Research Article

JQRLE 1(1) (2024)

Journal of Qualitative Research in Language Education

https://journal.as-ssr.org/index.php/jqrle

Has AI cracked the code on English proficiency assessments? A look at how AI is revolutionizing writing evaluations

Muhammad Galuh Elga Romadhoan[a*]

[a] Graduate Program in ELT, Universitas Islam Malang, Malang, Indonesia

ARTICLE HISTORY

Received: 16 May 2024
Revised: 25 June 2024
Accepted: 29 June 2024
Published: 30 June 2024

Corresponding Author:
Author Name*:
Muhammad Galuh Elga Romadhoan
Email*:
galuhelga23@gmail.com

DOI:

© 2024 The Authors. This open access article is distributed under a (CC-BY License)

ABSTRACT: This scholarly inquiry fuses the rapid growth in artificially intelligent writing systems with associated challenges in assessment tools that gauge English compositional skills. To put it another way, in some contexts, many AI-written texts often exhibit human quality. However, inconsistencies have been found which can problematize identification in differentiating human- from machine-penned compositions. It follows through with the claim that the availability of AI writing instruments implies score potential inflation for any type of standardized English language proficiency examination, which in turn challenges its validity. With this view in mind, 5-7 language assessment experts have been approached in this qualitative study for interviews and document analyses to elucidate concerns connected with differentiating human writing from AI writing. Submissions from cutting-edge AI models are matched against known criteria to identify singular features related to coherence, style, and originality, including the handling of open-ended tasks. Thematic analysis was tease out key themes related to the variation in the quality of the composition. Results offer useful insights into possible future detection methods and assessment designs.

KEYWORDS: Artificial intelligence, writing assistance, English writing, writing proficiency, ELT

Introduction

The advent of Artificial Intelligence (AI) technologies and Larger Language Models (LLMs) have made significant debates in the ruling of human writing competencies since these AI systems were highly adept at creating text not only as smooth and logical but convincingly imitative of human writing patterns. These competencies, amidst exemplary collegiate standards, would naturally be missing in AI-generated writing with subtilities of comprehension and contextual underpinnings central to human experience, expertise, and critical thinking processes (Fui-Hoon Nah et al., 2023). Because of this very difference, the role played by human cognition in coming up with original material for textual meaning becomes very important. While AI might help cobble together text that looks deep and coherent, it simply cannot hold the depth and experience a human writer puts into their work (Creely, 2024; Mohamed et al., 2024). This matters in a host of fields demanding reverse-engineered acumen and serious contextual understanding—literature, philosophy, and the social sciences, among others. Moreover, Furthermore, a set of ethical and pedagogical problems has been brought to the fore by AI use in writing (Raheem et al., 2023). Suffice it to say that since AI might hurt the development of human writing skills—especially in the context of educational institutions—the mechanisms of assessment was have to be readjusted (Han et al., 2023; Rahmi et al., 2024). Still more fundamentally, it is considered that AI would inhibit the

development of human writing skills, especially within those educational institutions. This possibility rhetorically might change the whole methods of evaluations.

The assessment of written English has long been in the hands of a human judgment of essays, reports, and all other forms of written discourse (Shanorra et al., 2021; Toba et al., 2019). Organization, development of ideas, vocabulary used, correctness of ideas or grammar, and mechanical accuracy—the writing is minutely evaluated by humans (Hz et al., 2023). However, AI writing technologies throw prominent challenges into this evaluation process. In many instances, human raters have found it quite challenging to determine which text is written by a human and which is by a computer—especially when robotically rating solely on writing style, structural consistency, or language application. In addition, evaluators are limited to the fact that they cannot establish whether the particular essay is a real representation of the original ideas, research efforts, and analytical viewpoints of the supposed author and not just a rewording of already existing ideas previously compiled from other sources. (Hz et al., 2023; Smagorinsky et al., 2010).

The forecasted integration of AI into writing raises the need to reassess traditional assessment methodologies. As AI-generated texts get increasingly sophisticated, it challenging human evaluators were have to come up with novel strategies and complimentary tools to sort out the authenticity and originality of students' submissions. This brings out the inadequacy of current evaluative criteria applied in assessing student work and maybe substantially expands on their capacities by understanding AI's possibilities in effective impersonation of human writing ("Reviewing the Performance of AI Detection Tools in Differentiating between AI-Generated and Human-Written Texts: A Literature and Integrative Hybrid Review," 2024). The ethical considerations associated with AI usage in academia simply cannot be overlooked as the line begins to blur between assistance and academic dishonesty (Amirjalili et al., 2024). Towing from the above, it becomes very important that educational institutions start revamping assessment frameworks to accommodate these newly emerging challenges in order to ensure assessments continue remaining fair, valid, and reliable enough to measure a student's real writing competencies and intellectual inputs (Hz et al., 2023). Moreover, the increase of AI Writing has also raised concerns regarding artificially raising assessments of English proficiency. The test-taking population is going to use AI-written text instead of their own work, thereby compromised the credibility and integrity of the standardized English language proficiency evaluation. This trend creates serious problems for test administrators, who are required to negotiate technical and ethical decisions with regard to the extent to which a submitted writing sample was AI-generated (Labadze et al., 2023; Moorhouse et al., 2023). This has been exacerbated by sophisticated models of AI text generators, such as GPT-4, which generate content very similar to that of human writing and thus blur the line between a human and an AI-generated text. This would mean more robust methods of detection need to be deployed by administrators in such form as advanced algorithms and forensic linguistic analysis. Moreover, the ethical considerations in regard to such surveillance are deep. It becomes incumbent upon administrators to balance test integrity against requirements for respecting test-taker privacy and autonomy. AI detection tools could then be seen as intrusive or punitive in their intent and cause erosion of trust in the examination process (Weber-Wulff et al., 2023). Furthermore, the increasing reliance on AI in testing environments raises broader questions about the future of language assessment. As AI technologies progress, according (Abida et al., 2023). said that the criteria for assessing genuine language proficiency may need to be reevaluated to accommodate the impact of AI assistance. This evolving landscape underscores the importance of ongoing dialogue among educators, policymakers, and technologists to ensure that English proficiency assessments remain equitable, valid, and dependable in an

era of rapid technological progress (Alieto et al., 2024).

