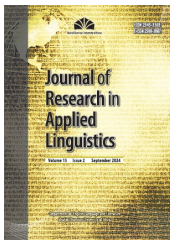
Journal of Research in Applied Linguistics

ISSN: 2345-3303 – E-ISSN: 2588-3887 – https://rals.scu.ac.ir

Please cite this paper as follows:

Bannister, P. (2024). English medium instruction educator language assessment literacy and the test of generative AI in online higher education. *Journal of Research in Applied Linguistics*, *15*(2), 55-72. https://doi.org/10.22055/rals.2024.45862.3214

Shahid Chamran University of Ahvaz



Research Paper

English Medium Instruction Educator Language Assessment Literacy and the Test of Generative AI in Online Higher Education

Peter Bannister

Doctoral School, Vice Rectorate for Research, Universidad Internacional de La Rioja, La Rioja, Spain; *peter.bannister@unir.net*

Received: 16/01/2024 Accepted: 24/08/2024

Abstract

Generative Artificial Intelligence (GenAI) is often portrayed as disruptive. Higher Education (HE) assessment is not exempt from this, although the implications for multilingual settings remain an area of limited exploration. Drawing on the scarce literature in English as a Medium of Instruction (EMI) assessment and EMI educator language assessment literacy (LAL), this study sought to explore EMI online HE educator LAL and awareness of GenAI's potential impact on established language assessment praxis. A sequential explanatory mixed-methods approach was used, comprising a survey on LAL self-perceptions (n=174) and semi-structured interviews (n=12). Findings illustrate general tendencies of low levels of LAL and practitioner unease towards GenAI-assisted academic misconduct. A heightened lack of confidence in GenAI tool usage detection efficacy and HE institutional capacity to respond to their evolving capacities in a timely manner was also found. It is therefore suggested that, given the complexity and continuing swift development of GenAI tools, the implementation of continuous professional development programmes focused on enhancing EMI educators' language assessment literacy and competence in using GenAI technologies is prioritised. These findings underscore the need for initiatives that not only improve technical skills but also address ethical considerations and strategies to uphold academic integrity in the face of emerging GenAI capabilities.

Keywords: Generative Artificial Intelligence; English as a Medium of Instruction; Language Assessment Literacy; Academic Integrity; Online Higher Education.

1. Introduction

The advancement of Generative Artificial Intelligence (GenAI) has led to widespread scrutiny of traditional conventions and paradigms in Higher Education (HE), in both its face-to-face and online manifestations (Bozkurt & Sharma, 2023). Globally, academic communities continue towards fully comprehending and addressing the potential ramifications and affordances of GenAI on teaching, learning, and assessment in different pedagogical settings (Dawson, 2021; Sullivan et al., 2023). English as a Medium of Instruction (EMI) is defined as a dual-focused education model attending to the learning of subject contents and the English language, in which it is not otherwise employed as the principle communicative vehicle (Macaro, 2018; McKinley & Rose, 2022). These settings represent a high-stakes battlefront (Pack & Maloney, 2023) in which the assessment of student second language (L2) academic proficiency arises as a complex challenge owing to GenAI enhanced text production capacities (Wang et al., 2023).

There are broader underlying philosophical, sociological, and anthropological issues also brought forth by GenAI advances (Sudajit-apa, in press). Such matters of contention include the gradual ebbing away of intentionality, creativity, and originality as assumed uniquely characterizing traits of biological human beings (Boden, 2004; Sharples & Pérez y Pérez, 2022); the retention of human control over ever-more capable, and perhaps eventually, superior machines (Russell, 2019); and the new inequitable frontier in the digital divide and the HE awarding gap potentially aggravated by advanced GenAI model premium access payment schemes (Zajko, 2022), despite claims to the contrary (Fido & Wallace, 2023).

In this present landscape, it is understandable that future-orientated speculation and prediction have characterized swathes of discourse in academia fueled by international mediatic attention particularly since the launch of ChatGPT in

November 2022, giving rise to contributions such as that of Bozkurt et al. (2023). However, here attention is firmly set on the present. Whilst specific in focus, this work strives to tackle some of the aforementioned broader issues at stake through the prism of educator language assessment literacy and teacher preparedness for the additional challenges of GenAI in EMI online HE (Derakhshan & Shakki, 2024).

2. Literature Review

In a critical review of the conceptualisation of the term AI in HE research literature published prior to 2021, Bearman et al. (2022) identify two saliant discourses: "the advent of unprecedented sociotechnical change and how higher education has an imperative to respond" and "how AI is altering the locus of authority and agency surrounding academic work" (p. 374). Henceforth, novel scholarly work continues to emerge as the academic community begins to develop a deeper understanding of the most recent publicly available GenAI tools.

