saved them time and made them feel more confident about the mechanics of their writing. These opinions are very similar to what other studies have found, for

example, Wilson et al. (2021) and Wang (2020) noted that students liked the speed and convenience of AWE tools.

However, the limitations of AI were also brought up by quite a few participants. Some said that the feedback from ChatGPT was too general or even confusing. There were cases where students felt the AI either misunderstood what they wrote or gave suggestions that didn't fit their topic or intention. A few mentioned they got mixed messages from the AI, especially when it came to organizing their arguments or fixing vague sentences. This kind of issue, where the AI can point out something is "unclear" but not explain how to fix it, is frustrating for learners. It matches what Gayed et al. (2022) and Dwivedi et al. (2023) wrote about: that AI still can't fully "read between the lines" or understand the deeper meaning of a text. It's good with rules and patterns, but it doesn't think like a teacher who knows the student and their writing style.

5.4 Combining AI and Teacher Feedback: What Students Think

One of the more surprising and positive things that came out of the thematic analysis was how many students liked the idea of combining both types of feedback. Even though most preferred the teacher's feedback, they didn't reject the AI; many said the two could work well together if used in the right way. A common idea was that AI could handle the technical part—checking grammar, fixing awkward phrases, giving quick suggestions—while the teacher could focus on bigger things, like logic, structure, and how the student's ideas are developing.

Several students described this as a "best of both worlds" situation. They said that AI feedback was fast and sometimes helpful for first drafts, but only the teacher could understand their intent, tone, and personal writing style. One participant even mentioned that their teacher gave advice "from the heart," which shows how important the human side of learning still is.

Others said that combining feedback would help them avoid missing small mistakes while still improving the quality and structure of their arguments.

This shows that students are not against AI, but they see it more as a helpful tool than a replacement for their teacher. Most of them made it clear that while AI has its place, it still needs to be guided by a teacher to be fully effective. These views support earlier studies (Zhai & Ma, 2021; Wang, 2020), where students said AI was useful but couldn't replace the experience, understanding, and personal touch that only a real instructor can provide.

5.5 Implications for EFL Writing Instruction in Kazakhstan

This study contributes to the limited research on AI in EFL assessment in Kazakhstan. As highlighted in the literature review, while international applications of AI in education are rapidly expanding, its practical use in Kazakhstani universities remains scarce and mostly theoretical (Kaldygozova et al., 2025; Dussembinova & Ramazanova, 2024). The results of this research suggest that AI tools like ChatGPT have potential for local implementation, especially in writing courses where feedback demand is high.

However, it also emphasizes that such tools must be used with caution. Teachers need proper training to integrate AWE tools effectively, and ethical considerations such as data privacy, bias, and authorship need to be addressed. Students should also be educated about the role of AI in learning, not as a replacement for effort or thinking, but as a support tool to aid revision and learning.

The strong preference for teacher feedback observed in this study also highlights the continued importance of human interaction in education. In the Kazakhstani context, where teacher-student relationships often play a vital role in student motivation and performance, AI will likely serve best as a supplementary tool rather than a standalone solution.

5.6 Final Thoughts

To sum up, this study confirmed that students find feedback essential for improving their writing and that they deeply value the support and clarity provided by their teachers. While AI feedback is seen as helpful for quick and basic corrections, it is not yet capable of replacing the human element in education, especially when it comes to guiding students in idea development, logic, and expression. Combining both sources of feedback, however, shows great promise and was welcomed by almost all participants.

Nevertheless, the potential prospects of AI integration in the educational process and writing assessment in particular can positively influence student learning outcomes and reduce the overwhelming workload for teachers. If used thoughtfully, with awareness of its limitations, ethical considerations, and the need for teacher oversight, AI could help create a more balanced and effective feedback system. The key lies in collaboration: not choosing between teacher and AI, but understanding how both can work together to support student success.