Then, there was previous research that dealt with the same topics in current research, as found by Toncic (2020) which conducted the research about how high school English teachers conceive of writing pedagogy and assessment in the classroom (Toncic, 2020). In this research, there was seven New Jersey teachers were interviewed about their current methods for teaching writing and evaluating students work. Then, they were then introduced to Grammarly and asked about incorporating such a tool. While teachers reported not explicitly teaching grammar, conventions like spelling factored into grades. Teachers felt AI checkers could serve as "personal assistants" to help students improve writing and ease teachers' workloads when marking mechanical errors. Open questions also surviving concern how these new technologies may impact policy regarding education and testing. Besides, Zhao et al. used a machine-generated score to the students' responses of picture-cued writing tasks (Zhao et al., 2023). This dataset created over 3000 Chinese middle school students' responses that describe 15 pictures. The approach uses cross-modal matching to measure the relationship between pictures and responses, evaluating content description. It also uses NLP to analyze linguistic features and quality. Results found the AI assisted scoring method can accurately assign scores. With high agreement to human ratings. Besides, Rahman et al (Hz et al., 2023) studied that examined the use of Grammarly automated writing evaluation (AWE) software to help improve university students' English writing skills. The findings showed that AI Grammarly was effective at detecting grammatical, spelling and punctuation errors but less so far content and organization. Students' perception was largely positive, with most feeling it improved their writing and motivation. Then, Rizqa et al (Rizqa et al., 2022). Explored pre-service teachers' experiences of cognitive conflicts during peer assessment facilitated by AI for an academic writing activity. An analysis of reflective journals found that peer feedback provided viewpoints but also led to cognitive conflicts when suggestions contradicted understanding or changed contexts. The last one which survey-based research elicits the views of English teachers on whether AI tools are useful to them as support for writing (Wulandari et al., 2024). Among the teachers, investigations showed that most agreed on the easily accessible tools improving skills in grammar and organization and increased their confidence in writing.

While previous research has examined the effectiveness of AI writing tools like Grammarly in helping improve students' writing skills and the perceptions of teachers and students on using such tools, the previous research has not fully explored the challenges of detecting AI-generated text and the implications of advanced AI writing for high-stakes English proficiency assessments. This current research, therefore, redresses these gaps through the qualitative analysis of challenges in detecting AI-written text faced by English language assessment experts and gives recommendations on possible adaptations to assessments in view of the increasing sophistication of AI writing technologies. This paper focuses on how advanced AI systems could probably deprive the validity of some important tests in the English Language. This research has two research questions: how the voices of English teachers give an English writing test in the AI era and how teachers overcome problems giving English writing tests in the AI era?

Research methods

This research involves qualitative research where the purpose of the selection of this method is to know subjectively the opinion of English teacher in giving English writing test to students. The qualitative approach is particularly suitable for this study as it allows for an in-depth exploration of the teachers' experiences, beliefs, and attitudes towards writing assessments. In this way, this research employs a phenomenological research design to better grasp the deeper understanding of the phenomenon being investigated. This ensured it

attains a better understanding of participants' lived experiences and perceptions. By focusing on the essence of participants' experiences, the researchers intend to unveil some of the latent meanings and themes that might come out of the phenomenon at hand. It contributed to broadening the knowledge base on the subject under study and had valuable implications for theory and practice in some field of study.

In the current study, 4 English teachers from different places across Indonesia were approached. All participants had different education backgrounds, starting from high school to higher education. Since the researchers had promised to keep the participants' identity both confidential and private, they cannot be named here. The diversity in the participants' backgrounds provided a broad spectrum of perspectives, enriching the study's findings. This variety also highlights the different approaches and methodologies employed in ELT across the region, contributing to a more comprehensive understanding of the subject.

Table 1. Participants Background

Participants (Code)	Gender	Institutional Level	Teachers' Experiences (Years)	Region
P1	Male	Senior High School	5 Years	East Java
P2	Female	Senior High School	4 Years	East Java
P3	Male	University	6 Years	Bali
P4	Female	University	6 Years	East Java

Semi-structured interviews in this research helped the researchers elicit deep personal information from the respondents. This research took place from April to June 2024, thus lasted three months. Observational methods were carried out on participants P1 and P2 in the process of English teaching. These observations were noted to provide information that is linked to data specific to the field. Furthermore, the focus groups discussion was held to further explore the views of participants regarding the assessment tools in testing the English writing ability of students. This provided an even greater understanding of their issues and experiences. A multi-method approach has been used to collect complete data in this research. Semi-structured interviews were able to garner more information from the participants since the open-ended nature of the questions encouraged respondents to express their personal experiences and insights. These observational methods gave the researchers a chance to get first-hand views of participants' practices and interactions within the natural teaching environment, hence useful contextual information. It enabled the researchers to dwell on some part of the participants' work in more detail and get a view as a whole with respect to the problems and considerations that incorporate administering English writing tests to students. The discussion in the focus group explained as to how this aspect fitted into the context of their testing policy and how it sat alongside other, perhaps competing policies.