2.1. GenAI in HE

Numerous authors have sought to examine the technology in terms of HE practical implications (cf. Baidoo-Anu & Owusu Ansah, 2023; Bañeres et al., 2023; Bearman & Ajjawi, 2023; Eager & Bunton, 2023; Su & Yang, 2023, amongst others). Affordances conveyed include personalized feedback, learner performance prediction, and enhanced student engagement. Their findings also draw attention to shortcomings, i.e. ethical and safety concerns, and untested effectiveness of a technology with data quality limitations (Yu & Guo, 2023). Moreover, in its current iteration, Tzirides et al. (2023, p. 30) unfavorably frame GenAI in HE teaching and learning praxis, given that "it undermines some of the key epistemological bases of modern science and reliable knowledge systems", and propose necessary user-facilitated "epistemic, empirical and ontology-based recalibration" for it to truly embody an effective and purposeful supporting function in HE.

Further investigation has documented stakeholder perceptions of GenAI in HE. Addressing the notion of increasing automatization and potential human obsoletion in the field, certain studies encouragingly report on the preferred key stakeholder conceptualization of GenAI as learning assistant as opposed to educator replacement at present and going forward (Chen et al., 2023). Other authors, nonetheless, draw attention to the perceived need for action within Higher Education Institutions (HEI) globally, recommending that "evidence-based" guidance and policies for responsible GenAI use continue be developed that focus on AI literacy and critical thinking enhancement (Chan & Lee, 2023, p. 22). Although, others openly raise concern about HEI capacity to respond to these challenges in a timely and comprehensive manner (e.g., Carrigan, 2023).

Turning to tool usage, Amani et al. (2023, p. 7-8) found that the majority of faculty and student respondents in the USA had utilized GenAI tools such as ChatGPT. Both groups coincided that, whilst useful for technical questions and the explanation of concepts, due to benefits such as "personalized learning and effective feedback", these may impact negatively on both learner "critical thinking and problem-solving" skills (p. 8). Data yielded from the study also conveyed a perception of learner tool usage for assessment amongst students with 63% of respondents expressing concern that such tools would be used to engage in academic misconduct.

2.2. GenAI and HE Assessment

Underlining the global generalizability of GenAI's potential impact, other geographically-diverse publications articulate similar conclusions. A limited number of studies extol affordances in enhancing assessment design (Yildrim-Erbasali & Bulut, 2023), efficacy in facilitating test correction (Chen, 2022), and as a means of learner cognitive offloading (Dawson, 2020). However, a greater number of publications declare that GenAI tools pose a cross-disciplinary substantial level of risk to academic integrity and that sector-wide collaborative action is needed (Dwivedi et al., 2023; Eke, 2023; Fergus et al., 2023; Nikolic et al., 2023). Dawson (2021) also emphasizes the difficulties in drawing a line between what may be constituted as AI help and AI-assisted e-cheating.

To that end, as expert consensus continues to be determined, the development of seemingly promising GenAI text classifiers was initially warmly welcomed in practice. Nevertheless, subsequent investigation has disputed the validity and efficacy of these tools at the time of writing (Dalalah & Dalalah, 2023; Elkhatat, 2023; Sadasivian et al., 2023), and warned that students who use English as a second or additional language may be particularly vulnerable to these flaws

(Liang et al., 2023). Moreover, Sharples (2022, p. 1125) claims that GenAI-assisted writing signifies that an education model dependent on summative written assessment "may have reached its apotheosis", signaling a move away from the oft-maligned and yet prevailing traditional methods such as the essay (Stobart, 2008). Propitious green shoots have begun to emerge in scholarship, although these are understandably still in their infancy. For instance, the exploration of different assessment media and moving towards authentic assessment (Rudolph et al., 2023), the creation of multidisciplinary working groups to develop and enhance HEI policy provision, and the fostering of AI assessment literacy proficiency for both staff and students (Bannister et al., 2023a).

2.3. GenAI and EMI HE Assessment

As change is called for at a macro-level, domain-specific works begin to shed some light on the bespoke challenges for different HE settings. As mentioned previously, EMI HE constitutes a somewhat saliant pedagogical context here. Learner GenAI tool usage in L2 writing proficiency assessment amongst international student cohorts raises serious authorial and linguistic questions of construct validity (Johinke et al., 2023). However, some authors do note that hybrid human-AI writing is to become the new norm as one of the key tenets of postplagiarism (Eaton, 2021). Despite this, there is a marked dearth of scholarship within the field that examines the impact of GenAI on EMI HE assessment, as may be expected at this early stage of proceedings (Bannister et al., 2023b). Intriguingly, however, historical publication trends also reveal a comparatively finite quantity of works that broach EMI assessment per se (Hultgren et al., 2022; Macaro, 2022), an avenue of investigation and praxis also formerly deemed "problematic" by Dearden (2014, p. 17).

