https://doi.org/10.53656/ped2025-9.06

Research Insights
Изследователски проникновения

A CASE STUDY ON UNIVERSITY STUDENTS' PERFORMANCE, ATTITUDES, AND AI USE IN COMPLETING AN ONLINE MULTI-STAGE EFL WRITING ASSIGNMENT

Dr. Blagovesta Troeva-Chalakova
New Bulgarian University (Bulgaria)

Abstract. The paper presents a small-scale case study on the implementation of a multi-stage writing assignment among forty-three students of English as a foreign language (EFL), enrolled in an online intermediate course. The purpose of the study is to evaluate the advantages of a combination of product- and process-oriented assessment, its benefits for students based on their performance and attitudes, as well as the unauthorized use of artificial intelligence (AI) tools during task completion. A mixed-methods approach is employed combining quantitative data from the online learning system in which the assignment is conducted with qualitative data from a student survey. The results suggest that introducing process assessment elements in writing assignments may offer greater learning value than direct product (essay) assessment. Unauthorized use of AI was not entirely prevented. It occurred among a small number of both strong and weaker students. Two online AI detection tools produced highly inconsistent results, while human judgment did not identify all cases flagged by both detectors. The analysis reveals teacher bias, with more cases detected among students who rarely attended classes compared to high-proficiency active learners. In view of reinforcing learning, the results highlight the need for further research on multi-stage assessment and the potential of a hybrid approach, in which AI use is reasonably integrated into the students' training, following strict guidelines.

Keywords: English as a second language; writing assessment; essays; process-based assessment; AI-generated texts; AI use detection

Introduction

While integration of artificial intelligence (AI) tools in teaching offers potential benefits, its impact on student evaluation has raised significant concerns within the academic community. One of the key challenges is the growing dependence of students on AI models to generate texts. Another is the need for educators to rethink both teaching strategies and evaluation methods, when attempting to ensure academic honesty and fairness in assessment.

Essays are among the most widely used assessment tools in foreign language courses (Dikli 2003). They are flexible, allowing educators to modify questions and rubrics depending on their instructional needs (ibid). However, they also allow for easier incorporation of AI-generated content without detection, especially in online courses. Traditionally, essay grading often evaluates only the final product, providing limited insight into students' writing process. This is a serious shortcoming as the latter consists of "the interplay of three recursive cognitive subprocesses (planning, translation, and revision) which interact with the writer's long-term memory and the writing task or task environment" (Hiyes & Berninger 2014).

This study investigates a structured, multi-stage approach to assessment aimed at both improving writing skills and minimizing AI reliance. The participants in this small-scale study were forty-three Bulgarian university students enrolled in an online Intermediate English-as-a-foreign-language (EFL) course for one semester. Most of them were high performers, with only a few experiencing difficulties with what Abbas (2017) calls common challenging areas such as language use or idea generation. They reported uncertainty regarding academic essay structure and organization, cohesion and coherence.

Addressing this need, the teacher introduced elements of the process-led approach when teaching them how to write advantages and disadvantages (A&D) academic essays. Planning received the greatest focus as it has been known to holistically improve college students' writing scores, being comprised of at least two elements: idea generation and organization (Limpo & Olive 2021). During the training sessions, specific features of the essay structure were highlighted and practiced. Students were scaffolded in drafting and organizing ideas, paragraph structuring, and maintaining coherence and cohesion. Such a multi-stage approach aligns with Vygotsky's concept of the Zone of Proximal Development (ZPD), where learners achieve higher levels of competence through guided practice and feedback. Personal feedback was given to the students on all the tasks performed. It was considered an essential part of the training as it "emphasizes a process of writing and rewriting where a text is not seen as self-contained but points forward to other texts the student will write" (Hyland 2010, p. 177). That is, it aims to help students become more independent writers.

The students' writing skills were assessed using a four-stage task mirroring some of the practice activities and culminating in writing an A&D essay. It marked a shift from the traditional assessment of a final product of writing, the decision being driven by three main objectives. First, it responded to students' needs, as they had reported and demonstrated strong language skills but weaker mastery of academic structuring and organization. Second, it aligned with the view that when working on different stages, students' benefits build up as they improve not only their writing but also their cognitive skills (AlTamimi 2020; Hayles 2024). Clark even argued that "introducing students with various stages of writing will be more beneficial

than highly focusing on their language structure" (2012). In addition, requiring students to develop their essays in stages was expected to eventually dissuade some students from submitting AI-generated texts. We found some support in such a proposition in August et al.'s (2024) suggestion to reduce reliance on AI tools by including tasks that require critical engagement such as documenting the writing process (e.g. through drafts, revisions, and reflections).

The research questions guiding this study were the following:

1. Did the students' results from the essay assignment correspond to the overall performance of the students on the course?
2. What were the students' perceptions of the adopted writing procedure?
3. Did students resort to unauthorized use of AI to generate their essays?

In answering these questions, the study tries to highlight the benefits of process-based assessment and the challenges of AI use, as there is limited evidence on whether multi-stage writing tasks can simultaneously support writing development and reduce reliance on AI tools, particularly in an online EFL context.

Literature Review

Product- and Process-Oriented Assessment

Traditionally direct assessment, carried out over a limited time, has been the dominating method of evaluating students' written assessment (Dikli 2003, Drid 2018). Drid (2018) defines direct assessment of writing, also known as "on-demand or impromptu writing" (Wolcott & Legg 1998, p. 10), as an approach in which students are required to create a complete text following a specific framework. It is often associated with the so-called product-based writing, where teachers predominantly evaluate the final product of students' work, even though "the methodology of teaching ESL writing has shifted toward process-based approaches over the last two decades" (Lee 2006, p. 307). Although this method allows for a faster and direct evaluation of writing skills, it does not manage to register the cognitive processes taking part in the creation of a text (Weigle 2002). It also is not very informative about the extent to which students have mastered the different processes involved in writing a coherent and meaningful piece. And although it was considered that unlike portfolio assessments, "questions of authorship or of collaboration do not arise" as "the writing is done under the instructor's eye" (Wolcott & Legg 1998, p. 25), this is no longer the case, particularly in online ESL courses.