CHAPTER 6

Conclusion

As the field of AI technology advances, its use in all spheres of life becomes unavoidable and AI in education provides more opportunities to innovate pedagogical process. From the results of this study, students expressed a strong preference for teacher feedback. Although the majority of students still view AI as a great supplementary tool for learning, they believe it is impossible to completely replace the human connection and more nuanced approach of a human assessor based on interpersonal connection and years of experience. The majority of students 30 out of 33 prefer teacher feedback over AI. The most optimized feedback was seen in the combination of AI and teacher feedback. In this case scenario, AI works best for quickly identifying linguistic mistakes, while teacher focuses on more contextual aspects of student writing, saving time and reducing the load. The majority of students still hold some distrust towards the effectiveness of AI in higher-order thinking analysis and its robotic algorithms. The reasons behind their preferences over teacher feedback are specificity, personalization, in-depth analysis, and expertise of an educator. They align with the advantages students previously highlighted. The reasons behind choosing AI feedback by comparatively small number of students are prompting, accuracy, objectivity, timeliness, and mechanical aspects. Another no less important point was unjudgmental nature of AI, which provides a comfortable environment for shy or reserved students who are afraid of being judged by the teacher.

These in-depth insights provide a first step towards understanding how students actually experience integration of AI in practice. While AI is being globally promoted, this technology is not yet widely applied across Kazakhstani higher education institutions. The feedback from this study suggest that students are open to using AI, but only if it’s implemented with proper

guidance, support, and ethical consideration. Teachers would need training, and students would need to be taught how to use these tools responsibly, not to rely on them blindly, but to see them as part of a bigger learning process.

6.1 Implications and Recommendations

The findings of this study suggest that integrating AI into writing instruction should be done carefully and purposefully. AI tools can be helpful, but they must be framed correctly. First, educators should clearly explain to students what AI can and cannot do. Second, teachers themselves need digital training to learn how to manage AI tools in their classes in a way that benefits learning rather than replaces it. Third, any use of AI in assessment or feedback must take into account ethical issues, such as student data privacy and authorship.

The hybrid feedback model proposed by students in this study could be especially useful in contexts where class sizes are large, and teachers are overwhelmed. Allowing students to use AI for preliminary feedback, followed by more personalized teacher comments, could improve the quality of feedback without increasing teacher workload. This approach can also help students develop greater responsibility for their own learning, which is one of the key goals of feedback literacy.

6.2 Limitations of the Study

Although the results of this study provide useful insights, it is also important to acknowledge its limitations. The study was conducted in one institution, with a relatively small number of participants, which means the findings cannot be generalized to all contexts. Also, it focused on only one AWE tool - ChatGPT, while other platforms may offer different types or quality of feedback. Another limitation is that the study explored student perceptions, not actual writing improvement over time, which would require a longer-term study design.

Future research could involve more participants from different universities and educational backgrounds. It might also compare different AI feedback tools or explore the long-term impact of combined AI and teacher feedback on students' writing performance. Another possible direction is to investigate how students use the feedback they receive; whether from teachers or AI, and how that affects their actual revisions and final drafts.

6.3 Final Reflection

To conclude, the findings of this study confirm that while AI feedback can serve as a helpful resource in the EFL writing classroom, it cannot replace the thoughtful, nuanced, and supportive feedback that teachers provide. Students value the emotional and intellectual connection with their instructors and view AI tools as something that can support, but not fully guide, their learning process. The most promising direction, therefore, is not to choose between human and machine, but to bring them together in ways that complement each other. With careful planning, ethical use, and proper training, AI can become a helpful addition to EFL education, improving the feedback process and supporting students on their path to becoming more confident and capable writers.

References

Al-Gharabally, M. (2015). The writing difficulties faced by L2 learners and how to minimize them. *International Journal of English Language and Linguistics Research, 3*(5), 42–49.

Aydin, S. (2012). Factors causing demotivation in EFL teaching process: A case study. *Qualitative Report, 17*, 101.

Balfour, S. P. (2013). Assessing Writing in MOOCs: Automated Essay Scoring and Calibrated Peer Review™. *Research & Practice in Assessment, 8*, 40–48.

Beatty, K. (2010). *Teaching & Researching: Computer-Assisted Language Learning (2nd ed.)*. Routledge. https://doi.org/10.4324/9781315833774

Black, P., & Wiliam, D. (2009). Developing the theory of formative assessment. *Educ Asse Eval Acc, 21*, 5–31. https://doi.org/10.1007/s11092-008-9068-5

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology, 3*(2), 77–101. https://doi.org/10.1191/1478088706qp063oa

Brooke, J. (1996). SUS-A quick and dirty usability scale. *Usability evaluation in industry, 189*(194), 4–7.

Carless, D., & Boud, D. (2018). The development of student feedback literacy: enabling uptake of feedback. *Assessment & Evaluation in Higher Education, 43*(8), 1315–1325.

