International Journal of Learning, Teaching and Educational Research
Vol. 25, No. 2, pp. 281-305, February 2026
https://doi.org/10.26803/ijlter.25.2.14
Received Oct 13, 2025; Revised Jan 12, 2026; Accepted Jan 29, 2026

Exposing ChatGPT-Assisted Plagiarism in Student Assessment in Higher Education: Linguistic and Non-linguistic Clues

Faisal Al-Maamari[*]
English and Translation Department,
Sultan Qaboos University,
Oman

Abstract. ChatGPT offers second language writers' limitless opportunities for engagement with this technology. This exploratory study focuses on Gen-AI academic plagiarism in the context of unsupervised written assessment and pursues the concept of opportunity in traditional academic fraud theory, in an attempt to evaluate its applicability to Gen-AI academic plagiarism. The research questions aimed to gather perceptual and textual observations by departmental faculty regarding their students' unpermitted use of ChatGPT in written assessments. Thirteen experienced faculty members in the English and Translation Department at a publicly funded university in the Sultanate of Oman completed a questionnaire asking them to identify clues of plagiarism evident in their students' written work. Additionally, assessment artefacts in the form of 15 student-authored literature reviews were examined in search of these clues. Using an inductive, mixed-methods approach, the analysis drew on faculty members' growing understanding of the affordances of large language models, coupled with their situated knowledge of their students' writing abilities in terms of the lexico-grammatical and discoursal features characterising their submitted texts. The findings were summarized in a model which highlighted the interrelationships amongst the various factors leading to writer disengagement principally manifested through language, subject-matter, and behavioural clues. The paper concludes by adopting a utilitarian, pragmatic perspective on academic plagiarism, with a view to transforming these limitations into opportunities for writer engagement and, ultimately, learning.

Keywords: ChatGPT and Gen AI-based Plagiarism Detection; Learning; Opportunity; Writer Disengagement; Written Assessment

[*]Corresponding author: Faisal Al-Maamari; *faisalf@squ.edu.om*

1. Introduction

Large language models (LLMs) are advanced artificial intelligence (AI) systems trained on massive datasets which empowers them to understand and generate human language in various forms. In spite of their recent emergence, LLM chatbots such as ChatGPT have been described as a "societal disruption" (Mennella & Quadros-Mennella, 2024, p. 3). Likened by some to the innovations such as the calculator, typewriter, word-processor, or AI-enhancing tools such as Grammarly, others argue that the challenges they pose to education are real and pervasive, including "the death of the essay, and the development of language and complex thought" (Culp Jr, 2023, p. 2). Itself an LLM, ChatGPT is a prompt-based conversational technology which allows users to interact with the interface in a human-like fashion and excavate a great deal of information.

Gen-AI offers users numerous benefits for second language writing including increased engagement on task, improved writing skills, formative and evaluative feedback, and exposure to diverse English in use (Barrot, 2023; Thao et al., 2023). They are also useful in content generation and the organization of ideas (Mahapatra, 2024). Tsai et al. (2024) also found dramatic improvements in ChatGPT-revised essays written originally by English major students. However, ChatGPT also raises some concerns, including second language writers' overreliance on it, coupled with diminished levels of creativity and expression, and accuracy of input (Thao et al., 2023).

Second language ChatGPT users further exhibit reduced attention to grammatical accuracy (Mahapatra, 2024), and the erosion of critical thinking and problem-solving skills (Rezaei et al., 2024). From an equitable assessment perspective, Tsai et al. (2024) stated that supervised ChatGPT use by teachers brought about "disproportionate improvement" in essay scores which were unreflective of students' true writing ability, also noting a concern relating it to potential fraudulent behaviour (Cotton et al. 2023; Jarrah et al., 2023). These new developments cast the notion of academic plagiarism in second language writing into new light, making ChatGPT-assisted plagiarism a research priority.

Although mainstream, traditional plagiarism is pervasive in academia (Burke & Sanney, 2018), ChatGPT has accentuated this through an intelligent writer, tipping off the scale to the benefit of ill-meaning, and apathetic students, who have been identified by research as the category most at-risk (Galindo-Domínguez, 2025; Playfoot et al., 2024). Students' confusion regarding the role of ChatGPT as a potential tool to generate content and improve their writing and as an assignment writer adds to this murky picture (Mennella & Quadros-Mennella, 2024). Superseding the conventional methods of academic fraud, newer methods of plagiarism have included the use of Gen-AI chatbots (e.g., Deepseek & GPT 3.5), plus the concomitant use of AI-paraphrasing or humanizing tools (e.g., Quillbot) (Yan, 2003), with the result that students are "often spoilt for choice" (Baron, 2024, p. 157). This suggests that in any academic assignment, a student may outsource content in various, unpredictable and unprecedented ways.

This uncertainty has also spread to the methods used for plagiarism identification and detection. As we speak, the literature affirms that no technology (be it AI or otherwise) exists that can offer bullet-proof Gen-AI plagiarism detection and protection (Baron, 2024; Pudasaini et al., 2024) with conventional and AI-powered anti-plagiarism tools reporting ineffectual reliability levels (Patel et al., 2011). Further, it is becoming harder for authorship attribution detectors to distinguish human and machine texts, simply because LLMs continue to develop with the same ferocity and potency (Clark et al., 2021; Uchendu, 2021). Teachers also "may have difficulty distinguishing between the students' work and the text generated by ChatGPT" (Barrot, 2023, p. 4), with instructor detection power is at best modest (Liu et al., 2023; Uchendu, 2021).

At one level, the availability of multiple methods of plagiarism, students' proclivity to abuse Gen-AI, and the inadequacy of anti-plagiarism technology and teachers to identify Gen-AI plagiarism present increased opportunities for Gen-AI plagiarism. At another level, they present opportunities to investigate them empirically. The aim of the study is to put the financial fraud triangle theory by Donald Cressey (1953) adapted by Burke and Sanney (2018) to the test by assessing the applicability of its opportunity construct in the AI new environment. It is assumed in this study that the opportunity construct provides increasing affordances for learners to commit Gen-AI plagiarism in written assessments.

Many studies seized this opportunity, but they were not informed by a similar theoretical orientation. For example, AlAfnan and MohdZuki (2023), Liu et al. (2023), Desaire et al. (2023), and others compared machine-generated texts to human texts to identify ChatGPT-identifying lexico-grammatical, discoursal and content features. Findings indicated the syntactic superiority of machine-generated texts. Other research reported inadequate pragmatic awareness and content-curriculum misalignment of output produced by LLMs (AlAfnan & MohdZuki, 2023; Desaire et al., 2023; Meishar-Tal, 2024).

To give one example, in Saudi Arabia, Almanea (2024) reported that her English as a Foreign Language (EFL) instructors associated ChatGPT outputs to possess a general writing style, characterized by being overly objective and impersonal. The instructors also reported ChatGPT outputs to consist of fabricated information and sparse content lacking in the use of academic references. Further, sentences were found to be exceptionally lengthy and syntactically complex, and ideas lacking in coherence. Vocabulary consisted of low-frequency words, which are often not aligned with the learners' language proficiency.

The AI lexico-grammatical, discoursal and content features reported above are anecdotal in nature, or tangential to the research reporting them. They were also not grounded in pedagogical practice in that they did not draw on teachers' actual pedagogical and assessment practices, which are significant to utilize not only because of the obvious limitations of current AI-Generated Content (AIGC) detection technology, but also owing to the increasing orientation in the extant literature to integrate teaching, learning and assessment. To address this gap, the present study is premised on the notion that "identify[ing] gaps between the teacher's expectations concerning students' linguistic proficiency and knowledge and the output generated by the chatbot" (Meishar-Tal, 2024, p. 708) is key to Gen-AI plagiarism identification. As such, it is a study bound by time and place and context and focused squarely on Gen-AI plagiarism. It specifically aims to explore instructors' situated knowledge of ChatGPT-assisted academic plagiarism in applied linguistics and literature courses and analyse localized students' writing practices regarding the same issue in a tertiary setting in the Sultanate of Oman. Finally, the study offers an empirical test of the opportunity construct in traditional academic fraud theory and its applicability in the Gen-AI era.

2. Literature Review

Academic plagiarism in writing is essentially the appropriation of materials from a source without proper attribution (Jarrah et al., 2023; Pecorari, 2001). While plagiarism has always been a strongly felt reality in academia, large language models (LLMs) such as ChatGPT accentuate this (Cotton et al. 2023), and perhaps sometimes unevenly to the advantage of struggling learners.