Contemporary discussions have highlighted the benefits of process-based assessment, reflecting the idea that writing is "a series of recursive stages entailing deliberate goals and choices on the part of the individual" (Flower & Hayes 1981). "Process" is meant to be "discovery in which ideas are generated and not just transcribed as writers think through and organize their ideas before writing and revising their drafts" (Lee 2006, p. 307). The most effective teaching of writing is

divided into different stages aiming to support systematically the development of skills needed (Sun & Feng 2009). Assessing in the process-oriented approach per se includes work on several stages of the writing process and involves polishing with the help of peer and teacher conferencing (Montague 1995).

In their meta-analysis of 115 writing intervention papers, Graham & Perin (2007) confirm that there is a statistically significant improvement in the quality of students' writing when students' writing process is scaffolded and supported. This includes instruction on how to use specific text structures, what strategies they need to apply, how to regulate them, and how to meet specific writing objectives. Similarly, Vega and Pinzon (2019) show that applying a process-based approach through guiding students through the various stages, such as planning, monitoring and self-evaluation, improve the text's content, organization and vocabulary, as well as students' confidence about writing in English. Lee (2006) finds out that process-oriented essay writing, including submission and peer discussion of drafts, leads to increased length, more complex sentences, increased holistic and analytic scores. Murray's (1972) thought that "instead of teaching finished writing, we should teach unfinished writing, and glory in its unfinishedness" highlights the value of working on various levels and aspects of the text before we reach excellency.

One effective way to account for the development of students' writing skills, while gaining insight into the processes involved, is through formative assessment tools such as portfolio assessment. It, however, does not fulfill the needs for a more controlled and concentrated process assessment (Lee 2006). Process-oriented assessment usually involves peer feedback and assessment, which in some circumstances, as in end-of-semester final tests, is not a likely part of the evaluation procedure. Another type, which we employed and explored in our experiment, is multi-stage essay writing. As it only partially responds to the characteristics of a process-oriented assignment, it could be more appropriate to describe it as "adaptations to direct assessment" (Wolcott & Legg 1998 p. 26) employing both elements of process and product writing. The term refers to prewriting strategies, which allow for preparations of the students about the topic and revisions of the draft within a timed period. The shortcomings of this approach lie in the limited possibilities for revising higher-order elements such as organization and development (Wolcott & Legg 1998 p. 26).

AI in EFL Writing Improvement

In the last few years, an increasing body of research has focused on the beneficial use of AI tools in the classroom. An improvement in writing scores via the help of tools like Grammarly, ChatGPT and Quillbot is reported in Dja'far and Hamidah (2024). Gültekin Talayhan and Babayiğit (2023) specifically investigate how AI tools enhance content development and text organization in EFL writing, identifying improvements in generation of ideas, vocabulary, and coherence. August et al. (2024) highlight various types of AI assistance, including brainstorming, spell-checking, and reference formatting. Pratama and Hastuti (2024) find significant

improvements in EFL students' descriptive texts due to AI provision of personalized feedback, while Joo (2024) discusses both the opportunities and weaknesses of such feedback. The use of ChatGPT has been shown to enhance ESL learning and ease teachers' administrative workload in Chinese context (Hung and Chen 2023). Klenbort (2023) compares it to past technological shifts and views it as an educational tool carrying considerable potential.

Challenges of AI in Academic Integrity

While AI tools offer significant benefits for EFL writing, their use also raises important ethical and practical challenges. Issues such as dishonesty, plagiarism, over-reliance on AI, and the challenges of detecting AI-generated texts have been reported in Mohammadkarimi (2023), Khalil and Er (2023), Vasilatos et al. (2023), etc. The problem in educational institutions is further complicated as many universities lack specific policy guidelines on AI use, and it is up to the instructors to determine what constitutes dishonest practices in the courses they lead (Caulfield 2023a in August 2024, p. 189). In the context of ESL teaching, using spell-checking tools may not be considered a violation, whereas submitting a fully AI-generated text seriously hinders students from learning (August et al. 2024). August et al. (2024, p. 191) stress that educators must establish clear guidelines for themselves and their students on how AI tools align with learning objectives.

An increasing number of scholars warn against dangers of decline of critical thinking (Gültekin Talayhan & Babayiğit 2023, August et al. 2024). The authors of the last two studies advocate for a balanced approach to AI use by educators and learners, and present a framework for developing guidelines for AI use in students' writings. It has been emphasized that AI tools need to be supporting and not leading the learning process (Godwin-Jones 2022).

AI tools can write human-like essays, and combinations of human and AI-generated texts are getting harder to detect (Yan et al. 2024). The discourse goes beyond the notion of plagiarizing, which Merriam-Webster Online Dictionary (n.d.) defines as stealing and passing off (the ideas or words of another) as one's own or using (another's production) without crediting the source. AI-generated texts are different in that they are impossible to compare to an original version, which makes the concept of "plagiarism" ambiguous (Vynck 2023).

While teachers can be trained to identify better such interference, still the outcome is not up to the level that can be desired, and we can expect the models to mimic better human writing styles in the nearest future (Yan et al. 2023). Detection is further hampered as "while a badly organized essay with many spelling mistakes is almost certainly human-written, a well-written, well-organized essay is not almost certainly AI-generated" (Yan et al. 2023). Besides, AI tools have demonstrated inconsistencies. For example, during a professional training session, a text about the university's mission written twenty years ago was incorrectly flagged as AI-generated (76%) by one of the popular AI-use detectors[1].