Despite this, fledgling research efforts in this regard are encouraged, as per the EMI Research agenda formulated by Sah (2022) after conducting a series of interviews with leading scholars in the field. Perhaps this task is even more urgent now considering the potential magnitude of the phenomenon at hand.

2.4. Language Assessment Literacy and EMI

Language assessment literacy (LAL) represents not only an apt point of entry, but also an area which has received comparatively more substantive empirical and conceptual attention in cognate disciplines such as EFL (cf. Tsagari & Vogt, 2017) and CLIL (cf. Liu et al. 2023; Andujar & Spratt, 2023). Notwithstanding reservations over the mis- and overuse of the term "literacy" of late (Nieminen & Carless, 2023), Fulcher's (2012) working definition, continues to be of relevance. He posits that LAL may be broadly thought of as the amalgamation of:

> knowledge, skills and abilities required to design, develop, maintain or evaluate, largescale standardized and/or classroom based tests, familiarity with test processes, and awareness of principles and concepts that guide and underpin practice, including ethics and codes of practice (Fulcher, 2012, p. 125).

There are a range of interpretations of this premise which offer legitimate nuance to this. For example, Taylor (2013) conceptualized LAL profiles encompassing eight core dimensions of LAL knowledge, skills, and principles, namely, knowledge of theory, technical skills, principles and concepts, language pedagogy, sociocultural values, local practices, personal beliefs/attitudes, scores, and decision making. In turn, Scarino (2013, p. 324) convincingly advocates for the contemplation of beliefs, attitudes, and personal conceptions in LAL, considering that practitioners are required "to integrate simultaneously the complex theoretical, practical and institutional dimensions of the assessment act and an understanding of self in relation to these". Furthermore, additional research has elucidated the developmental nature of LAL as opposed to an inseparable knowledge mass waiting to be acquired. To this end, several models have been conceptualized which aim to accentuate differing levels of competency (e.g., Harding & Krammel, 2016; Pill & Harding, 2013, as cited in Tsagari, 2020) with growing calls for stakeholder literacy profiles and involvement to be considered going forward (Harding et al., 2022). For a comprehensive systematic review of further LAL research publications, see Gan and Lam (2022).

Correlated empirical research conducted in EMI contexts is markedly scarce. Lasagabaster (2022) asserts that practitioners working in EMI HE were found to be generally underprepared irrespective of their geographical provenance, which conceivably extends to educator LAL, too. Similarly, Shahzadi & Ducasse (2022) found that EMI HE lecturers in Pakistan undertook training in which LAL was generally not addressed and this, together with limited knowledge of language assessment theory and practice, led to the conclusion that the participants had a "low level of LAL" (p. 105).

Mancho-Barés et al. (2022) examined both faculty and student LAL in an Engineering setting in Spain and concluded with the methodological affordances of genre analysis in such endeavors.

At the time of writing, one sole publication in relatively close thematic proximity to the present study was localisable. Farazouli et al. (2023) sought to clarify the mediating role of ChatGPT on educator assessment praxis. Faculty participants were found to fail texts whose human authorship they doubted, citing instances of "non-sensical statements about factual knowledge, strange use of synonyms and perceived translations of English terms, and repetitions of the prompt of the examination question" for this, although this was not always detected and passing grades were assigned to work produced by ChatGPT at times (p. 7).

2.5. Research Questions

Considering the present somewhat sparce scholarly landscape and the seemingly transformational disposition of the continuingly evolving implications of GenAI tools for EMI HE assessment, the purpose of this exploratory study is to make a dual novel contribution to the existing body of knowledge by not only examining international educator LAL but also their level of teacher preparedness regarding GenAI and assessment. To that end, the following research questions (RQ) were composed:

1. To what extent are international EMI online HE educators language assessment literate?
2. How cognizant are international tertiary EMI online lecturers of the potential impact of GenAI on established language assessment praxis?

3. Methodology

Given the complexity of the matter under investigation, a sequential explanatory mixed-methods design was determined as an effective means to gather and analyze the data yielded in concordance with both theoretical (Creswell & Plano Clark, 2011) and more recent cognate literature (e.g., Afshar & Ranjbar, 2021).

