DESIGN OF AN AI-SUPPORTED PROTOTYPE APPLICATION FOR ASSESSING WRITING SKILLS IN TURKISH AS A FOREIGN AND SECOND LANGUAGE[1]

Kadir COŞKUN[2]

Mustafa DURMUŞ[3]

Abstract: *Recent advancements in artificial intelligence technologies have introduced innovative educational approaches, making learning processes richer and more accessible. This study introduces an AI-supported web application that assesses the writing of learners of Turkish as a foreign or second language and functions seamlessly in hybrid and distance-learning settings. Within the developed application, instructors can create writing activities for learners in their classes, and learners can create texts using these activity guidelines and share them with instructors. Learners' written products are scored based on the activity guidelines using an analytical rubric, developed through artificial intelligence, covering form, content, and grammar. The rubric scores and artificial intelligence feedback obtained from the evaluation are presented to learners, who can revise their texts in line with the feedback they receive. Instructors can mark errors in completed learner texts, add explanations, and contribute to the system with rubric scores. In addition, the application was developed to create a dataset that instructors and institution managers can use when planning their teaching processes. It provides the compilation and recording of indicators such as the number of errors in learner texts, learner vocabulary, writing speed, word and sentence lengths on a process-based basis. It is believed that the effectiveness and sustainability of this model, which differs from current practices by aiming to provide a learning environment where learners and instructors come together to teach Turkish as a foreign and second language by taking advantage of the potential offered by artificial intelligence, will be increased with future research.*

[1]This article is derived from the doctoral dissertation conducted by the first author under the supervision of the second author and was presented at the 5th Student Symposium on Teaching Turkish as a Foreign Language.

[2] Res. Asst. Kastamonu University, Faculty of Education, Department of Turkish and Social Sciences Education, Kastamonu/TÜRKİYE, e-mail: kcoskun@kastamonu.edu.tr, ORCID: 0000-0001-5771-9954

[3] Prof. Dr., Hacettepe University, Institute of Turkish Studies, Ankara/TÜRKİYE, e-mail: mustafadurmus@hacettepe.edu.tr, ORCID: 0000-0002-5337-3113

Yabancı Dil ve İkinci Dil Olarak Türkçe Öğretiminde Yazma Becerisinin Değerlendirilmesine Yönelik Yapay Zekâ Destekli Bir Uygulama Prototipinin Tasarımı

Öz: *Son yıllarda yapay zekâ teknolojilerindeki hızlı gelişmeler, eğitim alanında yenilikçi yaklaşımların ortaya çıkmasını sağlamış ve öğrenme süreçlerini daha zengin ve erişilebilir hâle getirmiştir. Bu doğrultuda gerçekleştirilen çalışmanın amacı, Türkçeyi yabancı veya ikinci dil olarak öğrenenlerin yazılı ürünlerini değerlendirmeye yönelik yapay zekâ destekli bir web tabanlı uygulama modeli geliştirmektir. Geliştirilen uygulama içerisinde öğreticiler kendilerine ait sınıflarda yer alan öğreniciler için yazma etkinlikleri oluşturabilmekte, öğreniciler ise bu etkinlik yönergelerine uygun olarak metinler oluşturup bunları öğreticilerle paylaşabilmektedir. Öğreniciler tarafından oluşturulan yazılı ürünler, yapay zekâ tarafından çalışma kapsamında geliştirilmiş olan biçim, içerik ve dilbilgisi boyutlarını kapsayan analitik rubrik çerçevesinde etkinlik yönergeleri dikkate alınarak puanlanmaktadır. Değerlendirme sonucunda elde edilen rubrik puanları ve yapay zekâ geri bildirimleri öğrenicilere sunulmakta, öğreniciler ise aldıkları geri bildirim doğrultusunda metinlerinde revizyon yapabilmektedirler. Öğreticiler, tamamlanmış öğrenici metinlerindeki hataları işaretleyip açıklamalar ekleyebilmekte ve rubrik puanlamalarıyla sisteme katkı sunabilmektedir. Ayrıca öğreticilerin ve kurum yöneticilerinin öğretim süreçlerini planlarken kullanabilecekleri bir veri kümesinin oluşturulması amacıyla geliştirilen uygulama; öğrenici metinlerindeki hata sayısı, öğrenici söz varlığı, yazma hızı, sözcük ve cümle uzunlukları gibi göstergelerin süreç bazlı olarak derleyerek kayıt altına alınması sağlamaktadır. Yapay zekânın sunduğu potansiyelden yararlanarak Türkçenin yabancı ve ikinci dil olarak öğretiminde öğrenici ve öğreticilerin bir arada olduğu bir öğrenme ortamı sağlamayı amaçlayarak mevcut uygulamalardan farklılaşan bu modelin etkinliğinin ve sürdürülebilirliğinin ileride gerçekleştirilecek araştırmalarla artırılacağına inanılmaktadır.*

Anahtar Sözcükler: *Yabancılara Türkçe öğretimi, otomatik metin değerlendirme, dil değerlendirmesinde yapay zekâ, yazma becerisi.*

Introduction

The development of technology and the changing educational needs that come with it have made measuring learner success more effective and accessible. This has notably reshaped the role of web-based applications in education and initiated a transformation. These applications focus on critical functions such as evaluating learner performance, tracking progress, providing feedback, and personalizing the learning experience, and

offer a more dynamic educational environment than traditional measurement and evaluation methods. In order to overcome the limitations of traditional methods and make the most of the innovative opportunities offered by technology in education, integrating web-based applications into the educational process is of great importance. In this context, storing learner data on a central platform not only facilitates measurement and evaluation processes but also allows for the analysis of learning processes according to individual needs and the development of personalized pedagogical strategies (Bennett, 2011; Black & Wiliam, 1998). With this approach, web-based systems not only collect data but also become a powerful tool for the holistic evaluation and development of the educational process by ensuring continuous improvement through the obtained data. Thus, while learners' learning processes are supported with a more interactive and data-based approach, academic success and individual development can be monitored in a more transparent and measurable way. It is known that learners receiving feedback from different sources make the learning process more effective (Shute, 2008; Uyar, 2016). Web-based applications offer learners the opportunity to receive feedback not only from instructors but also from peers and different field experts. In addition, today's development of Natural Language Processing (NLP) models and innovations in artificial intelligence technology allow learners to evaluate their written products more effectively and to provide feedback (Koe et al., 2024; OpenAI et al., 2023).