Pedagogical Responses to AI Use

The discourse on fairness of assessment has led to a shift in the way we think about the reasons for evaluating students' knowledge, needs or achievements, and the way this is done. Voices arise in favour of an overall shift in the assessment methods by turning to in-class writing and/or oral presentations (Klenbort 2023; Hung & Chen 2023). The latter emphasizes the need to instruct students how to critically evaluate AI-generated content. Mohammadkarimi (2023) suggests strategies to address AI-driven academic dishonesty, such as incorporating problem-solving activities, using plagiarism detection tools, and encouraging students to express their own ideas.

Regarding ESL courses, especially those in an online format, some of the Mohammadkarami's (2023) solutions do not work perfectly. Excellent performance in speaking would not automatically correlate to excellency in writing. Also, in many cases, online ESL courses provide no options for in-class examinations. Critically evaluating AI-generated content is an essential skill that must be taught and acquired in any field of knowledge; however its application is limited when writing tasks require non-factual personal accounts and sharing of opinions. Plagiarism detection software fails to identify AI-generated content, while AI-detection tools are still unreliable, as pointed out earlier.

Problem-solving activities and tracking the development of a piece of writing through different stages seem to be some of the logical approaches to a fairer student assessment. While existing research highlights the benefits of process-based assessment and the challenges of AI use, there is limited evidence on whether multi-stage writing tasks can simultaneously support writing development and reduce reliance on AI tools, particularly in online EFL context.

Research Method and Educational Context

The participants in the case-study were 43 students (aged 19-34; 26 women, 17 men) from New Bulgarian University enrolled in an online Intermediate (B1.2.) EFL course – Council of Europe, 2001, in the fall semester of 2024. They belonged to two different course groups (n1=23 and n2=20), and presented a convenience sample, based on the students' attendance on the days assigned for testing.

To carry out the present study, a mixed-method approach was employed. Quantitative data about the students' grades and times of completing the tasks were extracted from the learning management system Moodle, in which the assignment was carried out. The data was analyzed using descriptive statistics (Mean, Median, and SD). Additionally, a post-assignment survey measured the students' perceptions of the multi-stage writing process, which were both statistically and thematically analyzed (Appendix).

A methodological challenge was the lack of AI-detection system within the Moodle platform, requiring the teacher to rely on her own judgement and external

AI-detection tools. Another challenge was that not all students who had enrolled in the course and completed the assignment had attended it or took part in subsequent sessions and exams. This resulted in not having all of them complete the survey and not having a basis for comparison regarding their overall English skills.

The traditional way of testing students' written skills at this level had been through an essay task in which they selected one of two given topics, usually in the form of an A&D essay. The written works had normally been asynchronically submitted online. Prior to implementing the new format, most students underwent a two-session training of how to write A&D essays. It included video materials demonstrating the requirements of the type, focusing on structure, content and cohesion; direct instruction from the teacher; and guided practice into writing paragraphs with a special focus on topic sentences, developing argumentation, and using cohesive devices. These learning objectives were highlighted as the major skills that were going to be evaluated on assessment day. Three of the practice tasks mirrored what was going to be required in the written assignment. The assignment was divided into four sequential stages:

Task 1. Brainstorming. In the first stage the students had to generate ideas on their chosen topic, the instructions requiring students to write only words and phrases rather than complete sentences at this stage. A word limit of 50 words was set.

Task 2. Writing an outline. At the next stage students needed to make an outline of their essay by choosing three advantages and three disadvantages from their brainstormed list, and add supporting details or examples about each (a model was provided as a reference). A limit of 100 words was set.

Task 3. Paragraph development. At this stage the students were asked to write the advantages paragraph, highlighting the topic sentence and linking words and phrases they had used, demonstrating their understanding of paragraph structure. Again, they had a limit of 100 words.

Task 4. Essay Writing. At the last stage the students needed to write their complete essay of about 180 – 230 words, incorporating the previously written advantages paragraph, modeling the disadvantages one on it, and adding an introduction and conclusion.

A major limitation in effectively implementing the task was the university's Moodle version, which lacked the functionality to prevent students from completing tasks out of sequence (i.e., finishing a task before completing the preceding ones). This potentially increased the risk of some students not following the order of the tasks given.

Details about students' performance including the scores per task, the overall time and chronology of each task's completion were extracted from the Moodle test platform. Students also completed a survey a week after the test, answering questions about different parameters of the task and their attitudes about the format of the multi-stage assignment.

Results

The results below are presented in relation to each of the study's research questions.

1. Did the students' results from the essay assignment correspond to the overall performance of the students on the course?

The results of the students were considerably high (Fig.1), which was expected as most of them had demonstrated high proficiency in English during the course. The rubric assigned equal weight to the following five assessment components: a) structure and organization, b) grammar, c) vocabulary, d) coherence and cohesion, e) argumentation. The learning objectives of the writing training session (a and d) were satisfactorily attained in most of the written works.

Basic statistics was performed with the help of an online statistics service[2] (Fig. 1).