3.1. Participants

Participants for the study were selected using non-random convenience sampling. A total of 417 email invitations were sent to international EMI instructors who work in online HE, inviting them to complete the questionnaire and partake in the second stage of individual online semi-structured interviews. From this, a total of 174 participants anonymously completed the online asynchronous questionnaire and 12 practitioners participated in the online interview, with some respondents stating that busy working schedules impeded commitment to the latter stage.

The study population included 108 women and 66 men and comprised participants from Spain (n=96), China (n=44), Mexico (n=14), Brazil (n=12), and United Arab Emirates (n=8). The respondents worked at a total of 13 different HEIs, all of which have an entirely or in part online teaching and learning provision. All respondents had a minimum of five years HE professional experience, and either had (n=122) or were working towards a PhD (n=34) or were in possession of a postgraduate level qualification (n=18).

3.2. Data Collection

Two complementary phases were established, namely the completion of a quantitative and theoretically validated questionnaire, Language Assessment Literacy Survey (Kremmel & Harding, 2020), and a subsequent qualitative round of semi-structured interviews. Prior to implementation, two experts in the field carried out an evaluative review to validate that the proposed research instruments used were fit for purpose, and only after having received approval did the data collection procedures commence.

3.2.1. Stage One

In the first stage, Kremmel and Harding's (2020) self-evaluation Language Assessment Literacy Survey was used to collect quantitative data. This validated data collection instrument was selected as it draws on the earlier work of LAL profiles conceptualized by Taylor (2013), state-of-the-art LAL frameworks research, and input from stakeholders.

Thus, as conceptualized by Kremmel and Harding (2020), questions relate to the following areas: developing and administering language assessments, assessment in language pedagogy, assessment policy and local practices,

personal beliefs and attitudes, statistical and research methods, assessment principles and interpretation, language structure, use, and development, washback and preparation, and scoring and rating. Furthermore, the developers of the survey conceptualise the profiles of three groups in their article, namely "language test/assessment (LTA) developers, language testing/assessment (LTA) researchers and language teachers" (p. 109). Figure 1 below illustrates the LAL needs profiles:

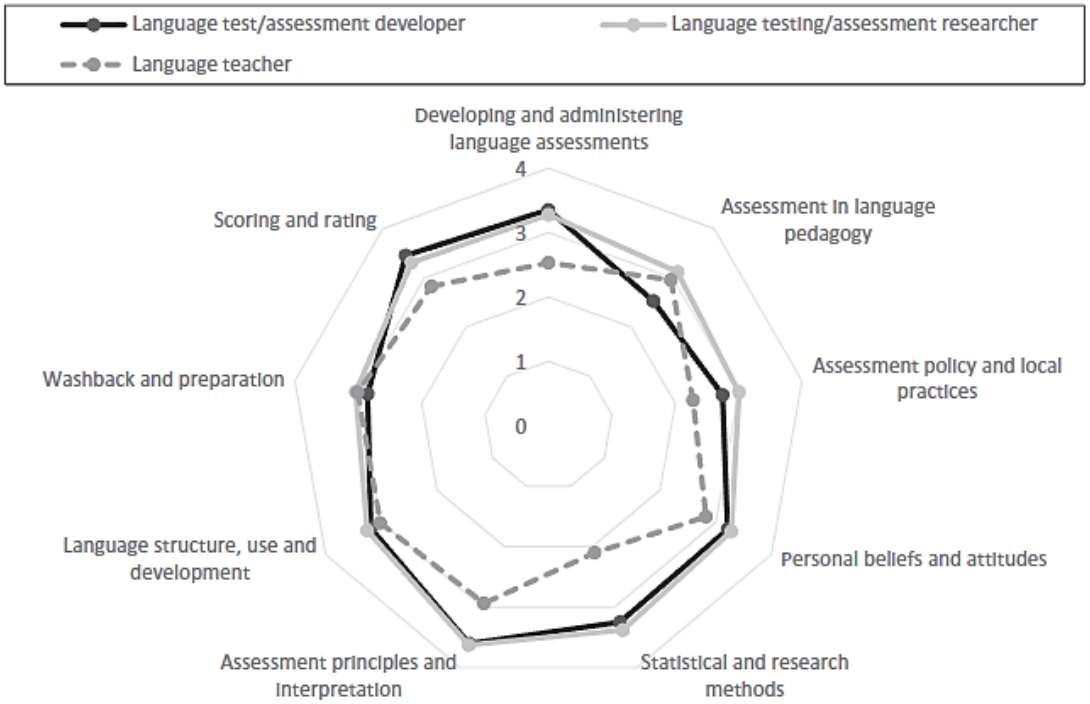


Figure 1. *LAL Stakeholder Needs Profiles*

Considering the respondents' professional experience, the expected profile would be that of language teacher. For each of the self-evaluation questions, respondents are required to submit a score as is detailed below in Table 1 (Harding & Kremmel, 2020):