2.1. Theoretical Framework: Pressure, Opportunity and Rationalization

To understand the current situation of Gen-AI academic plagiarism, I draw here on the financial fraud triangle theory by Donald Cressey (1953) and adapted by Burke and Sanney (2018) for traditional academic plagiarism i.e. before the advent of LLMs. The academic integrity triangle theory posits that there are three components of academic plagiarism, namely pressure, opportunity and rationalization:

First, students perceive some urgency about their grades; in other words, they perceive that they must attain a certain grade and will not be successful in that quest without resorting to academic dishonesty. Second, students are afforded an opportunity to cheat in a variety of formats. Third, students rationalize that it is acceptable to take advantage of these opportunities to cheat for any number of reasons. (p. 11)

The three components are self-explanatory, but the advent of LLMs requires a reconsideration of these components in their contribution to academic plagiarism. For practical considerations, however, this paper considers only the construct of opportunity, suffice it to mention that the pressure to plagiarize remains principally unchanged, which is to maximize academic success and extinguish failure. No further discussion is also made of the rationalization component, as students' cognitions were not of interest in this study. A reconsideration of the opportunity construct, however, is essential along several parameters in relation to methods of plagiarism and technologies of detectability, but also through factors related to teachers and learners.

2.2. Opportunity at the Level of Methods of Plagiarism

After AI has entered the race, not only methods of plagiarism have multiplied, but they have also blurred boundaries (Yan, 2003). Chronologically, before Gen-AI, methods popularly employed by students to commit plagiarism in externally writing assessments included appropriating information from traditional print sources, copying and pasting information from an internet source, the use of traditional paraphrasing websites, and purchasing papers from online paper mills. With the advent of LLMs, Gen-AI chatbots such as Deepseek and GPT 3.5, and paraphrasing technologies such as Gen-AI humanizing tools (e.g., Quillbot) have joined the mix, potentially offering plagiarists multiple and unprecedented mixed-methods to commit plagiarism, leading one researcher to comment that students are "often spoilt for choice" (Baron, 2024, p. 157.) In her study of Saudi undergraduate English major students, Almanea (2024) reported that ChatGPT's two conspicuous features, namely allowing distinct multiple responses for the same prompt and generating a word limit response, undermine detectability.

2.3. Opportunity at the Level of Anti-plagiarism Technology

Before Gen-AI, when instructors were unable to identify academic theft, conventional anti-plagiarism tools soon came to the rescue. Based on the literature, there is evidence to suggest these to be robust for that particular plagiarism type (Patel et al., 2011). However, currently, the literature affirms that no technology (be it AI or otherwise) exists that offers bullet-proof Gen-AI plagiarism detection and protection, leading Baron (2024) to conclude that we may have reached a stage whereby large language models (LLMs) "may inadvertently be supporting plagiarism rather than reducing it" (p. 151). Additionally, Patel et al. (2011) reported that Turnitin.com, iThenticate and PlagiarismDetect.com "can be cracked through different methods" (p. 623).

Principally, Artificial Intelligence Generated Content (AIGC) detection faces two formidable challenges related to the effectiveness of existing methods to curtail plagiarism and the constantly accelerating advancement in LLMs. The latest scholarship in AIGC detection fails to capture whether a specific text is machine-generated or human-written or what LLMs authored the machine texts (Clark et al., 2021; Uchendu, 2021). Khalil and Er (2023), for example, reported on a success rate of only 20% of Turnitin and iThenticate to identify 50 ChatGPT-generated essays, and based on their review Pudasaini et al. (2024) stated that AIGC detection "does not exist and may not exist" (p. 15). What complicates matters further is the existence of an action-reaction pattern between the plagiarism and anti-plagiarism apparatus, likened to a dog tail chase situation, or a "technical war" (Pudasaini et al. 2024, p. 15).

This also means that the increased availability of LLMs offers textual outputs which are increasingly "becoming more similar to human-written texts in styles" and will likely morph in the future (Uchendu, 2021, p. 8), enabling machine-generated texts to fly under the anti-plagiarism radar undetected. Pudasaini et al. (2024) reported that in the absence of any progression in AIGC detection and evasion techniques normally adopted by students such as recursive paraphrasing, authorship obfuscation and traditional word/phrase substitutions, any breakthrough may be a moot point (p.2), in effect leading any advancements in these technologies to effectively cancel one another.

2.4. Opportunity at the Level of Learners

There is no doubt as to the linguistic superiority of Gen-AI outputs compared to students' writing. Tsai et al. (2024), for example, found that the overall quality of ChatGPT revisions on essays originally written by English major students led to dramatic improvements in terms of vocabulary, grammar, organization and content. Numerous other studies confirmed this supremacy (Barrot, 2023; Mahapatra, 2024; Thao et al., 2023). Although not all students are willing to cheat, it is the most vulnerable who normally do, with some research concluding that the only predictor of AI use in written assessments was degree apathy i.e. a student's lack of academic motivation or lack of concern for study (Playfoot et al., 2024). Burke and Sanney (2018) also argue that some "students seem to thrive on putting more effort into figuring out how *not* do [sic] the assignment rather than just doing it" (p. 5, emphasis in original).

With the advent of Gen-AI, it could also be true that students attempt to thwart efforts to invest further in the learning process, as ChatGPT is seen as an easy way out. Further, requiring students to acknowledge its use presupposes fact-checking, which students perceived to be an extra burden, as the case was with the students in Mennella and Quadros-Mennella (2024). Further, students also tend to plagiarize when they perceive they will evade punishment and/or when they perceive the benefit of cheating to outweigh the cost of punishment (Playfoot et al., 2024).

These same students also tended to have misconceptions regarding Chat GPT. Mennella and Quadros-Mennella (2024) investigated college students' perceptions of ChatGPT for their assignments at a private comprehensive university in the US. Students had two perceptions: one group viewed ChatGPT as a potential tool to improve their writing and generate content for their assignment, whereas the second group viewed it as an assignment writer. Additionally, some held other misconceptions, often considering ChatGPT as a source of information and a search engine like google.

Finally, technology is not uniformly equitably accessible to all learners (Cotton et al. 2023), but even when accessibility to AI technology is not an issue, ChatGPT has both created and maintained a new kind of AI gulf in favour of expert users of technology (Fajt & Schiller, 2025). Additionally, ChatGPT has created the same gulf between originally high-achieving learners and originally low-achieving learners, the latter of whom may be more inclined (even pressurized) to indulge in readily available opportunities to use AI for the wrong purpose: to short-circuit necessary learning. Simultaneously also, this unfair abuse of ChatGPT by the latter group may bringing them *on par* with the high-achieving learners, which is really unreflective of their true ability (Tsai et al., 2024).

2.5. Opportunity at the Level of Teachers

According to Barrot (2023), "teachers may have difficulty distinguishing between the students' work and the text generated by ChatGPT" (p. 4). For example, the two instructors recruited to blind grade the essays in the study by Tsai et al. (2024) were unable to identify the essays generated by AI. In other studies, instructor detection power was modest. For example, 43 ESL language instructors were able to identify argumentative AI-generated essays from argumentative human essays within an accuracy of only 61.6% (Liu et al., 2023). Uchendu (2021) also found that their human reviewers were able to discriminate various political news media such as CNN and 10 large language models only slightly more than random chance level (50%).

Clark et al. (2021) assessed non-expert human evaluations of GPT-2 and GP23 generated texts in three domains, namely stories, news and recipes. Their findings indicated that humans' evaluations were at random chance level. Subsequently, upon offering humans some assessment support i.e. detailed instructions, annotated examples for human and machine texts or paired comparisons of human and machine-generated texts, human evaluations improved only slightly.

Finally, Liu et al. (2023) found that their 43 ESL language instructors were able to identify argumentative Gen-AI essays from argumentative human essays within an accuracy of only 61.6%, but that the accuracy rose only to 67.7% after minimal self-training in AI linguistic features and their reflections on them. Therefore, another opportunity is to support teachers' professional judgment and expertise in assessment. For example, Meishar-Tal (2024) argues that the key to plagiarism identification is "to identify gaps between the teacher's expectations concerning students' linguistic proficiency and knowledge and the output generated by the chatbot" (p. 708).

2.6. Previous Studies on Gen-AI Writing Style

Another line of research focused on describing Gen-AI writing style (AlAfnan & MohdZuki, 2023; Desaire et al.; 2023; Liu et al.; 2023), one of whose goals is to support teachers for teaching and formative assessment purposes. This research compared human-authored texts and machine-generated texts in order to discern their linguistic lexico-grammatical and text-level features.

Regarding the discourse macro-level characterization, machine-generated texts often provided terminology and abbreviations without any definitions or explanations (AlAfnan & MohdZuki, 2023) and exhibited many similar examples and many repetitive expressions (Liu et al., 2023). Meishar-Tal (2024) also characterised machine-generated texts by "the use of concepts not covered within the curriculum or through the presentation of illustrations conflicting with the local cultural context" (p.4). Almanea (2024) participants associated ChatGPT outputs with a general writing style, characterized by being overly objective and impersonal. The content consisted of fabricated and sparse information, often lacking in the use of academic references. Further, sentences were found to be exceptionally lengthy and syntactically complex, and ideas lacking in coherence.

Considering the lexico-grammatical features, human-authored texts were characterized by the presence of spelling and grammar errors and the reference to personal experience (Liu et al., 2023). Desaire et al. (2023) also noted that scientific writers used more subordinators and coordinators, and used more numbers and proper nouns and acronyms, while ChatGPT used general words when referring to names of scientists. By contrast, vocabulary in machine-generated texts consisted of low-frequency words, which are often not aligned with the learners' language proficiency (Almanea, 2024).