However, some studies also include negative statements about online assessments. In some studies, factors such as the lack of technological competence of instructors and learners, the preference for using paper and pencil, and concerns arising from online exams emerge as the reasons for reservations against these applications (Khan & Khan, 2019; Ocak & Karakuş, 2021; Zou et al., 2021). This situation shows that the individual characteristics of users are an important factor in the use of web-based applications. In addition, web-based applications have been addressed in various studies in the literature within the scope of behaviors against academic honesty, such as difficulties in ensuring exam security and detecting plagiarism in assignments.

1.1. Academic Integrity and Web-Based Applications

The concept of academic honesty refers to the ethical rules that researchers, learners, and educational institutions follow in the production of knowledge, the sharing of the knowledge produced, and the measurement of success in the learning process.

The most common violation of academic rules in exams conducted to measure learner success in educational institutions is learners cheating during the exam (Cizek & Wollack, 2017, p. 3). In studies conducted on cheating in exams, it has been observed that individual factors such as age, gender, intelligence, academic success, department, and anxiety levels of individuals do not have a significant effect on cheating behavior (Imran & Nordin, 2013; Polat, 2017). In addition, it has been stated that the importance of the exam and inadequate supervision during the exam are factors that can be effective in cheating behavior (King et al., 2009; Maramark & Maline, 1993). In addition, learners' stress levels and value judgments affect cheating behavior (Barnett & Dalton, 1981).

The widespread use of the internet has played a role in the increase of plagiarism, one of the behaviors that violate academic honesty, in learner assignments (Howard & Davies, 2009; Scanlon & Neumann, 2002). With the Covid-19 pandemic, assignment submissions and exams in many educational institutions around the world are being carried out over the internet, and the necessary checks in this process cannot be carried out due to systemic deficiencies or instructors' inadequate use of technology, and the

number of behaviors that violate academic honesty has increased even more (Abdelrahim, 2021; Comas-Forgas et al., 2021; Janke et al., 2021). The frequent encounters with behaviors that violate academic honesty in online applications with Covid-19 have led researchers to exam monitoring systems. In this direction, some studies have used face tracking systems (Kasinathan et al., 2022), and in some studies, microphones and users' devices have been tracked in addition to face recognition (Atoum et al., 2017). However, some studies on surveillance applications have shown that these applications have a negative impact on learner achievement (Alessio et al., 2018). In addition, it is discussed whether these applications, which record images and audio in the learner's environment and track the learner's device and/or internet browser activity, violate the learner's privacy and therefore to what extent they are ethical (Coghlan et al., 2021). Some researchers argue that well-designed courses inherently reduce opportunities for cheating, thus minimizing the necessity for invasive monitoring tools (Rowe, 2004).

1.2. Web-Based Applications for Assessing and Developing Writing Skills

Research highlights the positive impact of online tools on learners' writing skills (Çangal et al., 2025; Sakkir & Dollah, 2019; Shih, 2011). Some studies have explored collaborative writing through social networks and wiki-based platforms, assessing skill development via pre-test and post-test methodologies.

While these applications benefit learners, they do not provide instructors with detailed metrics such as lexical diversity, sentence length, textual complexity, revision frequency, or writing time. Consequently, dedicated tools for writing assessment are essential. Unlike social networking and wiki-based platforms, specialized applications allow for a more granular analysis of learner texts. Table 1 summarizes existing writing assessment tools and their functionalities.

Tool Type	Tool Name	Functions	Turkish Language Support
Spelling and Grammar Checkers	WhiteSmoke	Detects and corrects spelling, grammar, and punctuation errors; provides translation and plagiarism detection.	No
	Ginger Grammar Check	Detects and corrects spelling, grammar, and punctuation errors.	No
	Scribens	Detects and corrects spelling, grammar, and punctuation errors.	No
	LanguageTool	Detects and corrects spelling, grammar, and punctuation errors.	No
	PaperRater	Detects and corrects spelling, grammar, and punctuation errors, including plagiarism detection.	No
	GradeProof	Uses AI tools to detect and correct spelling, grammar, and punctuation errors.	No
	Duplichecker	Detects spelling and grammar errors and provides plagiarism detection.	Yes

Tool Type	Tool Name	Functions	Turkish Language Support
Style Suggestion Tools	Hemingway Editor	Helps improve text readability.	No
	TextExpander	Creates user-customizable templates for specific writing purposes.	No
AI-Based Tools	Google Bard	Detects spelling, grammar, and punctuation errors; suggests stylistic improvements; translates text.	Yes
	ChatGPT	Detects spelling, grammar, and punctuation errors; suggests stylistic improvements; translates text.	Yes
	QuillBot	Rewrites texts, translates languages, and generates various types of creative content.	Yes
	Wordtune	Rewrites texts, suggests stylistic improvements, and provides translations.	No
	Criterion	Detects spelling, grammar, and punctuation errors; provides feedback on learner writing and assigns holistic scores.	No
	E-rater	AI tool used in TOEFL iBT assessments; detects spelling, grammar, readability, and style errors.	No
	Grammarly	Detects and corrects spelling, grammar, punctuation, readability, and style errors.	No
	ProWritingAid	Detects and corrects spelling, grammar, punctuation, readability, and style errors.	No
	Writefull	Rewrites texts in an academic tone and provides style and academic writing suggestions.	Yes

Table 1. *Tools for Assessing Written Productions (as of April 2025)*

A close inspection of Table 1 shows a pronounced gap in rule-based Turkish coverage. All of the mainstream spelling-and-grammar checkers—WhiteSmoke, Ginger, Scribens, LanguageTool, PaperRater, GradeProof—either ignore Turkish completely or restrict themselves to interface translation; none of them analyse Turkish morphology or syntax. The single exception in this category is Duplichecker, whose Turkish module highlights basic spelling and agreement errors, yet offers no deeper analysis (e.g., verb-tense choice, punctuation rationale) and produces no formative comments.