Table 1. *Summary of Questionnaire Response Values*

Numerical Value	Significance
0	not knowledgeable at all
1	slightly knowledgeable
2	moderately knowledgeable
3	very knowledgeable
4	extremely knowledgeable

Given the focus of this study, a bespoke addendum to the configuration was made by the author of this study to the instrument. Thereby, in a subsequent phase, respondents were asked to rate themselves once again using the same question items, but on this occasion, to do so in terms of the potential impact that GenAI may have on each of the dimensions as outlined previously.

A total of 174 participants responded to the survey. Total anonymity was afforded with the first stage of the questionnaire comprising initial questions to create a bespoke anonymous participant ID for later use in data analysis.

3.2.2. Stage Two

One-to-one online semi-structured interviews were employed in this subsequent stage. Such an approach has been proven to be an effective means of collecting data when addressing complex matters (Ruslin et al., 2022) such as is the matter at hand. Whilst allowing for preparation of questions which may be deployed with a degree of flexibility, it

allowed for open-ended response data collection and spontaneous unstructured lines of questioning to explore respondent ideas in greater depth (Adams, 2015).

The question list elaborated by Vogt and Tsagari (2014) was used as a starting point for the semi-structured interviews. Subsequent questioning then turned directly to matters pertaining to RQ2. 12 EMI HE lecturers participated in the semi-structured interviews.

3.3. Data Analysis

Quantitative data were analyzed using descriptive statistical techniques in SPSS Statistics (version 29.0). Weighted frequency and percentages were examined to summarize and describe the quantitative results. Moreover, content analysis was utilized to systematically evaluate and interpret the qualitative data gathered from the open-ended questions included in the data collection process for this study. Specifically, open coding and thematic analysis techniques were used to identify overarching themes and patterns within participants' responses to these open-ended questions (Braun et al., 2019; Braun & Clarke, 2023).

4. Results

4.1. International Online EMI HE Educator LAL

Having conducted the procedures of data analysis of the first round of the Language Assessment Literacy Survey, participants were found to have a generally low level of LAL compared to the needs profile for language teachers as detailed previously. There were however a limited number of outliers (n=28), that is to say 16.09% of respondents, all of whom were female and held a PhD, were found to be language assessment literate as is illustrated in Table 2 below:

Table 2. *Questionnaire Phase 1 Quantitative Data Analysis Results*

LAL Dimension	Numerical Response	WF	P
Developing and administering language assessments (14 questions)	0	0	0.00
	1	4	2.29
	2	47	27.01
	3	115	66.09
	4	8	4.59
Assessment in language pedagogy (6 questions)	0	4	2.29
	1	5	2.87
	2	51	29.31
	3	96	55.17
	4	18	10.34
Assessment policy and local practices (6 questions)	0	20	11.49
	1	73	41.95
	2	55	31.60
	3	26	14.94
	4	0	0.00
Personal beliefs and attitudes (4 questions)	0	0	0.00
	1	3	1.72
	2	69	39.65
	3	87	50.00
	4	15	8.62
Statistical and research methods (4 questions)	0	29	16.66
	1	60	34.48
	2	53	30.45
	3	28	16.09
	4	4	2.29
Assessment principles and interpretation (4 questions)	0	12	6.89
	1	70	40.22
	2	64	36.78
	3	18	10.34
	4	10	5.74

Language structure, use, and development (5 questions)	0	0	0.00
	1	19	10.91
	2	40	22.98
	3	103	59.19
	4	12	6.89
Washback and preparation (4 questions)	0	0	0.00
	1	11	6.32
	2	46	26.43
	3	94	54.02
	4	23	13.21
Scoring and rating (3 Questions)	0	0	0.00
	1	3	1.72
	2	44	25.28
	3	119	68.39
	4	8	4.59

*Note. Definitions of numerical response values found in Table 1. WF= weighted frequency. P= percentage.

Respondents demonstrated that assessment policy and local practices together with assessment principles and interpretation were the dimensions in which they were least confident. In-class work such as language structure, use, and development and washback and preparation were however generally rated higher. Assessment in language pedagogy gleaned a mixed response from participants, seemingly indicating a varied range of formative experience amongst participants.