2.7. Research Gap and Research Questions

Faculty in higher education have been so far portrayed as incapable of detecting Gen-AI academic plagiarism. Barrot (2023), for example, states this, highlighting that some "teachers may have difficulty distinguishing between the students' work and the text generated by ChatGPT" (p. 4). Whilst this may be partially true, especially so given the multiple latent opportunities at the level of teacher, learner, methods of plagiarism and detection reviewed above, this paper argues that when given the opportunity teachers are able to discern their own characterizations of texts and identify Gen-AI plagiarism when research is grounded in their own instructional contexts building on their own situated knowledge (Lave & Wenger, 1991).

Faculty can draw on their professional assessment knowledge of genres in their associated disciplines and also identify their characteristic lexico-grammatical features (Rogerson & Basanta, 2016), which is essential in detecting and curtailing plagiarized work (Bretag & Mahmud, 2009). This presents a unique research opportunity, which this study aims to seize, capitalizing on the link between "the teacher's expectations concerning students' linguistic proficiency knowledge" and "the output generated by the chatbot" (Meishar-Tal, 2024, p. 708) and contextualizing the findings from the perspective of the opportunity construct of traditional academic fraud theory. To do so, the study addresses the following research questions:

RQ1: What strategies do academic faculty teaching in (applied) linguistics and literature courses in a tertiary setting in the Sultanate of Oman perceive their undergraduate students to employ in order to camouflage their unpermitted use of ChatGPT in externally written assignments?
RQ2: What linguistic indicators give away ChatGPT-assisted plagiarism in the context of student-authored literature reviews?

Based on the wide-ranging prevalence of LLMs and thus the in-distinguishability and humanness of their outputs (Uchendu, 2021), the study foregrounds a human perspective to addressing academic plagiarism in written assessments, thus advancing an emic perspective in the conduct of the study. A human perspective is timely in the absence of any reliable anti-Gen-AI plagiarism technology, and it is also congruent with the increasing emphasis between teaching, learning and assessment.

3. Methodology

The study adopts a mixed-methods methodology, where the focus is on complementarity (Johnson & Onwuegbuzie, 2004). This entailed that insights gained from one method were sought to support the insights gained from the other method. It specifically employed qualitative data in the form of observations of experienced faculty of their students' Gen-AI written assignments, notably ChatGPT 3.5, using a custom-made questionnaire including a number of free response questions. This was complemented by textual data in the form of a sample of students' written literature reviews (LRs) submitted to an academic writing course, where students acknowledged they granted permission to the instructor to analyse their written work for this research in the condition of anonymity. The data were collected concurrently and in parallel such that instructors' observations and the LRs were collected together and then were analysed separately in order to offer a more integrated picture of Gen-AI plagiarism.

The questionnaire, which was content validated by three panellists in applied linguistics and further clarified and developed, sought to focus on the faculty's observations of the linguistic and behavioral strategies which their students employed to camouflage their use of ChatGPT. It consisted of three sections: Section A collected background information on the target faculty including their teaching experience, area of specialization, and gender. Section B included a question on the non-linguistic strategies which faculty observed their students to use during external assessments.

Finally, Section C elicited the linguistic strategies which the faculty observed their students to incorporate in the write-up of the different assignments. Additionally, in the hope of empirically identifying "what students do" and "not [just] what [teachers] say they do" (Walker, 2010, p. 41), the analysis of the LRs sought to focus on the linguistic clues which undergraduate students employed, and which have inadvertently given away their unfair use of ChatGPT.

3.1. Data Sources and Sampling Methods

This study draws on instructors' observations and students' written artefacts as key data sources to address students' unpermitted use of Gen-AI chatbots in written assessments. Thirteen of the 27 faculty teaching in a Department of English and Translation at a university in Oman were surveyed at the end of Spring 2024, approximately two years after the emergence of ChatGPT 3.5. Purposive sampling was adopted, surveying instructors teaching in courses in which written assessments figured prominently. This produced a relatively small sample size because instructors teaching in other non-writing heavy courses (e.g. translation, language skills and theoretical linguistics) were excluded from the study.

The learner data in the form of 13 literature reviews (LRs) were acquired from students enrolled in an academic writing course during Fall 2023 and Spring 2024. Each LR was approximately 2500 words, making it difficult to analyse a larger sample. Additionally, these LRs were purposively sampled i.e. selecting only those demonstrating evinced use of Gen-AI based on the students' own teacher professional judgment. The decision was also based on previously written in-classroom essay assessments, which were attempted by the students inside the classroom, and acted as baseline data, benchmarked as generally indicative of these students' true writing ability, unassisted by Gen-AI.

The author taught the course several times before the introduction of AI in the setting, thus affording him informed insider and comparative perspective into the expected writing performance of the current and past students over the years, and at the same time allowing him heightened sensitivity to any traces of unpermitted

AI use, barring any use of AI-detection software in this study, which as reviewed earlier, is not effective (Pudasaini et al., 2024).

In essence, the use of data outsourced from the questionnaire and the LRs afforded two distinct types of insights. First, the faculty's observations regarding their students' recourse to Gen-AI chatbots in the written assessments were macroscopic in nature, focusing on instructors' perceived views of their students' use of Gen-AI across different writing genres such as literary reviews, essays, term papers, project papers, etc. This meant that faculty observations were not genre-bound but were general in nature. Second, the analysis of the student-authored LRs were microscopically focused, thus affording the study the much needed zero in on the lexico-grammatical properties of this particular genre. The interlocking nature of these two types of data allowed the complementary exploration of both perceptions and actual writing practices in relation to unpermitted Gen-AI use in written assessments.

3.2. Data Collection Procedure

A questionnaire was administered in May 2024 to faculty teaching in writing-heavy courses. By that time, targeted faculty and their students had already had two semesters of Gen-AI chatbots' experience. The literature reviews (LRs) were also selected from Fall 2023 semester (at the end of December) and from Spring 2024 (June 2024). Situated in the early days of Gen-AI, the study was necessarily exploratory. However, at the time, there was a heightened interest in sharing insights about this new phenomenon, and so rich data were collected in a relatively short-time span while the faculty's observations of their students' written work were still fresh in their minds having completed the final-term written assessments.

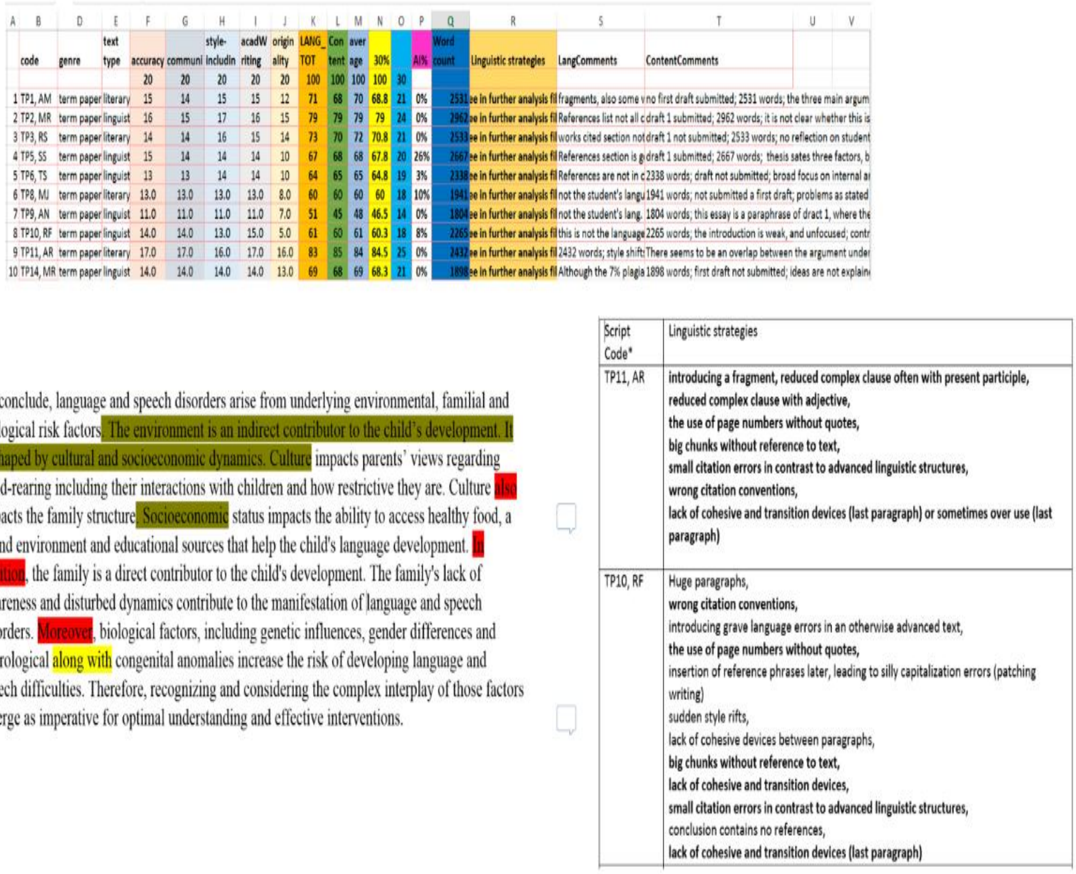
For the sample LRs, the researcher had access to these by virtue of them being a part of the course assessments. As a genre, the LRs were synthesized academic reports of lengths in the range of 2000 words on one academic research topic such as second language acquisition or speech disorders. Students typically produced these reviews outside of class based on eight academic sources (e.g., journal articles and book chapters). Their write-up was supervised by the course instructor in feedback-looped sessions during their actual classes.