No dedicated style-suggestion utility (e.g., Hemingway Editor, TextExpander) currently processes Turkish prose, so learners cannot obtain readability scores or sentence-level conciseness advice in their target language.

The AI-based tools present the most encouraging—though still provisional—picture. Large-language-model services such as Google Bard, ChatGPT, QuillBot, and Writefull can already generate, paraphrase, or rewrite Turkish text and flag obvious surface errors. Their feedback, however, is entirely machine-generated, delivered only to the learner, and is not aligned with any instructor-defined rubric; diagnostic accuracy varies, explanations are opaque, and learners cannot audit the underlying correction logic. In sum, purpose-built, instructor-mediated support for Turkish writing remains scarce: only Duplichecker offers rudimentary rule-based checking, no style tool handles Turkish, and

AI-based platforms are not directly integrated into formal educational assessment processes; moreover, their decision-making mechanisms largely operate through non-transparent (black box) structures.

The issues highlighted in the preceding sections jointly shaped the study's research design. Concerns about academic integrity in online environments underscored the necessity for a secure and transparent assessment workflow, while the documented absence of Turkish-specific writing tools made the development of a language-tailored solution imperative. Accordingly, the Method section details a design-based research (DBR) process that (i) embeds an AI-mediated, plagiarism-aware feedback loop and (ii) operationalises Turkish writing criteria through a purpose-built corpus, rule sets, and an analytic rubric. The following pages therefore describe the successive design iterations, ethical approval procedures, data-collection instruments, and analysis techniques adopted in this study.

2. Method

The model of this study, which aims to develop writing skills and track the process by providing writing activities for learners of Turkish as a foreign or second language in alignment with task-based language teaching, is based on the design-based research model. Design-based research is a methodology that focuses on solving real-world problems while simultaneously contributing to theoretical knowledge (Barab & Squire, 2004). This methodology enables the continuous improvement of learning environments and facilitates the development of more effective educational materials through iterative processes (The Design-Based Research Collective, 2003).

A review of the literature reveals that various tools exist for the computer-assisted evaluation of written productions in foreign or second language teaching, particularly in English (Burstein et al., 2004; Warschauer & Ware, 2006). However, the lack of an automatic assessment system specifically designed for Turkish has been identified as a significant gap in the field. Therefore, the features of similar applications were examined, their strengths and weaknesses were identified, and a new application was designed and developed accordingly.

2.1. Application Design

A learning environment suitable for teaching Turkish as a foreign and second language was modeled by reviewing the literature and examining applications aimed at developing writing skills in different languages within the scope of available technological resources. The operational process of the developed model is illustrated in Figure 1.
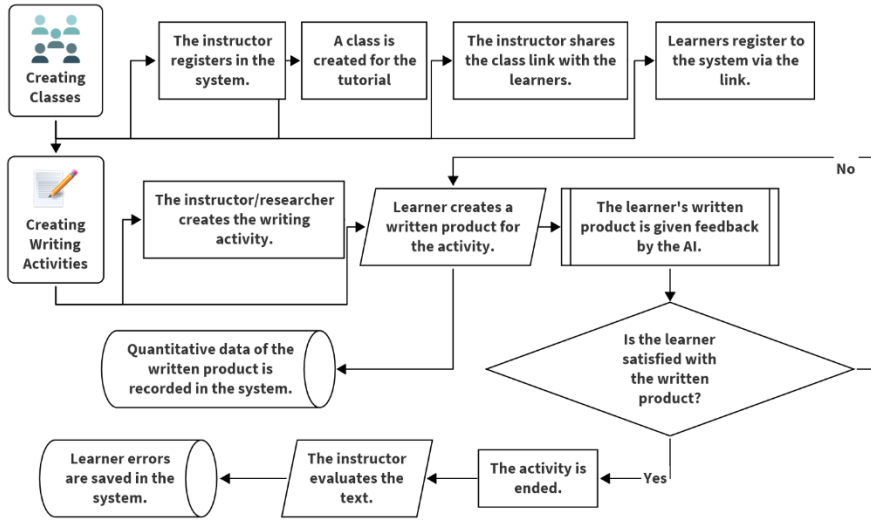
Figure 1. *Process Flowchart of the Designed Model*

The process flowchart in Figure 1 outlines the operational steps of the designed system. The first step in this process involves the instructor registering in the system. Once the instructor completes registration, a classroom is created, and the instructor is provided with a unique registration link accessible from their panel. The instructor shares this link with learners, enabling them to register in the application. As a result of these steps, a user group consisting of both the instructor and learners within the same class is formed.

The second step of the process involves the creation of writing activities and the learners' production of written texts for these activities. At this stage, learners generate texts based on the provided activity guidelines. Once a learner submits a text, the system automatically extracts the following data:
- Total word count in the text
- Number of unique words in the text
- Total number of sentences
- Total number of paragraphs
- New words introduced compared to previous texts and their count
- Punctuation marks used in the text
- Verb usage (categorized separately for A1 and A2 levels)
- Noun usage (categorized separately for A1 and A2 levels)
- Preposition and conjunction usage (count for each preposition and conjunction displayed)
- Usage rate of key vocabulary from the activity guidelines
- Average word length and average sentence length
- Total time spent writing the text
- Number of revisions made by the learner
- AI-supported rubric scoring of the text
- AI-generated corrective feedback on the text

247

Once the learner submits the text, it is sent for analysis to an AI model supporting the Turkish language, specifically the GPT model (GPT-4-0613 by OpenAI was used during the prototype phase). The response from the AI is stored in the system and displayed on a dedicated learner interface, allowing the learner to review their text and make revisions if necessary. Each version of the revised text is separately stored in the system.

The final step of the writing process involves the instructor reviewing the submitted text, assigning a score, and marking errors, ensuring both human and AI-driven evaluation are integrated into the assessment process.

During the development of the application, a corpus was created to identify nouns, verbs, conjunctions, and prepositions in learner's texts. Additionally, a user interface was designed to enable both learners and instructors to interact with the application and perform various tasks seamlessly.