4.2. GenAI Impact on International Online EMI HE Educator LAL

The second round consisted of the repetition of the self-evaluation questionnaire in light of the potential impact of GenAI on international online EMI HE LAL. The data here illustrate a general deterioration in the level of respondent confidence and proficiency even amidst those who were previously deemed to be assessment literate in accordance with the established needs profile, as is detailed in Table 3 below:

Table 3. *Questionnaire Phase 2 Quantitative Data Analysis Results*

LAL Dimension	Numerical Response	WF	P
Developing and administering language assessments (14 questions)	0	31	17.81
	1	76	43.67
	2	67	38.50
	3	0	0.00
	4	0	0.00
Assessment in language pedagogy (6 questions)	0	16	9.19
	1	40	22.98
	2	111	63.79
	3	7	4.02
	4	0	0.00
Assessment policy and local practices (6 questions)	0	98	56.32
	1	51	29.31
	2	25	14.36
	3	0	0.00
	4	0	0.00
Personal beliefs and attitudes (4 questions)	0	38	21.83
	1	46	26.43
	2	66	37.93
	3	24	13.79
	4	0	0.00
Statistical and research methods (4 questions)	0	28	16.09
	1	55	31.60
	2	53	30.45
	3	36	20.68
	4	2	0.00

Assessment principles and interpretation (4 questions)	0	21	12.06
	1	40	22.98
	2	92	52.87
	3	15	8.62
	4	6	3.44
Language structure, use, and development (5 questions)	0	0	0.00
	1	6	3.44
	2	49	28.16
	3	114	65.51
	4	5	0.00
Washback and preparation (4 questions)	0	18	10.34
	1	33	18.96
	2	90	51.72
	3	26	14.94
	4	7	4.02
Scoring and rating (3 Questions)	0	17	9.77
	1	46	26.43
	2	98	56.32
	3	13	7.47
	4	0	0.00

* Note. Definitions of numerical response values found in Table 1. WF= weighted frequency. P= percentage.

The juxtaposition here is marked across almost all of the dimensions with the exception of language structure, use, and development. This shift represented in responses highly skewed towards lower values highlights general tendencies of uncertainty, lack of expertise and experience across the study population irrespective of gender, geographical provenance, or qualifications.

These tendencies were also reflected in the second stage of the data collection procedure in the semi-structured interviews. Due to the qualitative nature of the data here, a deeper understanding that builds on these initial findings has been ascertained thanks to the thematic analysis undertaken. Table 4 below offers a summary of the main themes identified and examples of these are subsequently illustrated in the proceeding lines:

Table 4. *Overview of Thematic Structure of Qualitative Data*

Main Themes	Subthemes
1. More GenAI tools vs. Limited Knowledge of Impact	1.1. Rapid advancements in AI capabilities outpace understanding of educational impact
	1.2. Ethical and legal concerns surrounding GenAI usage
2. Lack of Effective Local Detection Tools and Procedure	2.1. Plagiarism detection software not optimized for AI-generated content
	2.2. Faculty need training on AI detection methods
3. Increased Educator Evaluation Time	3.1. Additional workload and time burden for educators
	3.2. Plagiarism detection software not optimized for AI-generated content
	3.3. Faculty need more training on AI detection methods
	3.4. Impact on instructional time and teacher-student relationships
4. Delayed and Ineffectual HEI Policy Response	4.1. Slow development of HEI policy compared to rapid AI advances
	4.2. Lack of consensus on how to address AI in HE assessment
	4.3. Ineffective policies and guidelines in addressing issues
5. Future Job Security	5.1. Concerns about automation and job displacement
	5.2. Importance of upskilling and reskilling for faculty and student future job roles
6. HE Paradigm Disruption	6.1. Pressure to reform traditional educational models and assessments
	6.2. Questioning role of educators in GenAI-driven world

One such prominent theme was the growing number of seemingly ever-more capable range of GenAI tools and the lack of awareness as to their potential impact on academic integrity in this pedagogical setting was raised as an area of particular concern by several interview participants:

> "Language, grammar, and structure are our main focus, so if this can all be done in under a minute by one of these AI apps, it pretty much defeats the point of EMI HE" (Interview 611).

> "This is so much more than ChatGPT and we just can't keep up with what they can all do. Academic integrity for us in the field has never been more vulnerable and at times I feel like we are defenseless- I just don't know enough" (Interview 1128).

Furthermore, the lack of effective detection tools and procedures locally and institutionally to identify GenAI academic misconduct in student submissions was identified as a theme:

> "I naively thought that it was game-over when I found a couple of sites that claimed to detect AI-generated text and then Turnitin came out and announced similar features. But in practice, looking into it myself, they aren't as good as I thought. If a student really wants to, they could cheat these systems quite easily" (Interview 356).

> "I tried using the detection software and after a couple of tries, it said what I had written was AI-produced" (Interview 434).