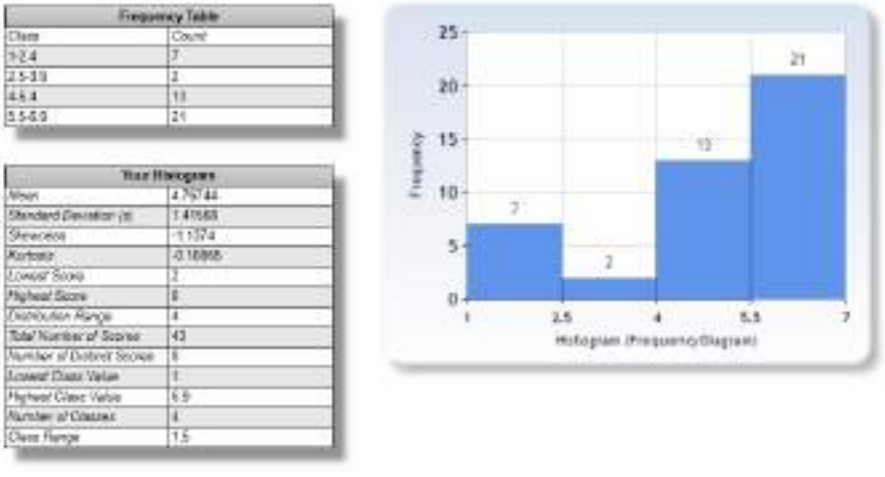

Figure 1. Histogram of the students' scores in a 6-grade scale system

The Mean (4.77) and Median (5.50) scores indicate that most students performed well. The data reflects their strong language skills and effectively structured writing process. The high grades are likely due to the students' higher actual levels, as many had chosen to enroll at the minimum required exit level. The SD (1.42) suggests that the poor grades are rather outliers and do not reflect the typical performance of the two groups of B1.2. students. Students with very low scores (0 – 7 points) were predominantly those who did not complete the four stages of the task and/or submitted essays flagged as AI-generated.

As writing an A&D essay was the only graded assignment of its kind, the students' achievement was studied in the context of their performance throughout the course. The analysis studied the differences between a) students' grades on the

essay assignment, b) their grades on the final written test, and c) their course final grade (Figure 2). The final written test included writing a piece in response to a question and was less structured than a), whereas the final grade was composed of 5 elements, with the oral presentation (the only face-to-face, students' cameras-on event) having the highest impact because of the online nature of the course.

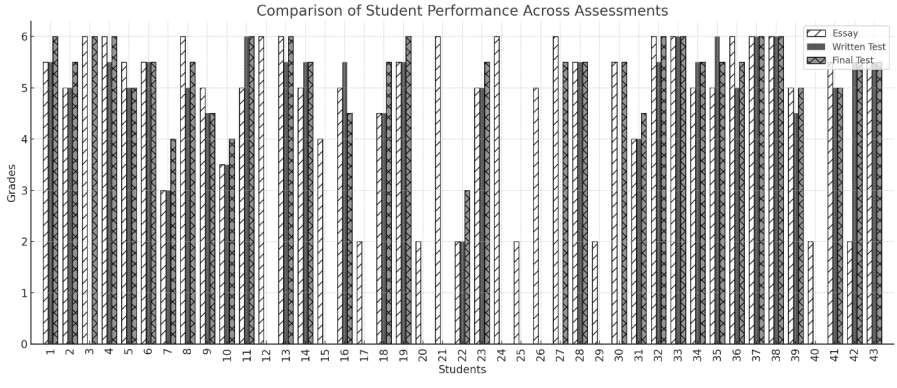

Figure 2. Comparison of student performance across three assessments (in a 6-grade scale system)

Nine students received 0-7 points, which equaled a poor grade (or 2). All of these were identified by the teacher as having submitted AI-generated texts for their own. Six were students who had not attended the course, of whom five did not do any other test, did not attend the examination session, and did not receive a final course grade. Three of the nine students attended the course, irregularly. One of them was a low-achievement student, two had average-to-good skills. All these students were offered a second chance to write an essay, which was used only by the two latter students, who subsequently received good grades.

Coinciding essay assignment and final test grades (Table 1) were very common (50% of 28 students who had a final test grade). Cases with lower essay grade and higher final test grade with a difference of 0.5 were only 10.71%, while those with higher written assignment grade and lower final grade with a difference of 0.5 made up 25%. The final test included tasks on various skills, only one of which was writing a composition. The fact that 92.86% of students had only minimal differences (no or 0.5 difference) between the two grades confirms that the essay performance of the students was not much different from their achievements in a multiple-component final test.

Table 1. Differences between grades on the essay assignment and the written test grade.

Difference between essay grade (a) and final test grade (b) (n=28)	Number of students	%
same grade (a=b)	14	50%
0.5 up (a<b)	3	10.71%
0.5 down (a>b)	7	25%
1 up (a<b)	2	7.14%
1 down (a>b)	2	7.14%

When compared to the overall course grade (Table 2), the analysis revealed that 90.32% of the cases showed either no or only a minimal difference (0.5 grade), with 38.71% of the grades being identical. More students scored higher on the overall grade, which is possibly due to some of its components, such as grades on speaking and participation. However, only 3 students received a grade that was 1 level higher, and no students had a difference of more than 1 grade.

Table 2. Differences between grades on the essay assignment and the final overall grade

Difference between essay assignment grade (a) and final overall grade (c) (n=31)	Number of students	%
same grade (a=c)	12	38.71
0.5 up (a<c)	9	29.03
0.5 down (a>c)	7	22.58
1 up (a<c)	3	9.68
1 down (a>c)	0	0

In answer to our first research question, we found that in most cases the grades on the A&D multi-stage assignment did not significantly differ from the overall performance of the students during the course. In addition, there was a close correspondence between the high values of the grades and the high proficiency of the students throughout the course. Low or zero values will receive special attention in the discussion of the third research question.

2. What were the students' perceptions of the adopted writing procedure?

Students' attitudes towards the four-stage assignment were studied by means of a survey, which tapped into their experience of both the training and the assignment tasks. The information was gathered from 37 of the 43 students, as it was voluntary, and not all students attended classes on the following week.

Some students did not answer some of the questions, therefore, some results are reported in numbers.