3.3. Data Analysis

The questionnaire data were analysed in November 2024, and the literature reviews (LRs) analyzed in two periods (at the end of December 2023 and June 2024), corresponding to the semester from which they were pooled. All the data were entered into NVivo 12 and analyzed with respect to the research questions. The analysis of the LRs was attempted first and were completed in three interim stages (Figure 1). First, they were initially marked using the course analytic assessment matrix comprising content, organization, task achievement and language. This was a standard coursework procedure, whose findings fed naturally into this study. At this stage, special attention was also paid to the similarity and AI indices provided by Turnitin.

Second, each paper was read again, colour-coding the textual segments demonstrating evinced unpermitted use of Gen-AI. Students' performance in the previous essay assessments afforded a comparative angle in the analysis (Miles & Huberman, 1994), thereby highlighting 'deviant' or atypical student writing. This was performed to ensure that segments marked as candidates for Gen-AI use were in fact representative, since they were grounded in relation to the performance baseline data, stressing the situated nature of the study. Additionally, the LRs were coded by means of the commenting function in Microsoft Word.

Finally, qualitative notes denoting the salient patterns emerging from textual highlights and commenting were recorded for each LR. The subjectivity inherent in the analysis of the LRs conducted solely by the researcher may be an issue; however, it is argued that the study's adoption of an emic insider perspective (Mercer, 2007) afforded by the instructor's familiarity with these students especially vis-à-vis their performance and writing style in the in-class essays was instrumental, as it afforded the analysis the depth and richness required for such a design. Additionally, to mitigate researcher bias, a subset number of LRs (N = 5) were independently analysed by the researcher at two different time intervals, and the analysis showed consistent patterns. At the end, the LRs were anonymized (LR1, LR2, etc.), and were transported onto NVivo 12 for more systematic coding.



A	B	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V
	code	genre	text type	accuracy	communi	style-includin	acadWriting	originality	LANG_TOT	Content	average	30%		AI%	Word count	Linguistic strategies	LangComments	ContentComments		
				20	20	20	20	20	100	100	100	100	30							
1	TP1, AM	term paper	literary	15	14	15	15	12	71	68	70	68.8	21	0%	2531	ee in further analysis fil	fragments, also some v	no first draft submitted; 2531 words; the three main argum		
2	TP2, MR	term paper	linguist	16	15	17	16	15	79	79	79	79	24	0%	2962	ee in further analysis fil	References list not all c	draft 1 submitted; 2962 words; it is not clear whether this is		
3	TP3, RS	term paper	literary	14	14	16	15	14	73	70	72	70.8	21	0%	2533	ee in further analysis fil	works cited section not	draft 1 not submitted; 2533 words; no reflection on student		
4	TP5, SS	term paper	linguist	15	14	14	14	10	67	68	68	67.8	20	26%	2667	ee in further analysis fil	References section is g	draft 1 submitted; 2667 words; thesis sates three factors, b		
5	TP6, TS	term paper	linguist	13	13	14	14	10	64	65	65	64.8	19	3%	2338	ee in further analysis fil	References are not in c	2338 words; draft not submitted; broad focus on internal a		
6	TP8, MJ	term paper	literary	13.0	13.0	13.0	13.0	8.0	60	60	60	60	18	10%	1941	ee in further analysis fil	not the student's langu	1941 words; not submitted a first draft; problems as stated		
7	TP9, AN	term paper	linguist	11.0	11.0	11.0	11.0	7.0	51	45	48	46.5	14	0%	1804	ee in further analysis fil	not the student's lang.	1804 words; this essay is a paraphrase of dract 1, where the		
8	TP10, RF	term paper	linguist	14.0	14.0	13.0	15.0	5.0	61	60	61	60.3	18	8%	2265	ee in further analysis fil	this is not the language	2265 words; the introduction is weak, and unfocused; contr		
9	TP11, AR	term paper	literary	17.0	17.0	16.0	17.0	16.0	83	85	84	84.5	25	0%	2432	ee in further analysis fil	2432 words; style shift:	There seems to be an overlap between the argument under		
10	TP14, MR	term paper	linguist	14.0	14.0	14.0	14.0	13.0	69	68	69	68.3	21	0%	1898	ee in further analysis fil	Although the 7% plagia	1898 words; first draft not submitted; ideas are not explain		

To conclude, language and speech disorders arise from underlying environmental, familial and biological risk factors. The environment is an indirect contributor to the child's development. It is shaped by cultural and socioeconomic dynamics. Culture impacts parents' views regarding child-rearing including their interactions with children and how restrictive they are. Culture also impacts the family structure. Socioeconomic status impacts the ability to access healthy food, a sound environment and educational sources that help the child's language development. In addition, the family is a direct contributor to the child's development. The family's lack of awareness and disturbed dynamics contribute to the manifestation of language and speech disorders. Moreover, biological factors, including genetic influences, gender differences and neurological along with congenital anomalies increase the risk of developing language and speech difficulties. Therefore, recognizing and considering the complex interplay of those factors emerge as imperative for optimal understanding and effective interventions.

Script Code*	Linguistic strategies
TP11, AR	introducing a fragment, reduced complex clause often with present participle, reduced complex clause with adjective, the use of page numbers without quotes, big chunks without reference to text, small citation errors in contrast to advanced linguistic structures, wrong citation conventions, lack of cohesive and transition devices (last paragraph) or sometimes over use (last paragraph)
TP10, RF	Huge paragraphs, wrong citation conventions, introducing grave language errors in an otherwise advanced text, the use of page numbers without quotes, insertion of reference phrases later, leading to silly capitalization errors (patching writing) sudden style rifts, lack of cohesive devices between paragraphs, big chunks without reference to text, lack of cohesive and transition devices, small citation errors in contrast to advanced linguistic structures, conclusion contains no references, lack of cohesive and transition devices (last paragraph)

Figure 1: Stages of data analysis of the students' literature reviews

During the time, the questionnaire data revealing the experienced faculty observations were subjected to the same rigorous analysis in NVivo. Faculty responses were read several times before they were coded. Single codes were initially created, but were subsequently converted into overarching categories, subsuming similar singular units of meaning. For instance, since the single codes initially emerging from the faculty responses pointed to language tokens such as misspellings, wrong word forms, capitalization errors, improved language style across two drafts, etc., this necessitated that those were grouped into meaningful codes (tree nodes), denoting for example code groups such as language style rifts, redundancy, language leap, etc.

4. Results

Following the iterative analysis, the results are presented here in two sub-sections: questionnaire data, principally related to the observations by faculty members regarding how their students camouflaged their unpermitted use of Gen-AI in externally-written assignments (Research Question 1), and learner data, which produced lexico-grammatical indicators, giving away students' unpermitted use of ChatGPT in their written literature reviews (Research Question 2).

4.1. Research Question 1: Questionnaire Data

During the analysis, the questionnaire data were interrogated for any indicators of unpermitted use of ChatGPT in undergraduate students' written assignments in general, and the findings pointed to the presence of three clues: linguistic, content-related and behavioural.

4.1.1 Linguistic indicators

The language-level markers indicating unpermitted ChatGPT use appeared in four sub-types, related to elegant style, language style rift, redundancy in expression, and language leap. First, the style elegance linguistic markers indicated undergraduate students' submitting prose which was elegant, error-free, marked by the use of esoteric academic jargon, unusual idiomatic expressions and complex syntactic noun phrases, unrepresentative of the students' linguistic ability.

Second, in the style rift linguistic markers, the language drops to a substandard level, showing free variation in writing style, intermittently oscillating between the use of highly sophisticated vocabulary and complex sentences on the one hand and simplistic and informal style on the other, and irregularly transitioning from perfect grammar to occasional odd language errors, suggestive of texts that schizophrenically combine both Gen-AI lexico-grammatical features with students' own EFL writing level, sometimes punctuated by students' possible deliberate insertion of wrong structures or phrases (see under Table 2 for an illustration of these problems from the textual perspective).

Third, the redundancy in expression linguistic markers involved undergraduate students' supplying prose characterized by repetitiveness. This saw writing go around in circles essentially saying very little, moving to the use of repetitive linguistic structures, and the almost identical paragraph structure. Finally, at the level of language leap linguistic markers, faculty perceived the prose to be of a higher level than that of students' previously-submitted written work, principally showing a sudden disappearance of a particular language problem, and the sudden shift from error-free writing to student-level writing across different submissions, usually with vast differences between an imperfect first draft and a final draft showing almost perfect editing. For example, a student's first draft contained structural inaccuracies in the form of fragments, comma splices or run-on sentences, and the final draft was impeccable, showing no evidence of these even though the instructor's initial feedback did not address them.