This links with a further theme of increased time needed for marking and reporting:

> "Well, if you actually have the time to conduct multiple writing tasks, assessed or not, then you can get to know your level of students writing. In larger groups of course this is not practical. It does all take a lot longer to actually mark something a student has produced that's not under strict exam conditions though" (Interview 633).

> "I've caught a couple of students already and quite simply I didn't give them feedback or a grade. What they had written just didn't add up. Marking now does take a lot longer than it used to and even then, I'm sure there will be a couple that slip through" (Interview 1012).

Although respondents pointed out that HEIs were taking action, the theme of delayed and ineffectual institutional policy response for the particularities of EMI online HE contexts was identified:

> "My institution has promised to make a new academic integrity policy to deal with all this, but in reality, on the front line it has taken what seems like forever to come about and in reality, it's little more than window dressing" (Interview 578).

> "I was sure that one of my students had cheated, when I tried to elevate it following the policy, I was pretty much ignored" (Interview 175).

> "Every time we ask about the whole preventive policy, we are told that it is coming soon. I tell my students that we will come down on them heavily but in reality, I don't think we have the internal means in place to properly deal with how widespread this could actually be" (Interview 848).

A somewhat less saliant theme was also identified, thereby, future job security within the field of EMI HE:

> "And at the bottom of all this is the fact maybe in ten years' time, who knows if we will just be out of a job… maybe HE will be less about human intervention when you have an army of AI bots who can teach, assess, and evaluate" (Interview 794).

> "I always used to think how silly those people were that bleated on about being replaced by machines, but from the little I know now, I think at some point in the future part of our work may be outsourced to machines" (Interview 1262).

Finally, the theme of general HE paradigm disruption was also identified as:

> "What is clear right now is that we are witnessing shifting sands of our understanding of authorship, and critical engagement, and then of course learning, and assessment" (Interview 207).

5. Discussion

5.1. Applicability, Breadth of Impact, and Contribution

Findings are relevant to the context in which the study is grounded as a potential snapshot as to the LAL status quo in a cross-section of international HEIs. The lack of preparedness illustrated in the findings in response to RQ1 is in line with cognate studies indicated previously (e.g., Lasagabaster, 2022; Shahzadi & Ducasse, 2022). A need for more extensive continuous professional development opportunities and engagement both on assessment, and the impact of GenAI on established assessment praxis is thus highlighted.

Results here support calls for HE stakeholder AI literacy development (Brew et al., 2023) and the conclusions of Bearman and Ajjawi (2023, p. 1167), who advocate for defining "tacit and explicit rules of the game'' and "meaningful interactions with AI systems". This could refer to assessment security and the preservation of academic integrity, and the automation of certain assessment procedures, such as assessment design or scoring, as a means for lecturer cognitive offloading (Dawson, 2020). This time-saving premise is juxtaposed with the additional time required to address potential AI-assisted academic malpractice in submissions. The findings also reiterate the need to comprehensively address how these tools may be used to enhance assessment design, delivery, and evaluation in pre-service and in-service formative programmes (Chen, 2022; Yildrim-Erbasali & Bulut, 2023).

Furthermore, the potential breadth of impact may contemplate an audience which goes beyond EMI HE practitioners and researchers. The lack of localized agreed procedure to detect the specific text-level traits of AI-assisted academic misconduct and how to handle this, the elevated time needed to evaluate student submissions, and the lack of confidence in AI detection tools may, in fact, be indicative of more generalized cross-disciplinary tendencies both in face-to-face EMI HE settings and non-EMI HE context in its different manifestations, or even at different educational levels. Therefore, these findings would seemingly corroborate results from other more recent studies and support calls for sector-wide collaboration named earlier to tackle the risk to academic integrity (Dwivedi et al., 2023; Eke, 2023; Fergus et al., 2023; Nikolic et al., 2023).

5.2. GenAI and Online EMI HE Educators LAL

Results key into the two salient discourses in AI HE research of AI-induced authority locus alteration and HE imperative response as identified by Bearman et al. (2022). In online EMI HE, albeit there have been bespoke institutional policy responses to regulate the use of GenAI, these often fall short or do not adequately cater for the linguistic and authorial implications of such use in this given context (Johinke et al., 2023; Pack & Maloney, 2023). As was raised by a semi-structured interview participant, HEI response time was characterized as a further obstacle doing little to assuage concerns in practice which align with Carrigan's (2023) remarks outlined previously.