Çankaya, P. (2018). Demotivation factors in foreign language learning. *Journal of Foreign Language Education and Technology, 3*(1), 1–17.

Chapelle, C. A. (2009). The Relationship Between Second Language Acquisition Theory and Computer-Assisted Language Learning. *The Modern Language Journal, 93*, 741–753. https://doi.org/10.1111/j.1540-4781.2009.00970.x

Cheng, X., & Zhang, L. J. (2024). Examining Second Language (L2) Learners' Engagement with AWE-Teacher Integrated Feedback in a Technology-Empowered Context. *Asia-Pacific Education Researcher*. https://doi.org/10.1007/s40299-024-00877-8

Cloete, A. L. (2017). Technology and education: Challenges and opportunities. *HTS: Theological Studies, 73*(3), 1–7.

De la Vall, R. R. F., & Araya, F. G. (2023). Exploring the benefits and challenges of AI-language learning tools. *Int. J. Soc. Sci. Humanit. Invent, 10*, 7569–7576.

Dussembinova, Z., & Ramazanova, K. (2024). Artificial intelligence – opportunities and prospects for application in higher education. *Вестник Alikhan Bokeikhan University, 2*(61).

Dwivedi, Y. K., Kshetri, N., Hughes, L., Slade, E. L., Jeyaraj, A., Kar, A. K., ... & Wright, R. (2023). Opinion Paper: "So what if ChatGPT wrote it?" Multidisciplinary perspectives on opportunities, challenges and implications of generative conversational AI for research, practice and policy. *International Journal of Information Management, 71*, 102642.

Fathi, J., Greenier, V., & Derakhshan, A. (2021). Self-efficacy, reflection, and burnout among Iranian EFL teachers: the mediating role of emotion regulation. *Iranian Journal of Language Teaching Research, 9*(2), 13–37.

Fu, Q. K., Zou, D., Xie, H., & Cheng, G. (2022). A review of AWE feedback: types, learning outcomes, and implications. *Computer Assisted Language Learning, 37*(1–2), 179–221. https://doi.org/10.1080/09588221.2022.2033787

Gao, J. W., & Ma, S. (2020). Instructor feedback on free writing and automated corrective feedback in drills: intensity and efficacy. *Language Teaching Research*. Advance online publication. https://doi.org/10.1177/1362168820915337

Gayed, J. M., Carlon, M. K. J., Oriola, A. M., & Cross, J. S. (2022). Exploring an AI-based writing Assistant's impact on English language learners. *Computers and Education: Artificial Intelligence, 3*, 100055.

Gibbs, G., & Simpson, C. (2005). Conditions under which assessment supports students' learning. *Learning and Teaching in Higher Education,* (1), 3–31.

Glover, C., & Brown, E. (2006). Written feedback for students: too much, too detailed or too incomprehensible to be effective? *Bioscience Education, 7*(1), 1–16.

Haddad, W. D., & Draxler, A. (2002). The dynamics of technologies for education. In Technologies for education: Potentials, parameters, and prospects (pp. 2–17).

Hyland, K., & Hyland, F. (2006). Feedback on second language students' writing. *Language Teaching, 39*(2), 83–101. https://doi.org/10.1017/S0261444806003399

Kaldygozova, S., Shakenova, M., & Jilkishiyeva, M. (2025). Application of artificial intelligence in the system of monitoring and management of education quality in Kazakhstan. *The Bulletin, 414*(2), 152–173.

Kannan, J., & Munday, P. (2018). New trends in second language learning and teaching through the lens of ICT, networked learning, and artificial intelligence.

Kirkpatrick, A., & Schaller-Schwaner, I. (2022). *English as a lingua franca*. In Handbook of Practical Second Language Teaching and Learning (pp. 97–113). Routledge.

Kong, Y., Molnár, E. K., & Xu, N. (2022). Pre-and in-service teachers' assessment and feedback in EFL writing: Changes and challenges. *Sage Open, 12*(3), 21582440221126672.

Lai, J. W., & Bower, M. (2019). How is the use of technology in education evaluated? A systematic review. *Computers & Education, 133*, 27–42.

Levy, M. (1997). *CALL: Context and conceptualisation*. Oxford: Oxford University Press.

Linderbaum, B. A., & Levy, P. E. (2010). The development and validation of the Feedback Orientation Scale (FOS). *Journal of Management, 36*(6), 1372–1405.