2.2. Corpus Development Based on Language Proficiency Levels

The designed application aims to automatically identify nouns and verbs in learner texts and classify them according to their proficiency levels. To achieve this, a corpus was developed to enable the system to detect nouns, set phrases, and verbs within texts.

Since the initial implementation of the system was planned for A1 and A2 level learners at Hacettepe TÖMER, the Istanbul Turkish Coursebooks for Foreigners—A1 (Bölükbaş et al., 2023a) and A2 (Bölükbaş et al., 2023b)—were used as reference materials. The verbs and nouns found in these coursebooks were extracted from the vocabulary lists provided at the end of each unit.

To enhance the accuracy and reliability of identifying and classifying verbs based on proficiency levels, the extracted root verbs were conjugated for each verb tense and personal pronoun in both affirmative and negative forms. These conjugated forms were tagged with their corresponding tense, proficiency level, and unit information, allowing for precise classification of verb usage in learner texts.

The developed corpus consists of:

- 1,417 nouns
- 25 set phrases
- 19,289 verbs, including:
- 317 root verbs
- 18,972 conjugated verb forms

In total, the corpus contains 20,731 linguistic elements, serving as a foundational resource for the system's text analysis and evaluation processes.

2.3. Development of the Application Interface

The design of interfaces in learning applications significantly influences learner motivation and engagement (Lewis et al., 1998). Therefore, ensuring an interface design that is both intuitive and accessible is crucial for providing an effective learning experience.

A clean and user-friendly interface enables learners to navigate the application with ease (World Wide Web Consortium, 2024). One of the key factors contributing to user-friendliness is color selection (Faghih et al., 2013). To enhance readability and usability, hover effects were implemented to change button colors when the cursor is placed over them (Sklar, 2012, p. 475). Additionally, dark text on a light background was chosen to improve content readability (Sklar, 2012, p. 197).

Upon logging into the system, learners encounter a learner panel on the homepage, which allows them to track their activities and performance within the application. The learner panel page is illustrated in Figure 2.
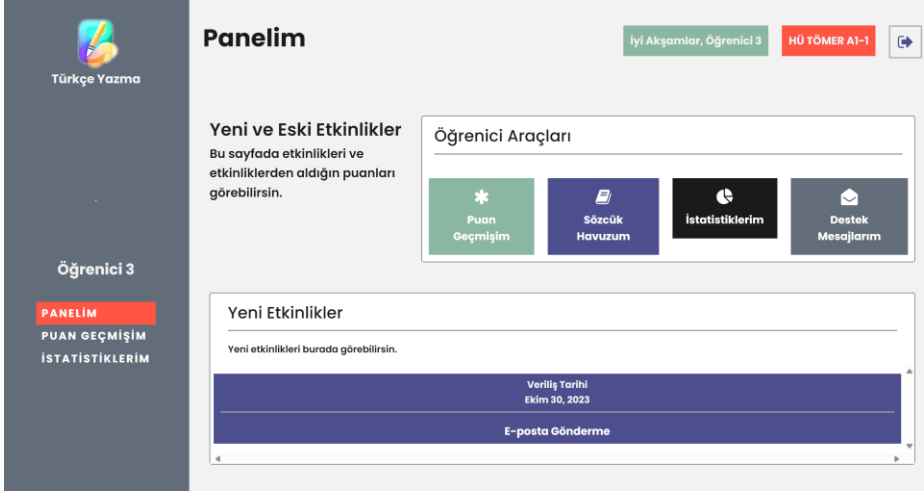

Figure 2. *Learner Panel Page*

Below, Figure 3 presents the score table page, where learners can view the evaluation results of their written texts.
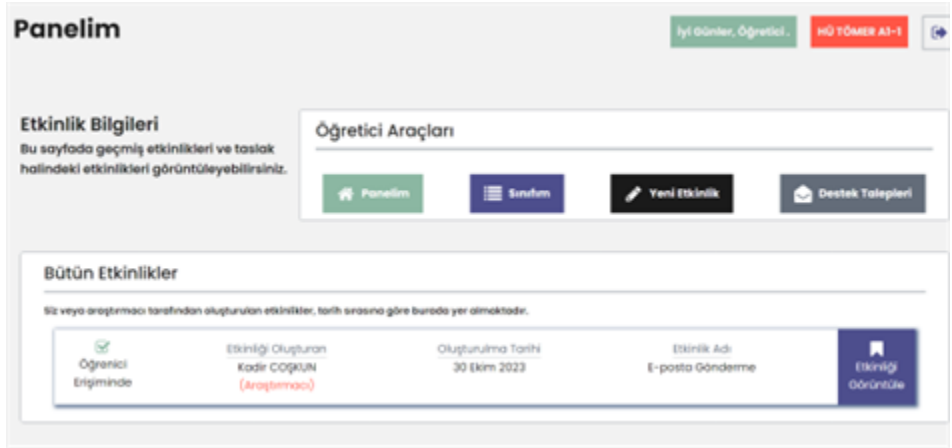

Figure 3. *Score Table and Progress Graph in the Learner Panel*

The instructor panel includes links that allow instructors to view their class lists and access learner-written productions. Instructors can navigate through the panel pages to access the class list, where they can view the data of a selected learner, or they can list activities to check the status of learner submissions for each activity. The instructor panel page is illustrated in Figure 4.