Most students (25/37) had attended the training on how to write A&D essays. Among those who missed it, most (10/12) reported they had reviewed the training materials available on Moodle. 30/37 students considered the training useful, seven did not answer this question. Most (35/37) reported an increased confidence in their abilities to write a similar type of essay in the future. Such responses confirmed that the content of the training had addressed areas which the students found challenging and assisted them in improving their writing skills.

We had a particular interest in finding out how the students looked at the major change introduced – the breaking of the writing process into four stages. Most (27/37) found it helpful. They appreciated that it helped them structure their essays (26/37), recall the necessary components (25/37), and refine their ideas (24/37). There was a small number who found this approach inconvenient, as it either slowed them down (6/37), or did not match their personal logic for structuring the essay (5/37).

Regarding the first task (brainstorming) and the second task (writing an outline), the students' overall attitude was positive with approval ratings of respectively 7/10 and 8/10. Task 3, which was writing an advantages paragraph, combined with highlighting the topic sentence and linking devices, was also rated with 7/10.

When asked in an open question what they would change in the writing procedure, a significant number (14/37) explicitly stated that they would not change anything. However, some students suggested reducing the number of steps (4/37), writing the entire essay all at once (3/37), or just the opposite – adding stages for the other parts of the essay (2/37), and focusing more on the draft before finalizing their work (2/37).

The survey provided insight into the extent to which students accepted the new evaluation method and identified some strengths and weaknesses, which will be discussed below and addressed in future assignment planning.

3. Did students resort to unauthorized use of AI to generate their essays?

We performed four types of analyses to identify AI-assisted writing: a) on the overall time of completion, b) on the chronology of completing the different stages, c) on the content, and d) on students' direct responses.

The time taken to complete the essay by the nine students who scored 0-7 points had Mean=42.05 min, Median=43.27 min, and SD=16.29 min. This contrasts sharply with the time invested by the other 34 students (12-22 points), whose data showed Mean=88.10 min, Median: 82 min, and SD=36.22 min.

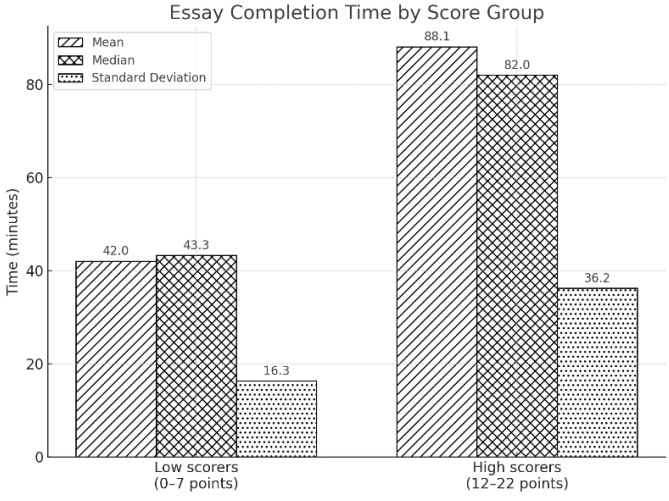

Figure 3. Essay Completion time by score group

The time that the lower-scoring students took was approximately half of the time needed by the others (Fig. 3). That might mean that they rushed through the tasks (some of them skipped some of the stages). The smaller SD also means that there was not much variability in the time they took. It could also mean potential reliance on AI tools to generate their essays quickly, which led to consistent completion times. The longer completion time of the stronger students (12 – 22 points) shows that they invested more effort working on the four stages of the assignment. Their SD is more variable, which reflects natural tendencies for students to work at different pace while completing a task.

The second insightful piece of data was Moodle's periodical time saving reports. We found that at least 3 students had pasted the completed essay in the final task soon after the task was accessible and then extracted from it the pieces needed for the previous tasks, working backwards. These were students who had not attended the training, probably did not know how to approach the tasks, and possibly used AI assistance. As students who had attended the training did not act in this way, this finding highlights the importance of detailed, step-by-step manner training in the skill of writing.

The cases flagged by the teacher as probably AI-generated stood out due to their native-like proficiency and fluency, sophisticated vocabulary, and intricate grammar. The structure closely followed the one given in the instructions. Seven out of the nine students were unknown to the teacher, and she had no other data about their command of English. In their case, resorting to AI assistance could be because of uncertainty and lack of sufficient knowledge and skills in writing. In one case,

the student had very low proficiency and was a close passer, completing the course with a fair grade. In two other cases, the students were good performers, availed themselves of the opportunity to write the essay a second time, and received very good grades. This proved to us that use of AI in writing texts is seen both in students with low performance and lack of skills to do the tasks by themselves, as well as by students with higher language proficiency.

After the test was manually assessed by the instructor, an AI detection tool ("ZeroGPT"[3]) was used to detect assisted writing of the final stage of the essay. A probability of over 75% was considered a positive one. The online tool detected 16 essays possibly generated by AI, 12 of them having 100%, one 96.93%, one 94.57%, one 87.77%, one 76.78% probability. Out of the rest, 16 essays were judged to have been written with 0% of AI assistance, and 11 essays with a probability between 9.14% and 61.52%.

Taking into consideration the unreliability of some AI detection tools, we employed a second test and ran all the essays through the platform "justdone.ai"[4]. This even complicated our findings as the discrepancies between the two AI detectors turned out to be substantial, probably due to differences in their training datasets or underlying algorithms (Fig. 4). We received doubtful results, e.g. 100% of essays being detected as AI-generated. Fourteen of them had been flagged as 0% AI-generated by the first detector, the latter coinciding with the teachers' own perception of the texts' originality. Our solution was to flag as AI-generated only those essays which showed more than 75% probability in both detectors' analyses. These coincided with the 16 essays identified by the first AI detector.