4.1.2 Content-related indicators

The content-related markers (Table 1) emerged in three sub-types revealing prose characterized by shallow content lacking in substance and marked by high redundancy (Group A), content misaligned to the subject-matter delivered during lectures (Group B), and content irrelevant to the requirements of the assignments (Group C).

Table 1: Content-related markers betraying students' use as perceived by faculty

Shallow Content	**Group A**	
	●	General Writing
	●	Too Generic
	●	Generic Reviews
	●	Generic Sentences
	●	Use of Generic Terms

Content Irrelevant to Class	**Group B**	
	●	Literary Text Not Discussed in Any Specific or Meaningful Way
	●	Quotations Purportedly from Text Are Not
	●	Well-Written Assignment Lacking Required Content Discussed in-Class

Content Irrelevant to Assignment	**Group C**	
	●	Not Following Assignment Instructions
	●	Ignoring Teacher Class Instructions (How To Write RP Proposal, Not Answering the Questions)
	●	Not Sticking to Instructions Given In-Class
	●	Submitting Text Vaguely Similar to the Question (Without Specific Details, Evidence or Examples)

This meant that the content making up the students' assignments was generic, superficial and/or irrelevant, indicating a level of disengagement form the learning outcomes students are expected to achieve. This comes at a time when the language quality employed in these assignments of a very high standard.

4.1.3 Behavioral indicators

Beside linguistic and content markers, the questionnaire data were equally interrogated for other non-linguistic indicators of unpermitted use of ChatGPT in their written assignments in general. The qualitative data revealed faculty perceptions of behavioral strategies their students employed took different personas, which can be grouped as them acting out in five ways to conceal this act: acting deaf (3 references from the data), acting outsmart (9 references), acting dumb (3 references), acting innocent (3 references) and playing the sympathy game (3 references).

In relation to *acting deaf*, one instructor reported the following:

> *"I could easily tell those who were dependent on ChatGPT because they would not listen. All my instructions including instructions on how to format the paper went into one ear and out of the other ear"* (I-13)

Another instructor reported that students kept

> *"insisting on working on a specific topic without accepting any modifications in the nature of topic to be researched"* (I-7).

In relation to *acting dumb*, one instructor stated:

> *"I would say one of these ways is when they admit using ChatGPT but they hide the extent to which they used it. They always claim to have used it to paraphrase only when in fact they use it for much more than that."* (I-13)

Another reported that a student

> *"tried to explain that Turnitin and additional on-line detection software identified large parts of her writing as AI-generated only because she used Grammarly to check the grammar"* (I-5).

In relation to *acting innocent*, one instructor reported that

> *"One student stated that she felt hurt I questioned her academic integrity"* (I-5).

Additionally, one instructor stated that

> *"One student claimed that her perfect writing was a result of seeking help from one of her friends who provided her with valuable feedback"* (I-8).

In relation to *acting outsmart*, one instructor reported the following:

> *"Providing outlines for extended writing which suggest that the student is working individually but the outline has itself been AI generated, with a few additions."* (I-6)

Additionally, another instructor stated the following:

> *"Requesting feedback: in one instance, a student asked me several times for feedback on the same assignment to give the impression that he worked on the assignment by himself."* (I-8).

Finally, in relation to *playing the sympathy game*, one instructor stated the following:

> *"Others emphasize that they need a 2.3 GPA for the practicum. So they basically tell the instructors that any low grade will lead them to be prevented to take the practicum in an attempt to create a feeling of guilt in the instructors."*

Another also stated that

> *"Others use their long-term or chronic illnesses or family trouble as ways to gain sympathy in case plagiarism or use of AI was detected."* (I-1).

4.2. Research Question 2: Learner Data: Linguistic Indicators

The learner data generated from the textual analysis of the students' literature reviews (LRs) were analyzed in relation to RQ2: What linguistic indicators give away ChatGPT-assisted plagiarism in the context of the student-authored LRs?

To address this question, the learner data in the form of LRs (N = 15) were interrogated in terms of their lexico-grammatical features initially and then subsequently in terms of how they could have displayed ChatGPT-assisted plagiarism, always remembering that this was facilitated in reference to the baseline essays written by students in the classroom under examination conditions. The textual analysis of the LRs pointed to three basic indicators betraying students' unpermitted use of Gen-AI in their written work, namely in-text citation markers, language style markers and discourse markers.

In relation to in-text citation (N = 44), 10 out of the 15 LRs displayed extended texts (normally in the range of one to multiple paragraphs) barring the use of in-text citation or attribution. Also, 10 out of the 15 LRs displayed highly sophisticated texts in terms of language style, but with foundational errors in in-text citation (e.g., missing the year in APA, and/or writing the page numbers in the wrong format). It was also typical in 6 of the LRs to have conclusions, without sources. The same number of LRs displayed in-text citations, lacking direct quotations but with page numbers nevertheless included. Four LRs showed in-text citations were added *post-facto*, often with the result of deliberately introducing capitalization errors. There were also 3 LRs, showing paraphrases of complete sources in some parts.

Finally, there were references to sources including the name of their authors and excluding page numbers. Finally, there was one LR where page numbers were faked. All in all, these incompatibilities in conventional in-text citation was a major giveaway for the LR genre (See Table 2).

Table 2: Linguistic indicators of ChatGPT-assisted plagiarism in the literature reviews

In-text Citation Errors (N = 44)		
	●	big chunks of text excluding academic reference
	●	basic citation in otherwise advanced language structures
	●	citation (including page numbers, but excluding quotes)
	●	conclusions without citation
	●	insertion of reference phrases later, leading to silly capitalization errors
	●	a series paraphrase of the references
	●	references to sources (excluding page numbers and quotes)
	●	quoted texts excluding page numbers
	●	fake page numbers

In relation to language style rifts (N = 41), the LRs showed the following characteristics: the use of sophisticated participial present (N = 15) and participial past (N = 15) verb phrases (atypical of students' writing ability as judged from their previous assignments and overall language ability). Additionally, there was also the tendency for texts to exhibit obviously grave language errors in an otherwise advanced text.

To illustrate these two linguistic tendencies, representative unedited excerpts from LR10 and LR6 are provided below, with the target language forms underlined for present participial verb phrases, **bolded** for past participial verb phrases, and double-underlined for grave errors (i.e. silly capitalization errors):

> *"Lack of access to textbooks, educational opportunities, and quality childcare services are the main causes of speech and language delay related to poverty. … Academic achievement also reflects social and economic differences; Children with speech and language disorders are more likely to repeat a year of school … If a child feels that he is different from the rest of his classmates …, this makes him feel frustrated, affecting his educational level. According to the study conducted by Lara-Díaz et al. (2021) …, the type of the school, private or public, detected no statistically significant differences, indicating that maternal education and family support were more important factors."* (LR10)

> *"As discussed above, linguistic fossilization is "the long-term persistence of plateaus of non-target- like structures in the interlanguage of non-native speakers" (Fidler, 2006, p.399). a complicated phenomenon that is impacted by a wide range of internal and external factors. These factors interact in complex ways, influencing how people acquire languages and playing a vital role in language competency plateaus or stagnation. The complex interactions … highlight the nuances and difficulties* involved *in overcoming language fossilization."* (LR06)

In relation to text-level features (N = 23), the clearest marker was the size of the paragraphs, whereby 12 of the 15 LRs showed proportionately oversized paragraphs for the genre of LR. Other linguistic markers were either the lack (or sometimes the overuse) of cohesive devices or uniformly writing a concluding sentence for each paragraph. In relation to the oversized paragraphs in the sampled LRs, the researcher performed a stratified-random cross-validation of four LRs written by similar level students receiving similar grades (A; A-; D; D) before and after the introduction of Gen-AI in the context (Figure 2).