Figure 4. *Instructor Panel Page*

Below, Figure 5 illustrates an example of a class list from the instructor panel. In this section, instructors can view the names of learners enrolled in their class and access individual learner data. This feature allows instructors to monitor learner participation and track their progress throughout the learning process.
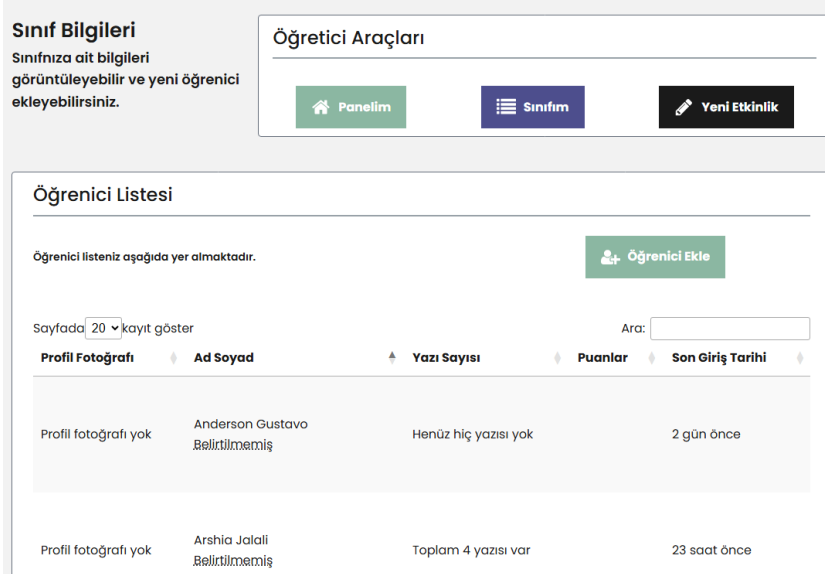
Figure 5. *Class List Page Visible to Instructors*

The writing activities page provides the necessary guidelines for learners to compose texts that align with the assigned tasks. Through this page, learners can complete their writing activities and submit their texts to the system. Instructors, on the other hand, can use this page to view a list of learner submissions for a specific writing activity.

A sample writing activity page containing the activity guidelines and a list of submitted learner texts is illustrated in Figure 6.

Figure 6. *Page Containing the Content of a Writing Activity and a List of Submitted Learner Written Productions*

To prevent plagiarism in learner-written texts, two primary measures have been implemented in the text creation interface.

First, if learners attempt to switch to a different tab or application during the writing process, the page is automatically closed. Additionally, the system does not allow texts copied from external sources to be pasted into the writing interface.

Furthermore, word suggestions on smartphone keyboards are disabled while composing written responses. For users accessing the system via computers, specific security measures have been put in place to prevent modifications to the source code.

Lastly, to detect users who copy and paste text despite these precautions, a typing speed detection function has been integrated into the application. If a user types at a speed of one or more characters per second, the system alerts instructors, enabling them to review the text for potential plagiarism.

2.4. Ethical Principles

This study did not require ethics committee approval as it did not involve the participation of human subjects or living organisms.

3. Evaluation of Learner Texts

The evaluation process for learner texts begins the moment a learner submits their text to the system. The initial assessment is generated by the AI model and is typically provided to the learner within approximately 30 seconds.

Considering that learners may wish to revise their texts based on the feedback received, they are allowed to edit and resubmit their texts up to three times. This iterative process aims to enhance their writing skills by enabling them to refine their work through guided revisions.

3.1. Text Evaluation Rubric

In evaluating written products, criteria such as the content, structure, and grammatical compliance of the text, as well as, if written in handwriting, the page layout and legibility of the letters, play a crucial role. Since many factors influence this evaluation process, analytical scoring is considered more accurate and detailed than holistic scoring, where a single overall score is assigned based on an impression of the text as a whole (Ghalib & Al-Hattami, 2015). Therefore, analytical rubrics were preferred to make the evaluation more transparent and measurable by defining detailed evaluation criteria and scoring each aspect separately.

Rubrics provide the opportunity to assess different features of the text based on predetermined criteria, allowing a more nuanced understanding of the learner's performance. As Wiseman (2012) highlights, analytical rubrics offer more precise feedback by focusing on specific components of writing, which in turn helps evaluators identify the strengths and weaknesses of learners in greater detail. This analytical approach supports a more effective evaluation process by emphasizing the developmental progress of the learner.

In the designed web-based application, since the learner texts are written in an electronic environment, criteria such as page layout and legibility of the writing are ignored, and a rubric was prepared in this context by focusing on the content, formal order and grammatical compliance of the text. The prepared rubric is given in Table 2.

Scoring Type	Item No and Description	Insufficient (1 point)	Acceptable (2 points)	Sufficient (3 or 4 points)
Form	1. The text has a title.	The text does not include a title. (1 point)	The text includes a title, but it is not aligned with the content. (2 points)	The text includes a title that is aligned with the content. (3 points)
	2. The text is formally structured.	The text is not divided into paragraphs. (1 point)	The text is divided into paragraphs, but the divisions are not made appropriately. (2 points)	The text is properly divided into paragraphs. (3 points)
	3. The text is appropriately organized.	The content is not properly ordered, and no meaningful connection is established between paragraphs. (1 point)	The content is mostly ordered appropriately, and a meaningful connection is established between paragraphs; however, some deficiencies	The content is presented in the correct order, and a semantic relationship is established between paragraphs. (3 points)

Scoring Type	Item No and Description	Insufficient (1 point)	Acceptable (2 points)	Sufficient (3 or 4 points)
			and/or mistakes are present. (2 points)	
Content	4. Sentences are clear and understandable.	Sentences are complex and difficult to understand. (1 point)	Sentences are generally understandable, but some parts cause confusion. (2 points)	Sentences are clear and easy to understand. (3 points)
	5. Provided keywords are used.	None of the provided keywords are used in the text. (1 point)	Some of the provided keywords are used in the text. (2 points)	All of the provided keywords are used in the text. (3 points)
	6. The content is relevant to the given topic.	The text is unrelated to the given topic. (1 point)	The text is partially related to the given topic. (2 points)	The text is fully relevant to the given topic. (3 points)
	7. The learner's written product is of sufficient length.	The text is too short and does not adequately cover the topic. (1 point)	The text is of sufficient length but does not fully cover the topic. (2 points)	The text is long enough to fully cover the topic. (3 points)
Grammar	8. Punctuation marks are used correctly.	Punctuation marks are used incorrectly throughout the text. (1 point)	Punctuation marks are mostly used correctly, but some errors are present. (2 points)	Punctuation marks are used correctly throughout the text. (3 points)
	9. Capitalization is used correctly.	Capitalization is used incorrectly throughout the text. (1 point)	Capitalization is mostly used correctly, but some errors are present. (2 points)	Capitalization is used correctly throughout the text. (3 points)
	10. Word choices are appropriate.	Inappropriate or meaningless words are used in the text. (1 point)	Words are generally meaningful and appropriate, but in some places better word choices could have been made. (2 points)	Word choices are appropriate throughout the text. (3 points)
	11. Word spelling is correct.	There are many misspelled words in the text. (1 point)	Word spelling is mostly correct, but some misspellings are present. (2 points)	Word spelling is correct throughout the text. (3 points)
	12. Verb tenses are correctly selected and conjugated.	Verb tenses are incorrectly selected or conjugated. (1 point)	Verb tenses are mostly correct, but some errors are present. (2 points)	Verb tenses are correctly selected and conjugated throughout the text. (3 points)
	13. Suffixes are attached to words correctly.	Suffixes are mostly attached incorrectly. (1 point)	Suffixes are mostly attached correctly, but some errors are present. (2 points)	Suffixes are correctly attached to words throughout the text. (3 points)
	14. Subject-verb agreement is correct.	Subject-verb agreement is incorrect. (1 point)	Subject-verb agreement is mostly correct, but some errors are present. (2 points)	Subject-verb agreement is correct throughout the text. (3 points)