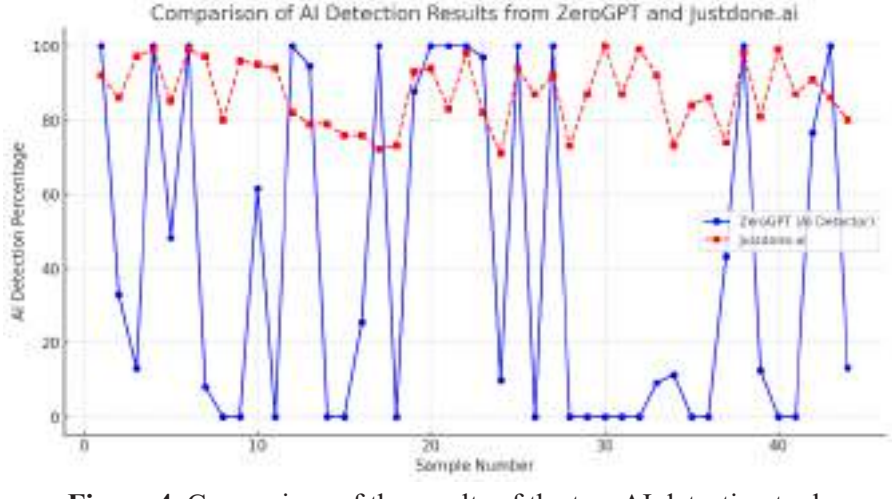

Figure 4. Comparison of the results of the two AI detecting tools

Comparing the positive identification by both the teachers and the online tools, we found a certain discrepancy (Table 3).

Table 3. Cases of AI-generated texts identified by the teacher compared to those identified by AI-detection tools

Did the teacher identify AI use?	Students who attended the course	Students who did not attend the course
Yes	2	7
No	7	2

Most cases which the human assessor identified in agreement with the AI-detecting tool were of students who had not been attending the course and therefore were unknown to the instructor. In contrast, the teacher failed to detect many of the cases when AI assistance was used by students who were known to the teacher. This data might mean that lack of detection was linked to the overall good impression of the instructor from the students' performance in class, which led to overlooking the signs of AI interference in their work.

When students had to self-report on their use of AI in the process of writing, most of the students who answered this question (26/33 or 84%) denied using it, with only 2 admitting using it to generate ideas rather than copying content directly, and 3 reporting use of Google translate (Figure 5). Nevertheless, when the teacher offered the students with points 0-7 to do the assignment a second time, giving feedback on the unauthorized use of AI tools, none of them objected to the feedback on each task, including results from the AI detector.

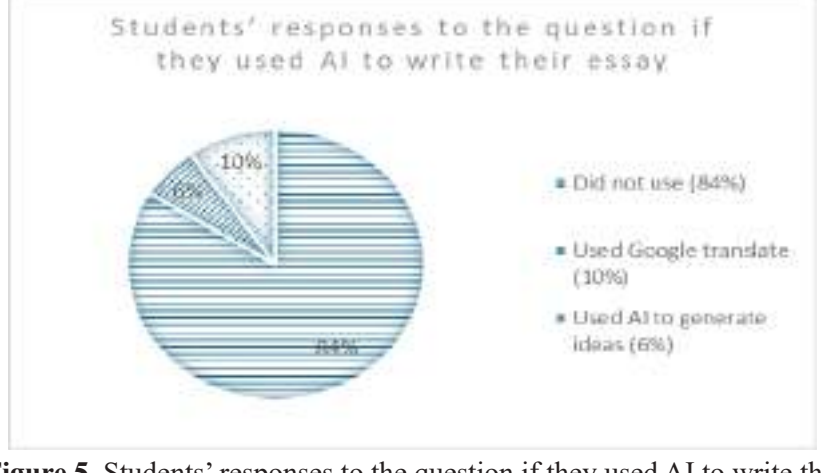

Figure 5. Students' responses to the question if they used AI to write their essay

Discussion

In this study we explored the introduction of a four-stage process of writing an A&D essay as a graded assignment in terms of its benefits to students' writing development, as well as their behaviour related to AI unauthorized use in the fulfillment of the assignment.

As the participants were mostly higher proficiency EFL students, it was expected that most of them would excel in this graded assignment. Therefore, the focus of the preparatory training lay on the areas of difficulty students themselves had put forward: organization and structure, and included comprehensive work on topic sentences, argumentation, coherence and cohesion. The training tasks mirrored the format of the written examination students were going to complete, aiding both their performance and their confidence-building.

Most students found the training useful and effective, and most of them approved of the switch to a less product-based and more process-based writing task as part of their assessment. The majority felt more confident in being able to write a similar type of essay in the future, and said they were reminded of the important structures they would otherwise forget about.

The students' scores aligned with other grades they received during the course, which means that this type of task did not cause devaluation of students' knowledge and abilities. However, some areas could be fine-tuned to accommodate different learning preferences and cognitive styles. First, while breaking the essay writing into stages worked well for most, a minority of students preferred a more fluid approach. If an option was given to either follow the staged method, if they needed structured help, or write more freely, especially if they were highly skilled writers, their different learning styles and capabilities could have been addressed better. Second, the value of each stage could be reconsidered as well. Timed brainstorming exercises could be done in class before the test, so that all students could collaborate to generate ideas, and use them under less pressure in their writing. Such an approach would have helped students who experience difficulties to generate ideas, hindering their writing process, as suggested by Kartikasari (2023).

Writing the advantages paragraph first was meant to address a common problem area - students' difficulty forming topic sentences, proper argumentation and paragraph coherence. Pinpointing some of these elements was thought to guide the students into writing properly structured main body of the essay. Students who had done a similar activity in class were made conscious of why this approach is beneficial, but those who had not might have missed to perceive its value. Significantly, those who tried to skip the initial phrases turned out to be the students who had skipped the training.