During the phase of data collection, researchers carry out multiple crucial actions. At first, the participants are given an overview of the study's goals and the data was provided with. Afterwards, researchers carry out separate interviews with each participant. After that, two participants are observed while they are learning English. Lastly, a group meeting is organized to confirm answers and guarantee data accuracy, reducing the chance of bias in the information.

While gathering data from the participants the team of researchers dedicated a timeframe ranging from half an hour to an hour for every interview they conducted. In pursuit of simplicity and adaptability the interviews were carried out over the internet with Google Meet serving as the platform. By doing it this way everyone had the chance to be part of it

without the stress of sticking to a specific time or place. To keep things clear for all involved the discussions happened in Indonesian so there was not any mix-ups or unclear points. The goal of the researchers was to shape an environment that felt easy and welcoming so that those taking part could offer precise and deep responses to the questions posed by the study. Understanding these methods matters a lot. It is because they play a big role in making sure that what the study discovers can be trusted and stands up across different cultures.

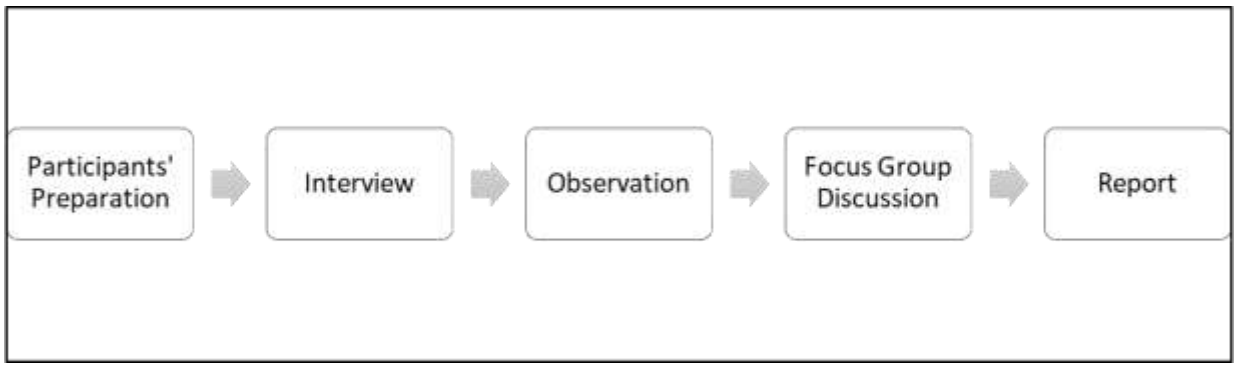

Figure 1. Data Collection Process

Table 2. Interview's Guideline

1.	As an English teacher/lecturer, how has the increasing prevalence of AI technology impacted the way you approach designing and administering writing assessments in your classroom?
2.	In an era where AI tools are capable of providing automated feedback and evaluation of student writing, how do you feel the role of the teacher's expertise and voice has evolved?
3.	Can you describe any specific strategies, techniques, or tools you've developed to ensure the authenticity and validity of your English writing assessments?
4.	What opportunities have you identified for incorporating AI tools and technologies into the English writing assessment process in a way that enhances, rather than replaces, the teacher's role?
5.	Looking ahead, what do you anticipate will be the greatest challenges and areas of focus for English teachers and lecturers when it comes to adapting writing assessments for the AI era?

Once they have gathered all the data the researchers move on to start looking at the numbers and information closely. At this key point they use a special way of understanding the data to pull out important points that help answer the questions they were trying to solve. The approach taken here dives deep into the data to discover hidden patterns and themes that were not obvious at first glance. Furthermore, the phenomenological reduction process entailed a meticulous examination of the interview transcripts to uncover the core meanings and essential structures underpinning participants' lived experiences with AI assisted writing assessments. In so doing, the researcher, through bracketing personal preconceptions and assumptions, found recurring themes, patterns within the data that shed light on what teachers were actually saying. The phenomena have been considering alternative interpretations of imaginative variations occurring across different points in space and time and, thereby, rendering space for the construction of a nuanced understanding of emergent challenges, strategies, and opportunities. The researcher integrated these multifaceted insights to develop a rich, textural, and structural description that captured the essence of the English teachers in this study as they wrote assessments in the

changed landscape of the AI era. Further validation was done through checking with the members, which ensured that the final phenomenological analysis truly represented the experiences of living teachers within the community. Definitely, such a rigorous phenomenological approach allowed investigating the phenomenon in great detail and giving a full, rich voice to the views of the teachers.