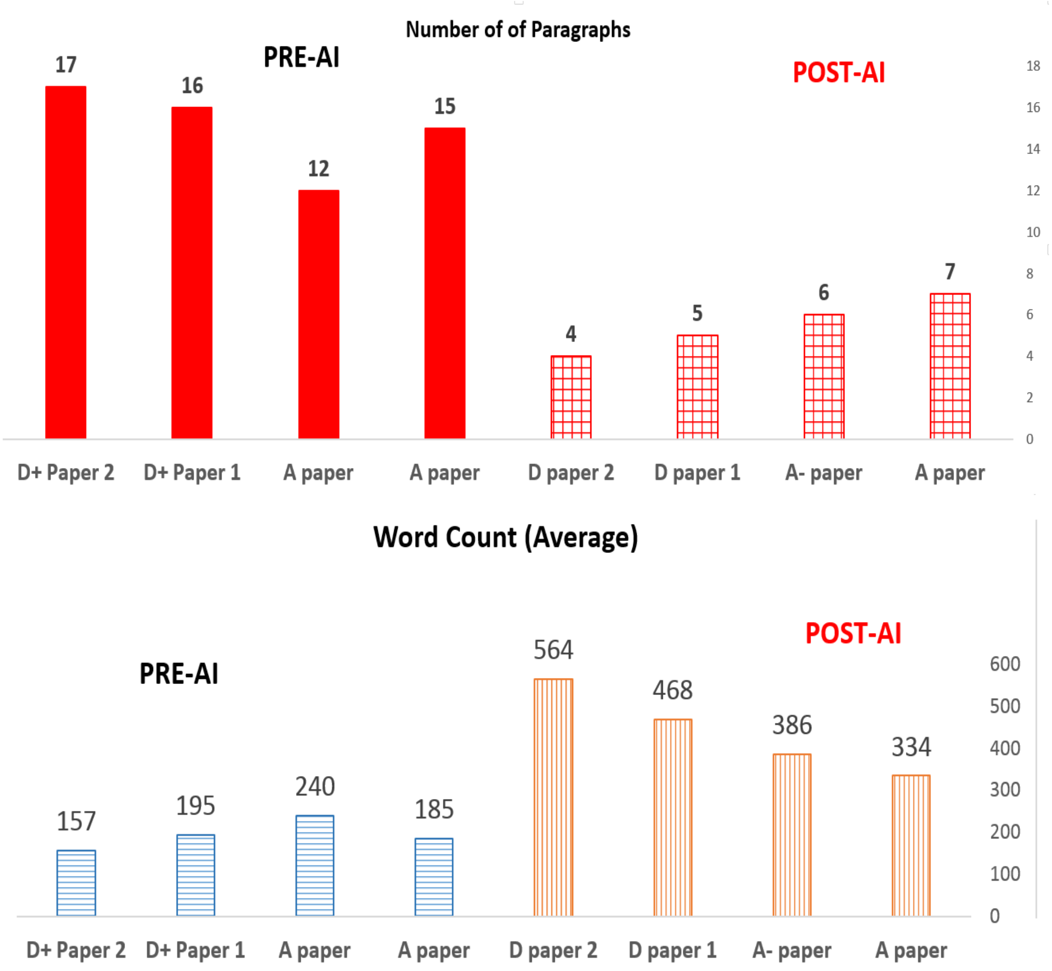


Figure 2: Cross-validation study of paragraphs size before and after the introduction of Gen-AI

Figure 2 shows remarkable differences. First, in the upper chart, before the introduction of Gen-AI, the number of paragraphs in the sampled LRs ranged from 12 to 17 in contrast to only 4 to 7 in the sample LRs which were written after the introduction of Gen-AI. It is important to note that the difference is so remarkable that there is no overlap in the two ranges.

Likewise, the average words per paragraph (represented by the lower chart) ranged from 157 to 240 in LRs written by students before the introduction of Gen-AI, compared to 334 words and 564 words in the LRs written after the introduction of Gen-AI, again denoting the vast difference between the two periods. This suggests that the discoursal feature of oversized paragraphs in the students' sampled LRs was not a chance level error of sample selection but may suggest a marked impact of Gen-AI technologies on writing quantity. In summary, the distribution of these linguistic markers in the LRs is presented in Table 3 below.

Table 3: Distribution of the three linguistic markers in the 15 LRs

Descriptive Statistics	**In-text Citation Markers**	**Language Style Rift Markers**	**Discourse Level Markers**	**Total**
Count	44	41	23	108
Average	2.93	2.73	1.53	
%	40.74%	37.96%	21.29	

Quantitatively and in descending order, the largest linguistic markers giving away students' unpermitted integration of ChatGPT in terms of the LRs were: in-text citation (N = 44, M = 2.93, and 40.74%) closely followed by language style rift (N = 41, M = 2.73, and 38%), and finally discourse-related (N = 23, M = 1.53, and 21.30%), suggesting the primary discoursal feature of in-text citation of the genre of the literature review was a major giveaway. These figures are substantial, showing that Gen-AI plagiarism is not a rhetorical manoeuvre, but an omnipotent reality.

5. Discussion

The study assessed the applicability of the opportunity construct derived from academic fraud theory (Burke & Sanney, 2018; Cressey, 1953) as a theoretical framework to understand and explain Gen-AI plagiarism. It attempted to do this through assessing whether faculty situated in a bound instructional context were able to identify a modern form of academic plagiarism, namely ChatGPT-assisted plagiarism in undergraduate written assessment, drawing on the expectations they have of their students' writing ability and the knowledge they possess of large language models' (LLMs) output capabilities (Meishar-Tal, 2024). It was hypothesized that with the introduction of LLMs such as ChatGPT, wide-ranging opportunities readily presented themselves to the students, thereby facilitating unpermitted use of LLMs in their written work.

The research questions were informed by means of questionnaire and textual data to address the clues which gave away undergraduate students' ChatGPT-assisted plagiarism in written assessment. Thirteen faculty in an English and Translation Department were surveyed by means of an open response questionnaire, and 15 externally authored student literature reviews (LRs) were analyzed in search for these clues. In effect, the study offered an empirical test of the potency of academic fraud theory to explain academic plagiarism in the new Gen-AI era. The findings lend support to the appropriateness of the opportunity construct derived from academic fraud theory as the best fit for the data and offer a preliminary model (Figure 2) predicting the interrelationship between the components leading up to Gen-AI plagiarism.

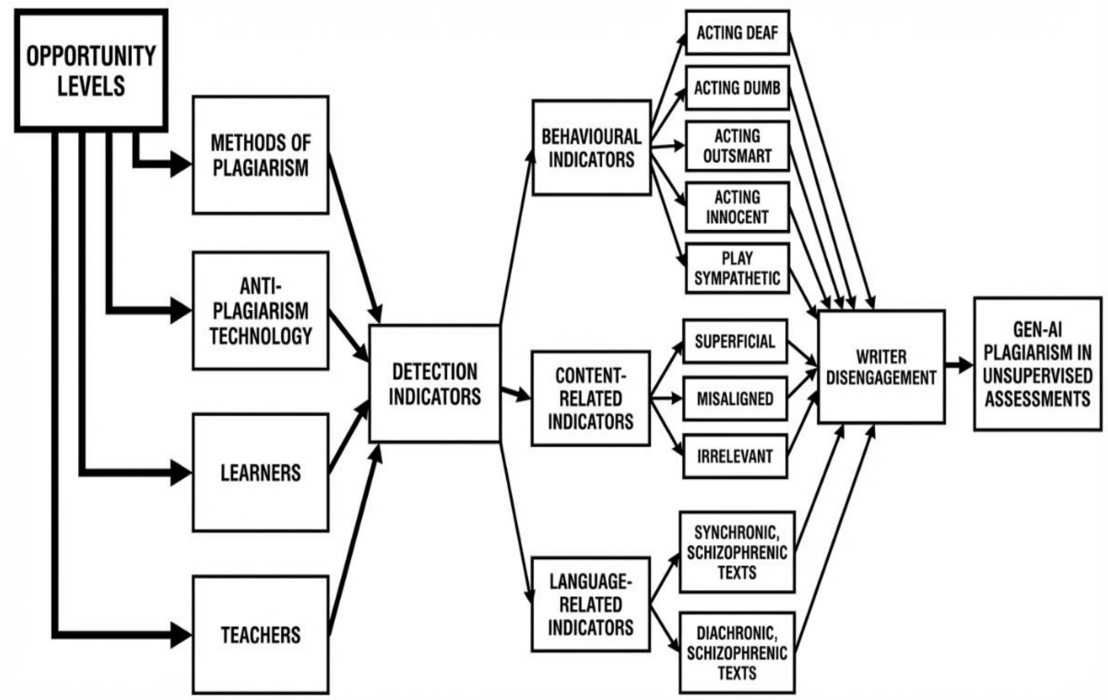


Figure 2: Proposed model of Gen-AI plagiarism

As shown in Figure 2, the results indicated that when they were left to their own devices with the Gen-AI bathtub some undergraduate students did not hesitate to commit plagiarism, benefiting from an existing opportunity manifest in the availability of a gap manifested at various levels related to teacher, learner, methods of plagiarism and anti-plagiarism technology. This was also facilitated by these students' recognition that the unsupervised use of Gen-AI for writing purposes is disrupting the long-held superiority of faculty and technology in the united detection of plagiarism, in the absence of any hard or concrete evidence in the form of originality index arbitration for its deterrence. Similarly, the findings indicated that not only were faculty able to identify plagiarism, but they were also successful in locating its traces at three levels: behavioural, content-related and language-related, therefore narrowing one of the opportunity gaps at their level.

In terms of behavioural clues, faculty reported students employed cunny and uncanny approaches, ranging from admittance of plagiarism to pretence to denial and complete concealment. At one extreme, students hid their plagiarism plans or pre-meditated decisions to commit plagiarism (e.g., not following assignment instructions or adamantly working on a topic they have pre-determined). Even when students admitted to committing plagiarism, which was rare, they invoked guilt in their instructors as a means of gaining sympathy (e.g., grades, graduation, family illness) or underestimated the extent of ChatGPT use (e.g., use claimed to be limited to paraphrasing and proof-reading technology such as Grammarly).