Mason, R. O. (2017). Four ethical issues of the information age. *In Computer ethics* (pp. 41–48). Routledge.

Oliver, R. L. (1980). A cognitive model of the antecedents and consequences of satisfaction decisions. *Journal of Marketing Research, 17*(4), 460–469.

Oxford, R. L. (2014). Linking theories of learning with intelligent computer-assisted language learning (ICALL). *In Intelligent Language Tutors* (pp. 377–388). Routledge.

Sari, E., & Han, T. (2024). The impact of automated writing evaluation on English as a foreign language learners' writing self-efficacy, self-regulation, anxiety, and performance. *Journal of Computer Assisted Learning*. https://doi.org/10.1111/jcal.13004

Swiecki, Z., Khosravi, H., Chen, G., Martinez-Maldonado, R., Lodge, J. M., Milligan, S., Selwyn, N., & Gašević, D. (2022). Assessment in the age of artificial intelligence. *Computers & Education: Artificial Intelligence, 3*, 100075. https://doi.org/10.1016/j.caeai.2022.100075

The Jamovi Project. (2024). *jamovi* (Version 2.6) [Computer Software]. Retrieved from https://www.jamovi.org

Van Ryzin, G. G. (2013). An Experimental Test of the Expectancy-Disconfirmation Theory of Citizen Satisfaction. *Journal of Policy Analysis and Management, 32*(3), 597–614. https://doi.org/10.1002/pam.21702

Wang, Z. (2020). Computer-assisted EFL writing and evaluations based on artificial intelligence: a case from a college reading and writing course. *Library Hi Tech.* Advance online publication. https://doi.org/10.1108/lht-05-2020-0113

Warschauer, M., & Healey, D. (1998). Computers and language learning: an overview. *Language Teaching, 31*(2), 57–71. https://doi.org/10.1017/s0261444800012970

Watling, C. J., & Ginsburg, S. (2019). Assessment, feedback and the alchemy of learning. *Medical Education, 53*(1), 76–85.

Wilson, J., & Roscoe, R. D. (2020). Automated writing evaluation and feedback: Multiple metrics of efficacy. *Journal of Educational Computing Research, 58*(1), 87–125.

Wilson, J., Ahrendt, C., Fudge, E. A., Raiche, A., Beard, G., & MacArthur, C. A. (2021). Elementary teachers' perceptions of automated feedback and automated scoring: Transforming the teaching and learning of writing using automated writing evaluation. *Computers in Education, 168*, 104208. https://doi.org/10.1016/J.COMPEDU.2021.104208

Zhai, N., & Ma, X. (2021). Automated writing evaluation (AWE) feedback: a systematic investigation of college students' acceptance. *Computer Assisted Language Learning, 35*(9), 2817–2842. https://doi.org/10.1080/09588221.2021.1897019

Zhai, X. (2022). ChatGPT user experience: Implications for education. *SSRN*. https://ssrn.com/abstract=4312418

Zupanc, K., & Bosnic, Z. (2016). Advances in the field of automated essay evaluation. *Informatica, 39*(4).

Appendix A

Informed consent letter

Consent Form for teachers' participation in a research study entitled as **"AI-Assisted Assessment vs Teacher's Assessment in EFL Writing Course-".**

This study is being conducted as a part of educational research for SDU University.

Dear Participant,

You are invited to contribute to this survey. Thank you in advance for taking the time to read this information leaflet. Please spend a few moments to fill it in and send it back to me by **email.**

Purpose: The purpose of this study is twofold: first it seeks to compare the performance of EFL students in writing with two different assessment methods - with the help of AI and instructor's assessment. Second, it aims to identify which method is more effective in improving students' writing skills.

Participant Selection: You were selected as a possible participant in this study because of your enrollment in a reading and writing course at Suleyman Demirel University. The student questionnaire is estimated to be completed by 50 students. Your participation would be greatly appreciated.

Please note, that this is not a test, so there are no right or wrong answers. I am interested in your personal opinion and would like to request that you answer honestly.

Confidentiality: Any information obtained for this study through which you might be identified will remain confidential and will be disclosed only with your permission. Your answers will be anonymized in the reporting of the results. Raw data will be maintained in a secure location and no identifying information will be used in the final dissertation or subsequent publications. No one, other than my supervisor, will have access to the raw data. All audio recordings, video recordings, and other documents, with the exception of the consent forms, will not be shared with any third party at the conclusion of this study.