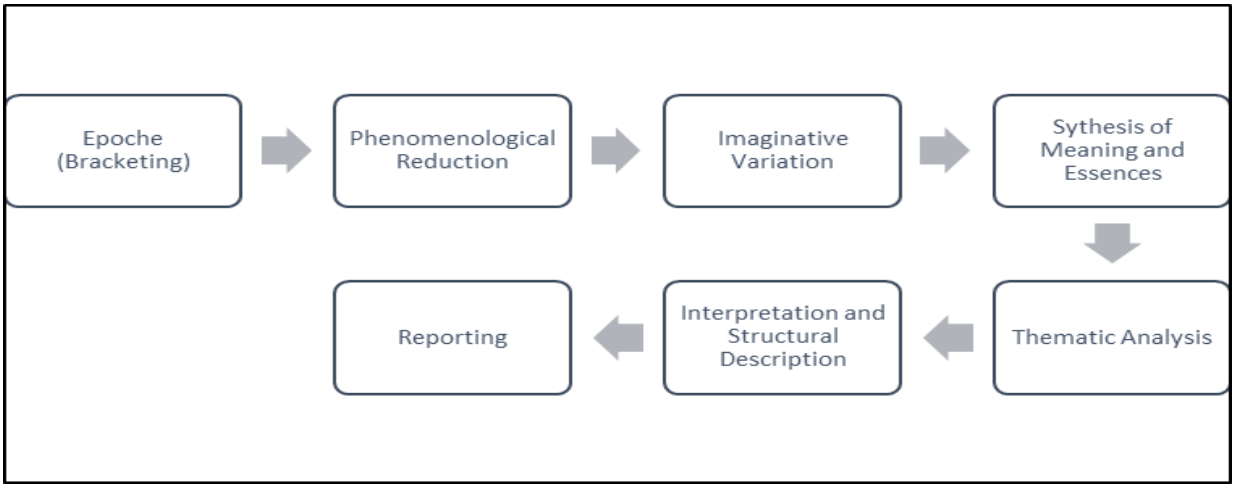

Figure 2. Data analysis process

Research findings and discussion

In the findings of this study, the researchers discovered various findings that have been found in the process of data collection and data analysis. The researchers grouped the findings from the participants into two major themes: the opinion of teachers in giving English writing tests to students, and how to overcome obstacles in the use of AI while teachers give student writing tests. The first theme focused on the views of the teacher participants regarding the process of administering English writing tests to their immediate students, the challenge it presented, best practice, and areas for improvement for such an approach to assessments. The second examines experiences and strategies participants have in navigating the use of AI technologies in student writing evaluation settings, with a focus on potential benefits incurred, limitations, and considerations that have to go into putting them into practice effectively.

The teachers opinion in giving english writing test

The study's findings elucidated a spectrum of perspectives among participating English teachers concerning the utilization of AI in the assessment of student writing. P1, an English teacher articulated reservation regarding the capacity of AI tools to accurately evaluate the nuanced dimensions of students writing, including creativity, critical thinking, and authorial voice. P1 underscored the significance of teachers' professional judgment in the evaluation process, emphasizing that teachers possess the ability to offer insightful feedback and guidance that fosters the enhancement of students' writing skills, like P1 excerpt:

> "While AI tools have potential to help streamline certain aspects of grading student writing, I remain skeptical for their capability to totally replace the human judgment, more so when dealing with higher skills. As English teachers, we are all trained to respond in ways that dig deeper into making students think more profoundly and engage ideas at a higher level of understanding. An algorithm could catch the missing apostrophes and run a paper against its training data for examples, but it will not be able to assess the subtle elements of style and voice that enliven a written work. What's more, it will not identify strengths and growth areas, uniquely, for any one student in that instinctive fashion that an educator can once said student is known. We don't just grade, but help students develop as writers over time through dialogue and guidance. It

threatens this fundamental human aspect of learning if we fall too deeply into the patterns of automated response." (P1)

Moreover, P1 commented that while the AI system was outstanding in picking up grammatical and structural mistakes, it failed to detect and appreciate the personality and originality of a particular student in writing. This can be a very serious deficiency, as it may forgo students' expressive abilities, resulting in homogeneous assessment. P1 contended that the human element in assessment was very important for the development of learners' minds and creativity. The role of the teacher was not just that of an examiner but also that of a mentor, motivator, and creator of an atmosphere in which students learn and feel comfortable, which cannot be replicated by any AI. Therefore, even though AI tools are useful and helpful to some extent, they can't replace such an important evaluation role of a teacher.

"While AI probably does work for a number of standardized tasks, this overreliance on algorithms in assessments is at the risk of losing a human element of assessment that is critical in writing development. We want each student to develop his or her own voice and creativity; we teach writing skills to ensure meaningful feedback and not just to catch grammatical mistakes. In that respect, especially the personalized guidance, support, and personalization of the teacher are a part of intellectual and creative growth—an area no AI can replicate. AI may at best supplement certain areas…but critical working evaluation has to stay with qualified human educators…" (P1)

In contrast, another English teacher, P2, did allow that AI-assisted writing assessments could have some benefits with respect to efficiency and consistency in grading. At the same time, P2 worried that AI might miss important contextual factors and that characteristic which is particular to students' writing. P2 believes that, to some extent, these limitations may be offset by some synergies available between human teachers and AI systems. This model allows some of those strengths of the AIs to be supplemented—notably objective grading and accounting for large volumes of work—by nuanced understanding and contextual awareness that the teachers bring to the assessment process.

"Although AI provides some advantages, such as increased consistency and speed in grading, it would be highly questionable whether this method could recognize major contextual details and student characteristics … .A collaborative model of human-AI would honor respective strengths—artificial intelligence for the first objective screening of huge workloads, and human judgment making definitive assessments with nuanced feedback…" (P2)

By integrating AI's capabilities with the professional judgment of teachers, this approach aspires to optimize the evaluation of student writing, ensuring that both efficiency and individualized attention are achieved. Moreover, P2 emphasized the importance of ongoing research and training to refine AI tools and enhance their sensitivity to the diverse range of student expressions and writing styes. This integrated method could serve as a robust framework for leveraging technology in educational settings, ultimately contributing to more effective and equitable writing assessments.

Furthermore, P3 and P4, lecturers in English, had a positive view regarding the adoption of AI in assessing students' writing. P3 showed great optimism in how AI could provide immediate responses and identify weaknesses in student writing. According to the participant, AI is trained to learn and be continuously improved so that the AI can respond to the nuisances and complexities typical of student writings. Another possible application of AI-assisted writing assessments shared by P4 is in reducing teachers' workload. The idea is that AI takes over routine evaluative tasks, so teachers can really focus on providing feedback and individual support to students. As such, P3 said:

"… I do envision algorithms offering rapid feedback on technical writing issues with continued training on student samples. Once refined, AI could strengthen the assessments in marking format and mechanics early, so that teachers can concentrate on deeper support. Subtleties still call for human discernment, yet prospects for scaled, personalized guidance through strategic

integration of automation are undeniable—so long as student learning remains at the forefront as these systems progress." (P3)

Moreover, both participants acknowledged the transformative potential of AI in educational contexts, positing that its integration could lead to significant enhancements in teaching and learning outcomes. P3 highlighted the importance of ethical considerations and the need to address potential biases within AI systems to ensure equitable assessments. Meanwhile, P4 suggested that AI could facilitate more consistent and objective evaluations of student writing, thus promoting fairness in the assessment process.