Use of AI was not avoided. A very small number of students skipped the first stages, pasted an AI-generated text into the last slot and then inserted parts of it to complete the previous steps. We found that these were students who had not

done the training, therefore were not aware of the strategies involved in writing an A&D essay. Probably, these were also students who had poor knowledge of English, which could not be confirmed as these students did not take part in the rest of the course. There was a possible AI reliance among a small number of strong students as well, as two of the students with an alleged use of AI and consequently low scores, made avail of the second opportunity of writing the essay and performed very well.

Detection of AI was made difficult by the fact that the two AI-detectors used produced considerably different results. Therefore, the instructor put greater trust in her own identification of AI-generated texts, being aware of possible misjudgment. Yan et al. claim that "humans rarely perform better than random guessing (60-65%) when asked to identify texts generated by modern Ais" (2023, p. 126, quoting Clark et al. 2021 and Ippolito et al. 2020). When evaluating non-scientific data such as language use in EFL classes, we might expect the percentage will be at least similar if not higher. In this study, the assumed threshold of over 75% probability coincidence in both AI-generated text detectors confirmed and coincided with the human evaluation, with the exception of a few cases the teacher had possibly missed. An important finding of the study was that lack of personal knowledge of the students increases the chances of a teacher identifying their work as AI generated. Therefore, teachers need to be aware of all personal biases in their attempt to provide fair assessment, in addition to being given specialized training to identify AI-generated texts.

We also recommend that, to ensure fairness, AI detectors should not be the only method of identifying academic dishonesty. The cautious flagging of students' works was informed by research which found that false positives are more common among texts written by non-native speakers than by native speakers (Jiang et al. 2024). Liang (et al. 2023) used several AI use detectors to analyse the essays of non-native speakers' taking the TOEFL examination. More than half the essays were wrongly identified as AI-generated versus classifying 90% of the essays of US native speakers as human-authored. Liang et al. (2023) explain this by the role of perplexity or complexity of vocabulary used. August et al. go as far as to claim that it "is not possible to identify with certainty whether a student used GenAI" (2004, p. 189).

Based on our findings, we could highlight several pedagogical recommendations. The writing process can be made more transparent and reduce AI misuse in graded assignments if they include drafting stages. Implementing students' reflection, e.g. on topic sentences as parts of the required paragraph structure, ensures the assessment helps students reinforce their knowledge and skills. Teachers should train students in responsible AI. Since some students are already using AI for writing support, addressing its appropriate use in academic settings could help clarify its benefits and limitations. This should be preferably done at the very beginning

of each language course. Universities should develop clear AI-assistance policies, especially regarding its use in coursework.

It is important to note that several limitations exist in the present study, one of them being the limited number of participants. Secondly, it would have yielded more precise information about the extent to which multi-stage assignments reduce the probability of students using AI tools, if more instances of them doing the same type of assignment were recorded and researched over time. A third drawback of the assignment processing was that students were not informed of their work going through an AI detecting tool. Future implementations should ensure transparency by informing students of AI detection procedures. They should also have the right to appeal against an AI use claim on their work. Four, there is not a 100% certainty if the unauthorized use of AI really took place, so we can rather talk about high probability. And finally, some of the students' development over the course could not be observed for lack of attendance.

Conclusion

As AI technology evolves, educators must strike a balance between leveraging AI's benefits and ensuring fair, transparent, and ethical assessment practices – a task quite demanding, especially in online courses. We found that changing the format of a written assignment from product-to process-based carried important learning potential but did not dissuade some online EFL students from submitting AI-generated work. In our study, however, those were a very small percentage and were mostly students who had not regularly attended the course and the writing training. Monitoring the writing development over time helps the teacher to assess their students' genuine progress. If assignments are not split into stages as in the present one, students may be asked to submit process logs (e.g. drafts or revisions) to prove their authorship. Future research should explore the opportunities for writing development through hybrid approaches that integrate students' original texts with AI-assisted feedback. As Hayles (2024) points out students need to "develop critical relationship with algorithmic cultures and to transparently show their contributions versus what the AI contributed". Further efforts should focus on the learning objectives and devising assignments in view of students achieving targeted skills, alongside improving teachers' capacities for AI-use detection, and investing time in educating students in responsible use of AI assistance.

NOTES

1. Personal archives.
2. Easy Histogram Maker is available at https://www.socscistatistics.com/descriptive/histograms/default.aspx

3. ZeroGPT detector is available at www.zerogpt.com
4. Justdone.ai detector is available at https://justdone.com/

REFERENCES

ABBAS, M. F.F., 2017. Assessing and Evaluating EFL learners' Ability in Writing Academic Essay. *ISELT*, vol. 5, pp. 257 – 261.

ALEXANDER, K.; SAVVIDOU, CH.; ALEXANDER, Ch., 2023. Who Wrote This Essay? Detecting AI-Generated Writing in Second Language Education in Higher Education. *Teaching English with Technology,* vol. 2.

AL-KHULAIDI, M. A. & ABDULKHALEK, M. M., 2022. Academic Writing Problems in L2 Settings: Realities and Need for Intervention. *Journal of English Studies in Arabia Felix*, vol. 1, no. 1, pp. 42 – 51.

ALTAMIMI, R., 2020. Beliefs and Practices of Teaching Writing by Teachers of English as Foreign Language: The Case of Context Impact. *Bulletin of the Faculty of Languages and Translation, Al-Azhar University,* vol. 25, pp. 196 – 220.