Refusal/Withdrawal: Participation in this study is entirely voluntary. Your decision, whether or not you decide to participate, will not affect your future relations with SDU University. Furthermore, you are free to withdraw from the study at any time without any further commitments.

Potential risks: Despite efforts to maintain confidentiality, there is a minimal risk that some information about participants' writing proficiency and performance may be disclosed during the assessment process. Engaging in assessments, particularly if critical feedback is provided, may lead to emotional discomfort or stress for some participants. There is a risk of bias in the assessment process, particularly if participants are aware of the assessment method (teacher assessment or AWE) before completion.

Benefits: Participation in the study provides an opportunity for participants to receive detailed feedback on their writing skills. This feedback aims to contribute to their learning and skill development in the area of writing. You will also gain insights into your writing proficiency through both teacher assessment and AWE system assessment. This understanding can be valuable for your academic and personal growth. Your participation contributes to educational research, helping to advance knowledge in the field of assessment methods.

Contact: If you have any questions about this study, you can contact me, Primary Researcher(s): Fazilova Alina
or my Scientific Supervisor: Smakova Kymbat, [PhD., Assist. prof. at SDU University, Kazakhstan] +1234567890 (cell phone), or at the office G318/5.

Appendix B

Questionnaire

Dear Participant,

Thank you for participating in the research, "A Comparative Study of AI-assisted Assessment and Teacher Assessment in an EFL Writing Course." Your feedback is valuable in helping us understand the effectiveness of different assessment methods in enhancing your writing skills. The information you provide will contribute to the ongoing improvement of our EFL writing courses.

Instructions:

1. Please read each question carefully.
2. Select the response that best represents your opinion and experience.
3. Your responses will remain confidential, and the data will be used for research purposes only.

NUMBERS from 1-5 represent the level of your agreement. Check the corresponding box.

1 - STRONGLY DISAGREE

2 - DISAGREE

3 - NEUTRAL

4 - AGREE

5 - STRONGLY AGREE

Thank you for taking the time to participate in our study. Your input is crucial in shaping the future of EFL writing assessments. If you have any questions or concerns, please don't hesitate to contact the researcher - Alina Fazilova (231302007@sdu.edu.kz)

SECTION I. General Feedback Preferences

1. I actively seek detailed feedback to improve my writing skills.
2. I find feedback valuable for identifying my strengths and weaknesses in writing.
3. I prefer feedback that provides specific examples or corrections.
4. Constructive criticism motivates me to work harder on improving my essays.
5. I am more likely to use feedback when it is easy to understand and is practical.
6. I believe feedback should focus on both content (ideas, arguments) and mechanics (grammar, style).

SECTION II. Comparative Evaluation of Teacher and AI Feedback

a. AI feedback

1. The feedback provided by the AI system was clear and easy to understand.
2. The AI system provided detailed explanations for its corrections or suggestions.
3. The AI feedback focused on both grammar and content effectively.
4. The AI system's feedback helped me improve my essay more than I expected.
5. I felt confident applying the feedback provided by the AI system to revise my essay.
6. The AI feedback was delivered promptly and conveniently.

b. Teacher Feedback

1. The feedback provided by my teacher was clear and easy to understand.
2. My teacher's feedback gave detailed explanations for its corrections or suggestions.
3. The teacher feedback focused on both grammar and content effectively.
4. The teacher feedback helped me improve my essay more than I expected.
5. I felt confident applying the feedback provided by my teacher to revise my essay.

6. The teacher feedback addressed areas the AI system did not cover.

SECTION III. Comparative Perception

1. I found the teacher's feedback more useful than the AI system's feedback.
2. I found the AI system's feedback more objective than the teacher's feedback.
3. I trust the teacher's expertise over the AI system's algorithms.
4. I trust the AI system's consistency over the teacher's subjective interpretation.
5. If I had to choose one feedback method, I would prefer:
    A) Teacher Feedback
    B) AI Feedback

SECTION IV. Open-ended questions

1. What aspects of the teacher's feedback did you find most helpful for improving your essay?
2. What aspects of the AI system's feedback did you find most helpful for improving your essay?
3. Were there any areas where you felt the feedback from either the teacher or the AI system was lacking? Please explain.
4. How do you think combining teacher and AI feedback could improve the feedback process for students?

Appendix C

Ethical Committee Research Approval