"AI could help promote fairer, more consistent evaluations by standardizing assessments based solely on quantifiable writing proficiency markers. This increased objectivity addresses concerns about biases impacting human scoring. A blended human-AI model harnessing standardized analysis alongside individualized mentor feedback could optimize evaluations to be strictly geared towards student growth, free from unrelated influences…" (P4)

Collectively, their insights suggest that, with appropriate development and implementation, AI could serve as a valuable tool in augmenting the efficacy of writing instruction and assessment.

The views from the English teachers give insight into both the opportunities and difficulties that come with incorporating AI into the grading process of the students' writing. The failure to develop a consensus from the skepticism through to optimism that this potentially showed in terms of what some members or delegates felt AI's role or function was into or about underscored the complexity of the problem, which needs more research and discussions. However, the optimistic views can be extended to the potentials of AI in making the assessment process more consistent and objective; thus, reducing teachers' workload and letting teachers' pay more concentration on personalized instructions (Dong, 2023; Pratama & Hastuti, 2024). Moreover, according to Hooda et al. (2022) mention, AI is capable of providing immediate feedback to students for a more iterative and responsive learning environment (Hooda et al., 2022). On the other hand, skeptical views have voiced out their fear that AI has a rather limited disposition in understanding the fine and rather subjective dimensions of writing—creativity, tone, and context. Critics state that if it heavily falls on AI, then these critical elements of writing would be devalued, adding that human oversight is necessary so holistic and accurate assessments are accurately accomplished. That spectrum of views suggests that putting AI into assessment is not something that can be done: it is something for which the potential benefits and possible drawbacks have to be carefully weighed (Hz et al., 2023). This makes further research, pilot programs, and educator professional development needed, not a pursuit of additional work simply for the sake of conducting it, but for the foundation of this often-murky territory that was enable the identification of the best and most fair methods possible to bring AI into the fold of student writing assessment.

It raises exactly the tension between measuring AI's objective, standardized abilities with nuanced, individualized human judgment in developing any form of writing. Indeed, P1 and P2 are dead on target when they identify where AI falls short in making sense of developed creative expression, critical thinking, and development over time. According to them, it is actually far from moving beyond basic tasks into the capabilities necessary for understanding more sophisticated processes of human creativity and the trajectory of the development of individual writers. On the other hand, P3 and P4 accept that the AI might develop ways of increasing feedback efficiency and consistency, propounding that with proper refinement, AI could be very instrumental in writing instructions This would need to be refines into AIs that are super-sensitive to subtleties of human expression and can respond with idiosyncratic feedback that sustains growth, thus working towards individual growth. On the other hand, Moroianu et al. (2023) argue that the implication of Artificial Intelligence in the

educational process is very much to be placed at the intersection of these two opposing viewpoints (Moroianu et al., 2023). How does one ensure that some of the more important capabilities Artificial Intelligence has to offer regarding processing volumes of text so much more quickly and consistently are not lost as National Teaching and Evaluation Service continues to meet some quite unique and developing needs through individualized, incisive, human feedback (Dergaa et al., 2023). In this respect, the future development of writing based on the symbiotic relationship between AI and human instructors so that the strength of one capitalizes on the potential of the other toward more efficient and individualized learning experiences.

One of the more productive paths forward that P2 proposes is the development of a model for collaboration between a human and AI, one that can take the best out of both entities' analytical strengths. For example, AI might be primed for preliminary screening and highlighting of special problems; these would then serve as a foundation for holistic formative feedback by teachers. Continuous training tries to sensitize the AI to variations of texts and situations. Moreover, there is a need to have engagement with the critical questions of bias, fairness, and student learning outcomes. According to Rezai et al (2022), the researcher could see if the hybrid evaluation methods optimized for student growth would lead to more equity of learning gain compared to the fully automated assessment, which always bends towards proficiency (Rezai et al., 2022). Finally, there are ethical issues of student data usage and minimization of the potential adverse impact on society that need to be discussed at great length.

Moreover, it can redefine educational assessment by the integration of a human-AI collaborative model in the learning environment. It was not just guaranteeing fairness but also idiosyncratic learning experiences (Roy & Paul, 2023). Studies on such hybrid models would be of utter importance in showing how technology is being used to drive fair educational outcomes (Dr. Neeraj Yadav, 2024). Hence, the balance in this dynamic educational environment should be ensured by the stakeholders, as student development and success are promoted, but the ethical issues so far translated would not be blind to the AI applications.

How to overcome obstacles in the use of ai while teachers giving students writing test

The participants also discussed strategies to reduce the challenges associated with using AI in assessments of student writing. P3 added that when using AI-assisted assessment tools in writing, there had to be a need for extensive professional development programs, focused on building the capacity of the teacher for better aligning his or her teaching practices with AI technology. This would better prepare or equip the teachers to optimum ways of using AI tools to enhance and complement the in-use/practice assessment methodologies.