In between were other behaviours in which the students denied plagiarism altogether, feigning innocence (e.g., feeling hurt the instructor questioned their integrity) or rationalizing the disproportionate improvement in their writing (e.g., due to peer assistance). Faculty also reported that students diverted instructors' attention and lured them into the belief that they were working on the assignment themselves (e.g., soliciting frequent feedback and partly submitting AI-based assignments). These behaviours signalled a lack of engagement from the writing process, or writer disengagement, where students did not engage in the task, and exerted little effort in completing it (Hınız & Çelik, 2025).

Regarding content generation, the literature highlights the potential benefits to student engagement that both learner-generated content (Lambert et al., 2017), and Gen-AI content generation for second language writing (Barrot, 2023; Thao et al., 2023). However, this study paints a contrary picture to that positive discourse. The evidence supporting students' uncritically borrowing content from LLMs, content perceived by academics to be superficial, misaligned to course content and irrelevant to assigned academic work, marks a more serious issue, which points to disengagement *from* rather than engagement *into* learning. These illustrate the inadequacies of ChatGPT as a generator of the kind of content which academics purview to be compatible with the content they profess in their courses.

The perceived discrepancy between disciplinary knowledge valued by academics and ChatGPT-generated content marks a curricular disjunction, of which students may not be cognizant. Theoretically, these deficits in content mark ChatGPT as a non-member of the same discourse community as academics, with parallels in the literature, including pragmatic inadequacies such as ChatGPT imbedding terminology without definitions or explanations (AlAfnan & MohdZuki, 2023), providing redundant examples and monotonous expressions (Liu et al., 2023), excluding personal experience (Desaire et al., 2023), using concepts irrelevant to the implemented curriculum or insensitive to the local cultural context (Meishar-Tal, 2024), and providing inaccurate information, lacking in references, and characterized by low content density (Almanea, 2024).

In terms of language-related plagiarism indicators, the findings are in accord with the published literature in highlighting the remarkable linguistic superiority of ChatGPT-generated texts compared to EFL student-authored written assignments. This is not surprising and a number of studies pointed to this already (e.g. AlAfnan & MohdZuki, 2023; Almanea, 2024; Desaire et al., 2023; Liu et al., 2023). However, what this study contributes to this body of research is a more consolidated characterization of ChatGPT-based writing style. It extends this linguistic dimension by documenting authorship in flux, through the appearance of schizophrenic texts, marking clear linguistic clues traceable to ChatGPT writing style. These schizophrenic texts demonstrate a leap evident in the language quality of the student's written work which can vary either synchronically (i.e. based on the instructor's initial assessment) or diachronically (over time).

Students producing synchronic, schizophrenic texts concoct hybrid texts whose defining lexico-grammatical features exhibit a linguistic oscillation between a vernacular style characteristic of EFL student writing on the one hand and a native LLMs' style on the other (vertical ChatGPT-assisted plagiarism). Conversely, students concoct diachronic, schizophrenic texts which are hybrid texts whose defining lexico-grammatical features exhibit a linguistic oscillation between a vernacular style characteristic of EFL student writing on one occasion (initial draft), and a native LLM style on another (subsequent drafts), denoting linguistic leaps brought about as an outcome of different submissions varying temporally (horizontal ChatGPT-assisted plagiarism).

Additionally, the use of mixed-methods offered a robust approach for the exploration and identification of Gen-AI plagiarism in unsupervised assessments. The findings indicated that both the perceptual and textual data complemented one another in the identification of key lexico-grammatical features indicative of ChatGPT-assisted plagiarism. Especially, the textual analysis of the genre of LR pointed to the presence of one genre-specific, linguistic indicator exposing ChatGPT-assisted plagiarism. This appeared in the form of in-text citation, an obligatory lexico-grammatical feature essential to the genre of LR with its discoursal function of surveying disciplinary theories and previous studies (Cheung, 2012; Davis, 2013). This finding suggests that unmarked lexico-grammatical features of a specific genre are more vulnerable to Gen-AI plagiarism, and paradoxically are the ones which pose a major giveaway exposing it. This also suggests that salient and obligatory lexico-grammatical features characterizing genres other than the literature review may be more predisposed to Gen-AI plagiarism, thus offering implications for generalizability.

To sum up, writing or writer disengagement, operationalized through the manner students short-circuited from language, content and behaviour engagement, is a major strong explanatory power indicator exposing ChatGPT-assisted plagiarism in external written assessments. This means that to engage in ChatGPT-assisted plagiarism is to disengage from learning. Both a student engagement (Hınız & Çelik, 2025) and a pragmatic perspective to academic plagiarism (McIntire et al., 2024) attach primary significance to deep learning, via creative struggle. Learning involves experimenting with ideas, and exploring subjects, leading to student engagement, which "has the capacity to build curiosity, a love for the subject, and a love for learning" (p. 3). However, when students' efforts are overfocused on outcomes through resorting to plagiarism at the expense of process through engagement with writing, the natural outcome is learner disengagement, which is inversely proportional to learning.

In second language writing assessment, student writers' not putting their own language to use, their overreliance on content not generated by themselves, but by ChatGPT, and their behaviour in pulling away from the writing task are salient markers of writing or writer disengagement, thus undermining the learning project in unsupervised use of LLMs, especially in content-based programmes where both content and language are intertwined. This marks writing activities in a learning and teaching context as sites of pull and push forces, representing opportunities for or against writer engagement. Through engaging in ChatGPT-assisted plagiarism, student writers pull out of, rather than push into, these opportunities, effectively disengaging from the writing process, and in so doing miss on, rather than seize, opportunities for language and content learning and development.

Being explorative in nature, the study is limited in two ways. First, it is possible for the findings to be undermined by the growing developments in LLMs, thus rendering a level of obsolescence to some. As LLMs are continuously trained on human data, the gap between machine-generated outputs and human-generated texts will reduce, weakening detectability from human assessors and AI-based anti-plagiarism technology. Related to this is the issue of generalizability as the result of the limited sample size of the LR assessment artefacts and the questionnaire respondents.

Because the detectability of ChatGPT-based plagiarism rests on the human assessors' knowledge of their students' writing ability, it is always context-dependent, making it difficult to extend the findings wholesale to other contexts. In spite of these limitations, however, this study's unique contribution to the anti-plagiarism literature is powerful in identifying the applicability of the concept of opportunity in traditional academic fraud theory in an era increasingly characterised by fervent use of Gen-AI, and in the provision of a model explicating aspects of Gen-AI academic plagiarism.

6. Conclusions and Implications

This mixed-methods study was able to offer a robust model of Gen-AI academic plagiarism and portray writer disengagement as a complex phenomenon comprising behavioural, content-related and language-related indicators in externally written assessments. The study draws two key implications regarding the nature of ChatGPT-assisted plagiarism in unsupervised written assessments. From a practical measurement perspective focused on the validity of what we assess as faculty, ChatGPT may be reimagined such that its use in second language writing assessment "would not be possible," "would not be effective" or "is a necessary component" (Playfoot et al., 2024, p. 9). To develop assessments that foster writer engagement, our feedback practices should target the limitations identified by this study at the behavioural, content and language levels in order to turn these limitations into opportunities through the adoption of a process-based assessment, building feedback in every loop of the assessment process.

From a theoretical learning perspective highlighting the formative added value of what we assess to student learning, a pragmatic view of academic plagiarism is essential. This stance acknowledges that because at this level students already have a firm understanding of what is ethically or morally wrong regarding academic plagiarism, students should view academic plagiarism as an adversary to learning, understood as the outcome of academic struggle and task engagement, and that for students to plagiarize means to cheat themselves "out of the very purpose for which they came to university in the first place" (McIntire et al., 2024, p. 19). Because of undergraduate students' overreliance on ChatGPT as a generator of content and the ever-increasing sophistication of LLMs, future studies could explore in-depth other forms of writer disengagement. Theoretically, the other two concepts making up academic fraud theory, namely pressure and rationalization, could benefit from further research in a world characterised by increasing use of ChatGPT.

7. Conflict of Interest

The author declares no conflict of interest.

8. Acknowledgements

The author declares that no form of Artificial Intelligence has been used in the authorship of this paper. The visual model in the Discussion section was produced using Oreate AI at https://www.oreateai.com/
The author would like to thank the reviewers of this paper who have offered extensive feedback to improve it.