AMIRJALILI, F.; NEYSANI, M. & NIKBAKHT, A., 2024. Exploring the Boundaries of Authorship: A Comparative Analysis of AI-Generated Text and Human Academic Writing in English Literature. *Frontiers in Education,* vol. 9.

AUGUST, E. T.; ANDERSON, O. S. & LAUBEPIN, F. A., 2024. Brave New words: A Framework and Process for Developing Technology-Use Guidelines for Student Writing. *Pedagogy in Health Promotion,* vol. 10, no. 3, pp. 187 – 196.

COTTON, D. R. E.; COTTON, P. A., & Shipway, J. R., 2024. Chatting and Cheating: Ensuring Academic Integrity in the Era of ChatGPT. *Innovations in Education & Teaching International*, vol. 61, no. 2, pp. 228 – 239.

DJA'FAR, V. H. & HAMIDAH, F. N., 2024. The Effectiveness of AI technology in Improving Academic English Writing Skills in Higher Education. *Journal of Language and Literature Studies,* vol. 4, no. 3, pp. 579 – 593.

DIKLI, S., 2003. Assessment at Distance: Traditional vs. Alternative Assessments. *The Turkish Online Journal of Educational Technology,* vol. 2, no. 3, 13 – 19.

DRID, T., 2018. The Fundamentals of Assessing EFL Writing. *Developing Psycho-Educational Practices Lab*, pp. 292 – 305.

GRAHAM, S. & PERIN, D., 2007. A Meta-Analysis of Writing Instruction for Adolescent Students. *Journal of Educational Psychology*, vol. 99, no. 3, pp. 445 – 476.

GÜLTEKIN TALAYHAN, Ö. & BABAYIĞIT, M. V., 2023. The Influence of AI Writing Tools on the Content and Organization of Students' Writing: A Focus on EFL Instructors' Perceptions. *Journal of Current Debates in Social Sciences,* vol. 6, no. 2, pp. 83 – 93.

HYLAND, K., 2010. *Second language writing*. Cambridge University Press.

HAYLES, N. K., 2024. Don't Ban AI from your Writing Classroom; Require It! *Poetics Today*, vol. 45, no. 2, pp. 59 – 265.

HUNG, J. & CHEN, J., 2023. The Benefits, Risks and Regulation of Using ChatGPT in Chinese Academia: A Content Analysis. *Social Sciences,* vol. 12, no. 380.

JOO, S.H., 2024. Generative AI as Writing or Speaking Partners in L2 Learning: Implications for Learning-Oriented Assessments. *Studies in Applied Linguistics & TESOL at Teachers College*, *Columbia University,* vol. 24, no. 1, pp. 54 – 59.

KARIMI, E., 2023. Teachers' Reflections on Academic Dishonesty in EFL Students' Writings in the Era of Artificial Intelligence. *Journal of Applied Learning & Teaching,* vol. 6, no. 2.

KARTIKASARI, P., 2023. Implementing Product and Process Writing Approach on ESP Course for Freshmen at University. *Jurnal Linguistik Terapan,* vol. 13, no. 2, pp. 1 – 6.

KHALIL, M., & Er, E., 2023. *Will ChatGPT Get You Caught? Rethinking of plagiarism detection.*

KLENBORT, S., 2023. Does ChatGPT Have a Place in the Classroom? *Eureka Street,* vol. 33, no. 3, pp. 29 – 31.

LEE, Y.J., 2006. The Process-Oriented ESL Writing Assessment: Promises and Challenges. *Journal of Second Language Writing*, vol. 15, no. 4, pp. 307 – 330.

LIANG, W.; YUKSEKGONUL M.; Mao Y., Wu E.; Zou J., 2023. GPT Detectors are Biased Against Non-Native English Writers. *Patterns (NY),* vol. 4, no. 7.

LIMPO, T. & Olive, T., 2021. *Executive Functions and Writing.* Oxford University Press.

KARIMI, E., 2023. Teachers' Reflections on Academic Dishonesty in EFL Students' Writings in the Era of Artificial Intelligence. *Journal of Applied Learning & Teaching,* vol. 6, no. 2.

MONTAGUE, N., 1995. The Process Oriented Approach to Teaching Writing to Second Language Learners. *New York State Association for Bilingual Education Journal*, vol. 10, pp. 1 – 13.

PRATAMA, R. M. D., & HASTUTI, D. P., 2024. The Use of Artificial Intelligence to Improve EFL Students' Writing Skill. *English Learning Innovation (Englie),* vol. 5, no. 1, pp. 13 – 25.

VASILATOS, C. ; ALAM, M. ; RAHWAN, T. ; ZAKI, Y. & MANIATAKOS, M., 2023. *HowkGPT: Investigating the Detection of ChatGPT – Generated University Student Homework through Context-Aware Perplexity Analysis.*

VEGA, L.F.S.; PINZON, M.M.L., 2019. The Effect of the Process-Based Approach on the Writing Skills of Bilingual Elementary Students. *Latin American Journal of Content & Language Integrated Learning*, vol. 12, no. 1, pp. 72 – 98.

WEIGLE, S. C., 2002. *Assessing Writing*. Cambridge University Press.

WOLCOTT, W. & LEGG, S. M., 1988. *An Overview of Writing Assessment: Theory, Research, and Practice*. Urbana, Illinois: National Council of Teachers of English, Urbana, Illinois.

YAN, D.; FAUSS, M.; HAO, J. & CUI, W., 2023. Detection of AI-Generated Essays in Writing Assessments. *Psychological Test and Assessment Modeling,* vol. 65, no. 1, pp. 125 – 144.

✉ **Dr. Blagovesta Troeva-Chalakova**
ORCID iD: 0000-0002-8477-2053
New Bulgarian University
Sofia, Bulgaria
E-mail: btroeva@nbu.bg