> "Any rollout of automated assessment tools must therefore go hand in hand with comprehensive teacher training in order to realize AI's potential benefits while mitigating risks. These include strategic skills in the checking of AI feedback, discarding its analysis when it may be insufficient, and making value judgments about how this supplementary information will be best integrated into their own practices. Professional development programs can be designed to truly empower instructors with knowledge to make AI a value-adding partner, not one that replaces them. Armed with expertise around these new technologies, teachers will then be best positioned to exploit automation's strengths without compromising student-centered pedagogies or important evaluative nuances. Such guidance is enhanced by AI in defeating poor assessment quality by being judiciously blended into the unique methodology of each instructor." (P3)

In addition, such professional development activities would concentrate on working with AI tools themselves, and highlight pedagogical approaches that reap from AI possibilities. This two-pronged approach makes sure that educators critically examine and are positioned to adapt AI recommendations to suit the particular needs of their students. This enabled teachers to adopt a more balanced and nuanced

approach toward assessment, armed with a better appreciation of the potential and limitations of AI. Further, the on-going support and collaboration events for educators to share best practices and challenges formed another basis for how AI was harnessed successfully within writing assessments.

Moreover, careful monitoring of the performance of AI systems and maintenance of human oversight would ensure good quality and fairness in the assessment process. This exemplifies the manner in which continued testing and fine-tuning for the setting of limits or biases as they manifest themselves in AI systems take place. This should be a balanced kind of approach where the educator and AI collaborate toward student learning and development.

> "Even sophisticated algorithms require diligent human oversight. Ongoing monitoring and assessment of AI is important to maximize benefits while reducing risks. It should include collaboration, not complete automation, for the assessment as a way to quickly identify and reduce emerging limitations or biases. There needs to be a continuous process of refinement as technologies continue to progress. Above all, the focus has to remain student-centered—a focus on the advancement of learning rather than merely automation. Most reasonable and promising is a balanced and judicious approach with AI guided by empowered and educated educators…."(P4)

AI integration in the educational environment should be looked upon as providing complementary tools that can help in enhancing capability, rather than replacing human judgment. Here, AI helps teachers realize insights that might otherwise get overlooked and, therefore, implement personalized and wholesome teaching strategies. More than that, such continued interaction between AI systems and human educators was aided in recognizing and rectifying the unconscious biases or errors in the functioning of the AI algorithms. Hence, this collaborative relationship between AI and teacher trigger an innovative and just educational environment that finally develop the students holistically.

The participants, particularly P1 and P2, both lecturers in English, thoroughly discussed ways of surpassing the challenges of using AI in the assessment of students' writing. As P1 put it, the role of the professional judgment of teachers should not be forgotten within the assessment scenario. Though P1 conceded that AI writing evaluation has potential advantages regarding efficiency, he indicated how one cannot move beyond the basic comforts of a teacher's expertise in delineating subtler dimensions of student writing accurately with respect to creativity, critical thinking, and voice.

> "…While AI shows efficiency advantages, it cannot surpass the critical role of any teacher's expert judgment in evaluation. We, as trained professionals, assess writing holistically, seeing dimensions beyond mechanics that are nuanced. It is in these dimensions—the creativity, thinking, and voice—arguably human discernment drawing from deep pedagogical knowledge and experience built over time. Preserving the first place of teacher expertise within all blended assessment models should be paramount. It is conceivable that AI can only complement this qualitative analysis, not replace it." (P1)

P1 then suggested that to offset the limitations in-built into AI, the teacher should remain central in the assessment process for the purposes of providing invaluable feedback and guidance, which foster improvement in writing skills for students. Besides, argued P1, the technology, though promising, is still at a stage of development and lacks capacity for an understanding of the subtlety of human expression and originality.

> "To truly address these limitations of AI at present, teachers must continue playing a pivotal role in direct engagement with the students. Through thought-making feedback, we can offer guidelines about that particular learner to progress for developmental improvement in writing skills. Standardized scores contribute very little to growth. Only personalized evaluations and mentorship from a human can capture the nuance of work or the individual behind it. While AI is leading to such idyllic humanistic understanding, there is still so much that remains to be done concerning creative expression, understanding, and originality. For now, while

automation may successfully complement consistent markers, nurturing students requires the expertise and care that only the educator with years of experience can impart. A blended model works best if the place of teachers remains integral to both assessment and learning experience." (P1)

Hence, integration of AI should not be perceived as replacing human evaluators but rather complementing their role. In this direction, what that could work toward is an approach that's collaborative in nature between the AI systems and educators leading to a much more balanced and wholesomely effective assessment framework, which is allow AI to focus on routine, objective aspects of valuation so that teachers can work more on subjective and complex elements. This may turn out to be a synergy that can transform the face of education in teaching and learning by associating the strengths of AI with human insight.

The views of P1, P2, P3, and P4 provided valuable guidance about strategies for negotiating the challenges associated with the integration of AI into writing assessment. The central theme was that, within any blended model, there is a need to preserve a central role and area of expertise for the teacher. As reiterated by P1, while educators have pedagogical knowledge, the qualitative elements such as creativity, their caliber remains as yet unachieved by AI. Broad training for teachers seems to be paramount in terms of ensuring that the human effect is maintained in this AI-infused climate. According to P3, professional development programs could empower teachers with strategic skills in ways that would help them make judicious use of AI as a supplement rather than its replacement. Training would be needed both in technical operation of the tools and pedagogical approaches on how thoughtfully an individual teacher should blend the AI's recommendations within his practice (Aggarwal, 2023; Park & Doo, 2024).