Scoring Type	Item No and Description	Insufficient (1 point)	Acceptable (2 points)	Sufficient (3 or 4 points)
	15. The learner has incorporated newly learned words into the text.	No newly learned words are included in the text. (1 point)	Some newly learned words are included, but they do not significantly contribute to the content. (2 points)	Newly learned words are included and meaningfully contribute to the content. (4 points)
	16. The learner has incorporated newly learned language structures into the text.	No newly learned language structures are included in the text. (1 point)	Newly learned language structures are included, but they do not significantly contribute to the content. (2 points)	Newly learned language structures are included and meaningfully contribute to the content. (4 points)
Total Score		16	32	50

Table 2. *Text Evaluation Rubric*

Table 2 shows the evaluation rubric, which includes 16 items classified under three different classifications: formal, content, and grammar, with three different scoring options: insufficient, acceptable, and sufficient. There are 3 items for the formal evaluation of the text, 4 for the content evaluation, and 9 for the grammatical evaluation. The prepared rubric was arranged to round the maximum score obtained from 16 items to the nearest decimal place to make the obtained scores easier to evaluate. With this arrangement, the highest score that can be obtained for the 15th and 16th items in the rubric, where the situation regarding the language learning process is scored, was changed from 3 to 4. Thus, the highest score that can be obtained in the developed rubric was updated to 50.

3.2. Evaluation of Learner Texts Using Artificial Intelligence

One of the important changes in education is the integration of artificial intelligence into educational processes. Artificial intelligence can provide learners with a customized learning experience through learning management systems, adaptive learning platforms and assessment tools. Artificial intelligence can guide instructors in understanding the individual learning needs of learners and developing more effective teaching strategies. For this reason, artificial intelligence has the potential to increase the quality of education if used correctly and effectively.

The role of AI in evaluating learner texts relies on Natural Language Processing (NLP) algorithms and machine learning models that analyze both grammatical structures and content within texts. AI-based text analysis identifies lexical diversity, sentence structures, and overall content, enabling the system to assess texts based on various criteria and generate reports for instructors. This enhances the efficiency and depth of text evaluations, allowing instructors to review the same text more comprehensively in shorter duration.

Although AI enables the rapid evaluation of large volumes of data, ensuring reliability and validity in large-scale automated scoring requires clearly defined criteria. For this reason, in this study, AI scoring was strictly limited to the rubric items. However, item 15 and item 16, which assess newly learned vocabulary and grammatical structures,

were excluded from the AI-based pre-evaluation, as AI cannot reliably score these aspects.

An example of AI-generated feedback on a learner's text is presented in Figure 7.

Figure 7. *A score and feedback generated by AI for a learner's written product, as seen on the administrator page (a) and front page (b)*
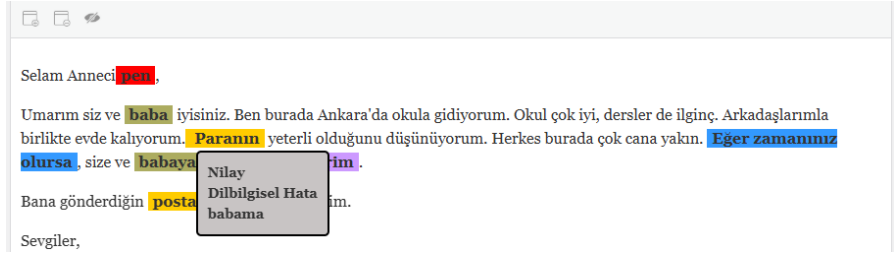
Figure 7 presents how AI-generated evaluations of a learner's written production are displayed within the back-end system and how these evaluations are presented to learners and instructors on the front-end interface.

The feedback generated by AI in Turkish is also translated into the learner's native language, as specified during their registration in the system. Given that AI-generated feedback may include complex grammatical structures, providing learners—especially those at the beginner level—with feedback in their native language is considered a beneficial feature to enhance comprehension and learning outcomes.

3.3. Instructor Evaluation of Learner Texts

The second and final stage of assessing learner-written productions involves evaluation by the instructor of the class in which the learner is enrolled. At this stage, the instructor can identify and mark errors within the learner's text and provide individualized feedback for each mistake. Additionally, using the prepared rubric, the instructor can assign a score to the text and offer further feedback based on the evaluation criteria.

The pages illustrating these evaluation steps are presented in Figure 8.

Figure 8. *The content of a student's written product (a), the scores given by the instructor (b), and the display of the scores on the front page (c)*

Similar to AI-generated evaluations, the scores provided by instructors are displayed using a five-tier rating scale (represented by stars) for each scoring category. This visualization aims to make the evaluation process more intuitive and easier to interpret for learners.

Once the instructor's evaluation is completed, the assigned scores and feedback become accessible to learners. To avoid confusion on the interface, the AI-generated evaluation results are removed from the front-end display after the instructor's assessment is finalized.