9. References

AlAfnan, M. A., & MohdZuki, S. F. (2023). Do artificial intelligence chatbots have a writing style? An investigation into the stylistic features of ChatGPT-4. *Journal of Artificial intelligence and technology*, *3(3)*, 85-94. https://doi.org/10.37965/jait.2023.0267

Almanea, M. (2024). Instructors' and learners' perspectives on using ChatGPT in English as a foreign language course and its effect on academic integrity. *Computer Assisted Language Learning*, 1-26. https://doi.org/10.1080/09588221.2024.2410158

Baron, P. (2024). Are AI detection and plagiarism similarity scores worthwhile in the age of ChatGPT and other Generative AI? *Scholarship of Teaching and Learning in the South (SOTL) in the South*, *8(2)*, 151-179. https://doi.org/10.36615/sotls.v8i2.411

Barrot, J. S. (2023). Using ChatGPT for second language writing: Pitfalls and potentials. *Assessing Writing*, *57*, 100745. https://doi.org/10.1016/j.asw.2023.100745

Bretag, T., & Mahmud, S. (2009). Self-plagiarism or appropriate textual re-use? *Journal of Academic Ethics*, *7*(3), 193-205. https://doi.org/10.1007/s10805-009-9092-1

Burke, D. D., & Sanney, K. J. (2018). Applying the fraud triangle to higher education: Ethical implications. *Journal of Legal Studies in Education*, *35(1)*, 5-43. https://doi.org/10.1111/jlse.12068

Cheung, Y. L. (2012). Understanding the writing of thesis introductions: An exploratory study. *Procedia-Social and Behavioral Sciences*, *46*, 744-749. https://doi.org/10.1016/j.sbspro.2012.05.191

Clark, E., August, T., Serrano, S., Haduong, N., Gururangan, S., & Smith, N. A. (2021). All that's' human'is not gold: Evaluating human evaluation of generated text. In *Proceedings of the 59th Annual Meeting of the Association of Computational Linguistics and the 11th International Joint Conference on Natural Language Processing* (pp. 7282-7296). https://doi.org/10.48550/arXiv.2107.00061

Cotton, D. R., Cotton, P. A., & Shipway, J. R. (2023). Chatting and cheating: Ensuring academic integrity in the era of ChatGPT. *Innovations in Education and Teaching International*, *61(2)*, 228–239. https://doi.org/10.1080/14703297.2023.2190148

Cressey, D. R. (1953). *Other people's money; a study of the social psychology of embezzlement*. Glencoe, Free Press.

Culp Jr, W. C. (2023). Artificial intelligence and ChatGPT: Bane or boon for academic writing. *The Journal of Education in Perioperative Medicine: JEPM*, *25(2)*, 1-3. https://doi.org/10.46374/volxxv_issue2_culp

Davis, M. (2013). The development of source use by international postgraduate students. *Journal of English for Academic Purposes*, *12(2)*, 125-135. https://doi.org/10.1016/j.jeap.2012.11.008

Desaire, H., Chua, A. E., Isom, M., Jarosova, R., & Hua, D. (2023). Distinguishing academic science writing from humans or ChatGPT with over 99% accuracy using off-the-shelf machine learning tools. *Cell Reports Physical Science*, *4(6)*, 1-11. https://doi.org/10.1016/j.xcrp.2023.101426

Fajt, B., & Schiller, E. (2025). ChatGPT in academia: University students' attitudes towards the use of ChatGPT and plagiarism. *Journal of Academic Ethics*, *23*, 1363-1382. https://doi.org/10.1007/s10805-025-09603-5

Galindo-Domínguez, H., Campo, L., Delgado, N., & Sainz de la Maza, M. (2025). Relationship between the use of ChatGPT for academic purposes and plagiarism: the influence of student-related variables on cheating behavior. *Interactive Learning Environments*, *33(6)*, 4047-4061. https://doi.org/10.1080/10494820.2025.2457351

Hınız, G., & Çelik, Ö. (2025). A bibliometric and content analysis of student engagement research in English language teaching. *Language awareness*, *34(1)*, 43-76. https://doi.org/10.1080/09658416.2024.2367970

Jarrah, A. M., Wardat, Y., & Fidalgo, P. (2023). Using ChatGPT in academic writing is (not) a form of plagiarism: What does literature say? *Online Journal of Communication and Media Technologies*, *13(4)*, 1-20. https://doi.org/10.30935/ojcmt/13572

Johnson, B. R., & Onwuegbuzie, A. J. (2004). Mixed methods research: A research paradigm whose time has come. *Educational Researcher, 33(7),* 14-26. https://doi.org/10.3102/0013189X033007014

Khalil, M., & Er, E. (2023, June). Will chatgpt get you caught? Rethinking plagiarism detection. In *International conference on human-computer interaction* (pp. 475-487). Cham: Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-34411-4_32

Lambert, C., Philp, J., & Nakamura, S. (2017). Learner-generated content and engagement in second language task performance. *Language Teaching Research*, *21(6)*, 665-680. https://doi.org/10.1177/1362168816683

Lave, J. & Wenger, E. (1991). *Situated learning: Legitimate peripheral participation*. Cambridge: Cambridge University Press.

Liu, Y., Zhang, Z., Zhang, W., Yue, S., Zhao, X., Cheng, X., ... & Hu, H. (2023). Argugpt: evaluating, understanding and identifying argumentative essays generated by gpt models. *arXiv preprint arXiv:2304.07666*. https://doi.org/10.48550/arXiv.2304.07666

Mahapatra, S. (2024). Impact of ChatGPT on ESL students' academic writing skills: a mixed methods intervention study. *Smart Learning Environments*, *11(1)*, 1-18. https://doi.org/10.1186/s40561-024-00295-9

McIntire, A., Calvert, I., & Ashcraft, J. (2024). Pressure to plagiarize and the choice to cheat: Toward a pragmatic reframing of the ethics of academic integrity. *Education Sciences*, *14(3)*, 1-25. https://doi.org/10.3390/educsci14030244

Meishar-Tal, H. (2024). ChatGPT: The Challenges It Presents for Writing Assignments. *TechTrends*, *68*, 605-710. https://doi.org/10.1007/s11528-024-00972-z

Mennella, T., & Quadros-Mennella, P. (2024). Student Use, Performance and Perceptions of ChatGPT on College Writing Assignments. *Journal of University Teaching and Learning Practice*, *21(1)*, 1-24. https://doi.org/10.53761/pgwk1a93

Mercer, J. (2007). The challenges of insider research in educational institutions: Wielding a double-edged sword and resolving delicate dilemmas. *Oxford review of education*, *33(1)*, 1-17. https://doi.org/10.1080/03054980601094651

Miles, M. B., & Huberman, A. M. (1994). *Qualitative data analysis: An expanded sourcebook* (2nd ed.). Sage publications: Thousand Oaks.

Patel, A., Bakhtiyari, K., & Taghavi, M. (2011). Evaluation of cheating detection methods in academic writings. *Library Hi Tech*, *29(4)*, 623-640. https://doi.org/10.1108/07378831111189732

Pecorari, D. (2001). Plagiarism and International Students: How the English-Speaking University Responds. In D. D. Belcher & A. R. Hirvela (Eds.), *Linking Literacies: Perspectives on L2 Reading-Writing Connections* (pp. 229–245). University of Michigan Press.

Playfoot, D., Quigley, M., & Thomas, A. G. (2024). Hey ChatGPT, give me a title for a paper about degree apathy and student use of AI for assignment writing. *The Internet and Higher Education*, *62*, 1-10. https://doi.org/10.1016/j.iheduc.2024.100950

Pudasaini, S., Miralles-Pechuán, L., Lillis, D., & Salvador, M. L. (2024). Survey on plagiarism detection in large language models: The impact of chatgpt and gemini on academic integrity. *arXiv preprint arXiv:2407.13105, 23,* 1137-1170. https://doi.org/10.48550/arXiv.2407.13105

Rezaei, M., Salehi, H., & Tabatabaei, O. (2024, February). ChatGPT, a Helpful Scaffold or a Debilitating Crutch for Academic Writing? In *2024 11th International and the 17th National Conference on E-Learning and E-Teaching (ICeLeT)* (pp. 1-5). IEEE. 10.1109/ICeLeT62507.2024.10493087

Rogerson, A. M., & Basanta, G. (2016). Peer-to-peer file sharing and academic integrity in the internet age. In T. Bretag (Ed.), *Handbook of academic integrity* (pp. 273-285). Springer, Singapore.

Thao, L. T., Hieu, H. H., & Thuy, P. T. (2023). Exploring the impacts of chatGPT in EFL writing student perceptions of opportunities and challenges in Vietnamese higher education. *Kognisi: Jurnal Ilmu Keguruan*, *1*(2), 107-124. https://doi.org/10.59698/kognisi.v1i2.175

Tsai, C. Y., Lin, Y. T., & Brown, I. K. (2024). Impacts of ChatGPT-assisted writing for EFL English majors: Feasibility and challenges. *Education and Information Technologies*, *29*, 22427- 22445. https://doi.org/10.1007/s10639-024-12722-y

Uchendu, A., Ma, Z., Le, T., Zhang, R., & Lee, D. (2021). Turingbench: A benchmark environment for turing test in the age of neural text generation. In *Findings of the Association for Computational Linguistics: EMNLP 2021* (pp. 2001–2016). Punta Cana, Dominican Republic. Association for Computational Linguistics. https://doi.org/10.48550/arXiv.2109.13296

Walker, J. (2010). Measuring plagiarism: Researching what students do, not what they say they do. *Studies in Higher Education*, *35(1)*, 41-59. https://doi.org/10.1080/03075070902912994

Yan, D. (2023). Impact of ChatGPT on learners in a L2 writing practicum: An exploratory investigation. *Education and Information Technologies*, *28(11)*, 13943–13967. https://doi.org/10.1007/s10639-023-11742-4