Additional integrations could be supported due to further work of collaborative communities, as P4 proposes. This, again, would have an associated value with its implementation improved iteratively over time, based on collective lessons learned. Core to this strategy would be the continued evaluation of AI systems under human supervision, including with effective detection and mitigation of new limitations or biases that might come up (Aggarwal, 2023; Mennella et al., 2024). Views generally point toward a synergistic model of human-AI if developed through evidenced-based, student-centered means. It is in this regard that teachers, through personalized guidance, remain at the very center of assessment, whereas AI handles only a standardized part to bring in more efficiency to the process (Hadi Mogavi et al., 2024). Within this context, strategies—especially sorted-out training programs and collaborative communities—are among the most critical facilitators toward actualizing the accruable benefits from a blended approach and mitigating risks related to educational quality, equity, and individual support to students (Marquez et al., 2023). Further research is conducted on how to investigate the optimal design and effects of those integrative partnerships. These complex strategies can be read about and provide a thoughtful framework from which to build in terms of integration of AI into pedagogy as it further advances. Diligent and collaborative efforts are required across both educational research and practice domains in order to ensure a human element remains integral.

Conclusion

The case was on the significant challenges that the English teacher encounters when grading a writing test in the era of advanced AI writing systems. Findings indicate that most cases have human assessors failing to differentiate between texts developed by humans and those created by AI, as these AI-powered writing tools become highly sophisticated in writing like a human being. This serious question arises in the credibility and integrity of standardized English tests when test-takers might opt to send AI-generated texts instead of their own work. This paper proposes revision of traditional assessment

methodologies of integration with AI into writing assessment for the purpose of overcoming these problems. As AI-generated texts continue to evolve, human evaluators was have to develop new methods and technologies that help in this process of authentication and verification of originality in student works. It may seriously revisit the current assessment criteria and better detection modes—which are mostly algorithms and forensic linguistic analysis—to realize, in talks between educators, policymakers, and technologists. This paper is a call for further research in the examination of ethical dimensions relating to AI usage in an academic setting, especially in terms of when this aid becomes a factor in academic dishonesty. That redefined what genuine language proficiency constitutes as AI technologies evolve. Further studies are needed to determine the way in which the approach to assessment frameworks within the institutions is going to change to avoid these new challenges and ensure such assessments in the future will be able to better reflect a student's actual writing ability and intellectual input.

Disclosure statement

The authors declare no conflict of interest

References

Abida, F. I. N., Kuswardani, R., Purwati, O., Rosyid, A., & Minarti, E. (2023). Assessing Language Proficiency through AI Chatbot-Based Evaluations. *Proceedings of International Conference on Islamic Civilization and Humanities*, *1*, 138–145.

Aggarwal, D. (2023). Integration of innovative technological developments and AI with education for an adaptive learning pedagogy. *China Petroleum Processing and Petrochemical Technology*, *23*(2), 709–714.

Alieto, E., Abequibel-Encarnacion, B., Estigoy, E., Balasa, K., Eijansantos, A., & Torres-Toukoumidis, A. (2024). Teaching inside a digital classroom: A quantitative analysis of attitude, technological competence and access among teachers across subject disciplines. *Heliyon*, *10*(2), e24282. https://doi.org/10.1016/j.heliyon.2024.e24282

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*, *9*. https://doi.org/10.3389/feduc.2024.1347421

Creely, E. (2024). Exploring the Role of Generative AI in Enhancing Language Learning: Opportunities and Challenges. *International Journal of Changes in Education*. https://doi.org/10.47852/bonviewIJCE42022495

Dergaa, I., Chamari, K., Zmijewski, P., & Ben Saad, H. (2023). From human writing to artificial intelligence generated text: examining the prospects and potential threats of ChatGPT in academic writing. *Biology of Sport*, *40*(2), 615–622. https://doi.org/10.5114/biolsport.2023.125623

Dong, Y. (2023). Revolutionizing Academic English Writing through AI-Powered Pedagogy: Practical Exploration of Teaching Process and Assessment. *Journal of Higher Education Research*, *4*(2), 52. https://doi.org/10.32629/jher.v4i2.1188

Dr. Neeraj Yadav. (2024). The Impact of Digital Learning on Education. *International Journal of Multidisciplinary Research in Arts, Science and Technology*, *2*(1), 24–34. https://doi.org/10.61778/ijmrast.v2i1.34

Fui-Hoon Nah, F., Zheng, R., Cai, J., Siau, K., & Chen, L. (2023). Generative AI and ChatGPT: Applications, challenges, and AI-human collaboration. *Journal of Information Technology Case and Application Research*, *25*(3), 277–304. https://doi.org/10.1080/15228053.2023.2233814

Hadi Mogavi, R., Deng, C., Juho Kim, J., Zhou,

P., D. Kwon, Y., Hosny Saleh Metwally, A., Tlili, A., Bassanelli, S., Bucchiarone, A., Gujar, S., Nacke, L. E., & Hui, P. (2024). ChatGPT in education: A blessing or a curse? A qualitative study exploring early adopters' utilization and perceptions. *Computers in Human Behavior: Artificial Humans*, *2*(1), 100027. https://doi.org/10.1016/j.chbah.2023.100027

Han, B., Nawaz, S., Buchanan, G., & McKay, D. (2023). *Ethical and Pedagogical Impacts of AI in Education* (pp. 667–673). https://doi.org/10.1007/978-3-031-36272-9_54

Hooda, M., Rana, C., Dahiya, O., Rizwan, A., & Hossain, M. S. (2022). Artificial Intelligence for Assessment and Feedback to Enhance Student Success in Higher Education. *Mathematical Problems in Engineering*, *2022*, 1–19. https://doi.org/10.1155/2022/5215722

Hz, B. I. R., Laiya, R. E., & Sarumaha, M. S. (2023). Navigating English Writing Proficiency Tests in the Era of Artificial Intelligence. *Journal of English Education and Teaching*, *7*(3), 480–498.