Additionally, once an instructor has evaluated a text, the learner loses the ability to edit it, ensuring that both the instructor's assessment and the original learner submission remain intact and preserved.

Conclusion

This study aimed to design a usable model for improving writing skills in learners of Turkish as a foreign or second language. A review of the literature revealed that similar applications exist for languages such as English, but no equivalent tool was available for Turkish. Therefore, the proposed learning environment addresses a real need in Turkish language instruction.

The operational process of the application begins with the instructor registering, creating a class, and inviting learners to join. Learners then produce written texts for writing activities, which are subsequently evaluated and scored by the instructor. The artificial intelligence model analyzes learner texts and provides feedback, contributing to developing learners' writing skills and providing effective evaluation opportunities for instructors.

Among the points considered in the application design are the creation of a corpus according to language levels, using a detailed rubric to evaluate learner texts, and providing feedback through artificial intelligence. These elements improve the learning experience by providing personalized feedback to the learner and effective evaluation tools to the instructor. All written products written by the learner during the process, changes made to their written products, feedback received for each written product, scores, and other quantitative values are stored in the system and made accessible to both the learner, the instructor, and if deemed necessary, institution administrators. By storing learners' data, it is possible to track their development during the process and observe their status more easily than other learners in the same class.

As a result, the designed learning environment offers an effective model for developing learners' writing skills. It provides a learner-centered learning experience suitable for today's technology. Although this study was limited to the design and development of the application, the prototype obtained has created a useful starting point for future research by revealing the basic building blocks. It is thought that the fine-tuning process of the model will be a critical step in increasing the accuracy of text evaluations performed by artificial intelligence and in analyzing the criteria related to newly learned words and language structures more reliably. It will be of great importance for future studies to experience this application with learners at different language levels, to examine its usability by instructors and learners in detail, to analyze the pilot application results and user feedback, to enrich it in terms of comparatively addressing the consistency between artificial intelligence and human evaluation, and to support it with an in-depth examination of technical issues such as data security and scalability. In addition, the fact that the current prototype has not been tested in real educational environments requires a comprehensive evaluation of instructors' adaptation, user experience, and long-term learning outcomes. Such comprehensive studies demonstrate that the developed prototype can provide valuable contributions to integrating the prototype into educational processes and developing learners' writing skills by revealing its effectiveness in teaching Turkish as a foreign and second language from a broader perspective.

Abbreviations

AI: Artificial Intelligence

NLP: Natural Language Processing

DBR: Design-Based Research

GPT: Generative Pre-trained Transformer

TOEFL iBT: Test of English as a Foreign Language – Internet-Based Test

References

Abdelrahim, Y. (2021). How COVID-19 quarantine influenced online exam cheating: A case of Bangladesh University students. *Journal of Southwest Jiaotong University*, *56*(1), Article 18. https://doi.org/10.35741/issn.0258-2724.56.1.18

Alessio, H. M., Malay, N., Maurer, K., Bailer, A. J. & Rubin, B. (2018). Interaction of proctoring and student major on online test performance. *The International Review of Research in Open and Distributed Learning*, *19*(5). https://doi.org/10.19173/irrodl.v19i5.3698

Atoum, Y., Chen, L., Liu, A. X., Hsu, S. D. H. & Liu, X. (2017). Automated online exam proctoring. *IEEE Transactions on Multimedia*, *19*(7), 1609–1624. https://doi.org/10.1109/tmm.2017.2656064

Barab, S. & Squire, K. (2004). Design-based research: Putting a stake in the ground. *Journal of the Learning Sciences*, *13*(1), 1–14. https://doi.org/10.1207/s15327809jls1301_1

Barnett, D. C. & Dalton, J. C. (1981). Why college students cheat. *Journal of College Student Personnel*, *22*(6), 545–551. https://eric.ed.gov/?id=ej253567

Bennett, R. E. (2011). Formative assessment: a critical review. *Assessment in Education: Principles, Policy & Practice*, *18*(1), 5–25. https://doi.org/10.1080/0969594X.2010.513678

Black, P. & Wiliam, D. (1998). Assessment and Classroom Learning. *Assessment in Education: Principles, Policy & Practice*, *5*(1), 7–74. https://doi.org/10.1080/0969595980050102

Bölükbaş, F., Yilmaz, Y. & Keskin, F. (2023a). *Yeni İstanbul uluslararası öğrenciler için Türkçe A1* (8. Baskı). Kültür Sanat Yayınevi, Press Global.

Bölükbaş, F., Yilmaz, Y. & Keskin, F. (2023b). *Yeni İstanbul uluslararası öğrenciler için Türkçe A2* (8. Baskı). Kültür Sanat Yayınevi, Press Global.

Burstein, J., Chodorow, M. & Leacock, C. (2004). Automated essay evaluation: The Criterion online writing service. *AI Magazine*, *25*(3), 27–36. https://doi.org/10.1609/aimag.v25i3.1774

Cizek, G. J. & Wollack, J. A. (2017). Exploring cheatings on tests: The context, the concern, and the challenges. In G. J. Cizek & J. A. Wollack (Eds.), *Handbook of quantitative methods for detecting cheating on tests* (ss. 3–19). Routledge Taylor & Francis Group.