Labadze, L., Grigolia, M., & Machaidze, L. (2023). Role of AI chatbots in education: systematic literature review. *International Journal of Educational Technology in Higher Education*, *20*(1), 56. https://doi.org/10.1186/s41239-023-00426-1

Marquez, R., Barrios, N., Vera, R. E., Mendez, M. E., Tolosa, L., Zambrano, F., & Li, Y. (2023). A perspective on the synergistic potential of artificial intelligence and product-based learning strategies in biobased materials education. *Education for Chemical Engineers*, *44*, 164–180. https://doi.org/10.1016/j.ece.2023.05.005

Mennella, C., Maniscalco, U., De Pietro, G., & Esposito, M. (2024). Ethical and regulatory challenges of AI technologies in healthcare: A narrative review. *Heliyon*, *10*(4), e26297. https://doi.org/10.1016/j.heliyon.2024.e26297

Mohamed, Y. A., Khanan, A., Bashir, M., Mohamed, A. H. H. M., Adiel, M. A. E., & Elsadig, M. A. (2024). The Impact of Artificial Intelligence on Language Translation: A Review. *IEEE Access*, *12*, 25553–25579. https://doi.org/10.1109/ACCESS.2024.3366802

Moorhouse, B. L., Yeo, M. A., & Wan, Y. (2023). Generative AI tools and assessment: Guidelines of the world's top-ranking universities. *Computers and Education Open*, *5*, 100151. https://doi.org/10.1016/j.caeo.2023.100151

Moroianu, N., Iacob, S.-E., & Constantin, A. (2023). Artificial Intelligence in Education: a Systematic Review. In *Geopolitical perspectives and technological challenges for sustainable growth in the 21st century* (pp. 906–921). Sciendo. https://doi.org/10.2478/9788367405546-084

Park, Y., & Doo, M. Y. (2024). Role of AI in Blended Learning: A Systematic Literature Review. *International Review of Research in Open and Distributed Learning*, *25*(1), 164–196.

Pratama, R. M. D., & Hastuti, D. P. (2024). The use of artificial intelligence to improve EFL students' writing skill. *English Learning Innovation*, *5*(1), 13–25.

Raheem, B. R., Anjum, F., & Ghafar, Z. N. (2023). Exploring the profound impact of artificial intelligence applications (Quillbot, Grammarly and ChatGPT) on English academic writing: A Systematic Review. *International Journal of Integrative Research (IJIR)*, *1*(10), 599–622.

Rahmi, R., Amalina, Z., Andriansyah, A., & Rodgers, A. (2024). Does it really help?

Exploring the impact of Al-Generated writing assistant on the students' English writing. *Studies in English Language and Education*, *11*(2), 998–1012. https://doi.org/10.24815/siele.v11i2.35875

Reviewing the performance of AI detection tools in differentiating between AI-generated and human-written texts: A literature and integrative hybrid review. (2024). *Journal of Applied Learning & Teaching*, *7*(1). https://doi.org/10.37074/jalt.2024.7.1.14

Rezai, A., Namaziandost, E., Miri, M., & Kumar, T. (2022). Demographic biases and assessment fairness in classroom: insights from Iranian university teachers. *Language Testing in Asia*, *12*(1), 8. https://doi.org/10.1186/s40468-022-00157-6

Rizqa, K., Drajati, N. A., & Putra, K. A. (2022). PRE-SERVICE TEACHERS'PERCEPTIONS ON PEER ASSESSMENT PROCESS IN LEARNING ACADEMIC WRITING: ARTIFICIAL INTELLIGENCE (AI). *ICONELT*, *3*, 50–58.

Roy, S., & Paul, S. K. (2023). Revolutionizing Education: How Artificial Intelligence is transforming the Learning Landscape. *Int. J. Trend Sci. Res. Dev*, *7*, 748–756.

Shanorra, T. T., Sofyan, R., & Sumbayak, D. M. (2021). A writing skill assessment of the first semester English department students of the Universitas Sumatera Utara. *VISION*, *17*(2).

Smagorinsky, P., Daigle, E. A., Allen, C. O., & Bynum, S. (2010). Bullshit in academic writing: A protocol analysis of a high school senior's process of interpreting Much Ado About Nothing. *Research in the Teaching of English*, *44*(4), 368–405.

Toba, R., Noor, W. N., & Sanu, L. O. (2019). The Current Issues of Indonesian EFL Students' Writing Skills: Ability, Problem, and Reason in Writing Comparison and Contrast Essay. *Dinamika Ilmu*, 57–73. https://doi.org/10.21093/di.v19i1.1506

Toncic, J. (2020). Teachers, AI grammar checkers, and the newest literacies: Emending writing pedagogy and assessment. *Digital Culture & Education*, *12*(1), 26.

Weber-Wulff, D., Anohina-Naumeca, A., Bjelobaba, S., Foltýnek, T., Guerrero-Dib, J., Popoola, O., Šigut, P., & Waddington, L. (2023). Testing of detection tools for AI-generated text. *International Journal for Educational Integrity*, *19*(1), 26. https://doi.org/10.1007/s40979-023-00146-z

Wulandari, F., Astuti, M. T., & Marhamah, M. (2024). Enhancing writing literacy teachers' through AI development. *Jurnal Onoma: Pendidikan, Bahasa Dan Sastra*, *10*(1), 246–256.

Zhao, R., Zhuang, Y., Zou, D., Xie, Q., & Yu, P. L. H. (2023). AI-assisted automated scoring of picture-cued writing tasks for language assessment. *Education and Information Technologies*, *28*(6), 7031–7063. https://doi.org/10.1007/s10639-022-11473-y