Coghlan, S., Miller, T. & Paterson, J. (2021). Good proctor or "big brother"? Ethics of online exam supervision technologies. *Philosophy & Technology*, *34*(4), 1581–1606. https://doi.org/10.1007/s13347-021-00476-1

Comas-Forgas, R., Lancaster, T., Calvo-Sastre, A. & Sureda-Negre, J. (2021). Exam cheating and academic integrity breaches during the COVID-19 pandemic: An analysis of internet search activity in Spain. *Heliyon*, *7*(10), e08233. https://doi.org/10.1016/j.heliyon.2021.e08233

Çangal, Ö., Çelik, M. E. & Başar, U. (2025). Yabancılara Türkçe öğretiminde yapay zekâ kullanımına yönelik öğretici görüşleri. *Aydın Tömer Dil Dergisi*, *10*(1), 57–97. https://doi.org/10.17932/IAU.TOMER.2016.019/

The Design-Based Research Collective (2003). Design-based research: An emerging paradigm for educational inquiry. *Educational Researcher*, *32*(1), 5–8. https://doi.org/10.3102/0013189X032001005

Faghih, B., Azadehfar, M. R. & Katebi, S. D. (2013). User interface design for e-learning software. *International Journal of Soft Computing and Software Engineering*, *3*(3), 786–794. https://doi.org/10.7321/jscse.v3.n3.119

Ghalib, T. K. A. & Al-Hattami, A. (2015). Holistic versus analytic evaluation of EFL writing: a case study. *English Language Teaching*, 8(7). https://doi.org/10.5539/elt.v8n7p225

Howard, R. M. & Davies, L. J. (2009). Plagiarism in the internet age. *Educational Leadership*, *66*(6), 64–67. https://eric.ed.gov/?id=ej834080

Imran, A. M. & Nordin, M. S. (2013). Predicting the underlying factors of academic dishonesty among undergraduates in public universities: A Path Analysis Approach. *Journal of Academic Ethics*, *11*(2), 103–120. https://doi.org/10.1007/s10805-013-9183-x

Janke, S., Rudert, S. C., Petersen, Ä., Fritz, T. M. & Daumiller, M. (2021). Cheating in the wake of COVID-19: How dangerous is ad-hoc online testing for academic integrity? *Computers and Education Open*, *2*, 100055. https://doi.org/10.1016/j.caeo.2021.100055

Kasinathan, V., Yan, C. E., Mustapha, A., Abdul Hameed, V., Ching, T. H. & Thiruchelvam, V. (2022). Proctorex: An automated online exam proctoring system. *Mathematical Statistician and Engineering Applications*, *71*(3s2), 876 – 889-876 – 889. http://philstat.org/special_issue/index.php/msea/article/view/320

Khan, S. & Khan, R. A. (2019). Online assessments: Exploring perspectives of university students. *Education and Information Technologies*, *24*(1), 661–677. https://doi.org/10.1007/s10639-018-9797-0

King, C. G., Guyette, R. W., Piotrowski & Chris (2009). Online exams and cheating: An empirical analysis of business students' views. *The Journal of Educators Online*, *6*(1), 1–11.

Koe, L. S., Kustandi, C. & Siregar, E. (2024). AI-driven feedback system: Implementing advanced NLP and openAI for online learning. *Jurnal Inovasi Dan Teknologi Pembelajaran*, *11*(3), 137–148. https://journal-fip.um.ac.id/index.php/jinotep/article/view/1326

Lewis, R., Stoney, S. & Wild, M. (1998). Motivation and interface design: maximising learning opportunities. *Journal of Computer Assisted Learning*, *14*(1), 40–50. https://doi.org/10.1046/j.1365-2729.1998.1410040.x

Maramark, S. & Maline, M. B. (1993). *Academic dishonesty among college students. Issues in education*. Office of Educational Research and Improvement.

Ocak, G. & Karakuş, G. (2021). Undergraduate students' views of and difficulties in online exams during the COVID-19 pandemic. *Themes in Science and Technology Education*, *14*, 13–30.

OpenAI, Achiam, J., Adler, S., Agarwal, S., Ahmad, L., Akkaya, I., Aleman, F. L., Almeida, D., Altenschmidt, J., Altman, S., Anadkat, S., Avila, R., Babuschkin, I., Balaji, S., Balcom, V., Baltescu, P., Bao, H., Bavarian, M., Belgum, J., . . . Zoph, B. (2023). GPT-4 technical report. Advance online publication. https://doi.org/10.48550/arXiv.2303.08774

Polat, M. (2017). Why do students cheat in examinations in Turkey? A meta-synthesis study. *Eğitim Bilimleri Araştırmaları Dergisi*, *7*(1), 203–222. https://doi.org/10.22521/jesr.2017.71.7

Rowe, N. C. (2004). Cheating in online student assessment: Beyond plagiarism. *Online Journal of Distance Learning Administration*, *7*(2), 1–10.

Sakkir, G. & Dollah, S. (2019). Measuring students' writing skills using Facebook group application in EFL context. *International Journal of Humanities and Innovation (IJHI)*, *2*(3). http://eprints.unm.ac.id/29923/

Scanlon, P. M. & Neumann, D. R. (2002). Internet plagiarism among college students. *Journal of College Student Development*, *43*(3), 374–385. https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=93619e5a4a7dac6591 a799732cd7d6a0d2fe635a

Shih, R.-C. (2011). Can Web 2.0 technology assist college students in learning English writing? Integrating Facebook and peer assessment with blended learning. *Australasian Journal of Educational Technology*, *27*(5). https://doi.org/10.14742/ajet.934

Shute, V. J. (2008). Focus on formative feedback. *Review of Educational Research*, *78*(1), 153–189. https://doi.org/10.3102/0034654307313795

Sklar, J. (2012). *Principles of web design* (5th ed.). *The web technologies series*. Course Technology Cengage Learning.

Uyar, Y. (2016). Yazma becerisini geliştirmeye yönelik araştırmalar: Son çeyrek asrın değerlendirilmesi. *Turkish Studies*, *11*(3), 2273–2294. https://doi.org/10.7827/TurkishStudies.9479

Warschauer, M. & Ware, P. (2006). Automated writing evaluation: defining the classroom research agenda. *Language Teaching Research*, *10*(2), 157–180. https://doi.org/10.1191/1362168806lr190oa

Wiseman, C. S. (2012). A comparison of the performance of analytic vs. holistic scoring rubrics to assess L2 writing. *International Journal of Language Testing*, *2*(1), 59-92.

World Wide Web Consortium. (2024, December 12). *Web Content Accessibility Guidelines (WCAG) 2.2*. https://www.w3.org/TR/WCAG22/

Zou, M., Kong, D. & Lee, I. (2021). Teacher engagement with online formative assessment in EFL writing during COVID-19 Pandemic: The case of China. *The Asia-Pacific Education Researcher*, *30*(6), 487–498. https://doi.org/10.1007/s40299-021-00593-7