This, and developing expert consensus on the precise GenAI implications on online EMI HE assessment praxis, constitute major areas of concern which ought to be urgently prioritized in both research and practice going forward. Whilst further confirmatory evidence has been obtained here to support the findings of Amani et al. (2023) regarding stakeholder concerns of GenAI tool usage as an additional means of e-cheating (Dawson, 2021), findings also accentuate the disruptive nature of the phenomenon and the lack of compatibility between GenAI tools and the status quo within HE. It also appears to offer certain credence to the potential paradigm shift hypothesized by Eaton (2021) to, in what she terms, an era of postplagarism, characterized by AI-human hybrid writing, amongst other key tenets. Such a stance does, however, need further measured consideration.

Data collected from both stages show tendencies of educator concern on the possibility of future human obsoletion within HE, keying into the concerns also highlighted in studies penned by Chen et al. (2023) and Odden et al. (2023). They additionally question the epistemological, ontological, and axiological assumptions which underpin such tools and their validity for the present HE educational paradigm. Certain concordance is found, therefore, between the results here and the conclusions of Tzirides et al. (2023) to this end mentioned previously. Building on these, qualitative data from an interview respondent warned of "misplaced student trust in ChatGPT responses in terms of academic accuracy and quality" (Interview 665). Moving away from a strict instrumentalist conception of GenAI tools, this seemingly draws on the fallacy of ad verecundiam, or of appeal to authority, in that the automated expert in which the

student places their trust is not in possession of the epistemic authority required to produce claims of sufficient reliability (Battersby, 2019; Koszowy & Walton, 2019).

In addition to emphasizing the apocryphal nature of GenAI, this participant response also sheds light on a further issue. Thereby, at present, such technologies are intrinsically limited, in that, despite producing human-like text in response to lines of questioning with varying levels of success, they do not truly know in the strict human sense of the word (Yufik et al., 2022). In contrast, the output formulated by GenAI is the end-product of the analysis of vast amounts of linguistic concurrences from text corpora from which it is able to predict desired responses based on familiarity (Deng et al., 2020). In other words, human knowledge may be defined as epistemological given that composite knowledge may be subjected to deconstruction and, subsequently, the potential links between different concepts may be established as they are in reality. This premise is counter to that of GenAI, which is solipsistic, as the technology is inherently bound by the limitations of the dataset(s) on which it is sustained.

In turn, these impose restrictive boundaries and confine the technology to the mere repetition and recognition of statistical patterns without being able to further substantiate on the source(s). A social conceptualization of GenAI for exploration based on dialogic human-user and chatbot interaction may still however be valid (Sharples, 2023) which may seemingly partially be in thematic proximity with education. Nevertheless, given that they are not sources of epistemic authority, they remain tools devoid of the ability to critique and reason through logic at present. Therefore, rather than regarding LLMs solely as oracles to be exploited in a purely pragmatic manner, consideration should be given to their outputs and conceptual underpinnings as catalysts for thoughtful and creative academic deliberation. Such an approach may be instrumental in guiding the expansive, collaborative pursuit of honing scientific comprehension going forward (Birhane et al., 2023). Nonetheless, knowledge construction as a core component of HE remains an entirely biological human activity.

What an interviewee called the “shifting sands of our understanding of authorship, and critical engagement, and then of course learning, and assessment” (Interview 1207) is related to the issue of human control over ever-more capable machines (Russell, 2019). It would also seemingly denote the challenges faced in staying the course within education to train informed future professionals who are capable of confidently articulating their own voice intentionally, creatively, and originally through academic discourse (Boden, 2004; Sharples & Pérez y Pérez, 2022). Furthermore, it emphasizes the necessity to redouble efforts to foster learner critical thinking and discernment as a cornerstone of HE pedagogy and assessment (Chan & Lee, 2023). In short, GenAI technologies ought to be conceptualized as a further legitimate compliment for all, irrespective of student affluence (Zajko, 2022), to enhance justly time-consuming and developmental learner-centric instruction and evaluation. In contrast, every effort ought to be made to ensure that GenAI tools are not exploited as a potential illegitimate fast-track shortcut to the finishing line that would arguably render the entire learning experience devoid of purpose.

5.3. Exploring the Pedagogical Potential of GenAI in EMI HE

While the findings undoubtedly underscore the challenges and apprehensions surrounding the integration of GenAI in EMI HE contexts, particularly concerning assessment integrity, it is imperative to approach this technological advancement with a judicious perspective. The qualitative data elucidates participants' recognition of GenAI's utility for "technical questions and the explanation of concepts" (Amani et al., 2023, p. 8), alluding to its potential as a pedagogical aid. Congruently, several scholars have expounded on the affordances of GenAI in fostering personalized feedback (Bañeres et al., 2023), learner performance prediction (Su & Yang, 2023), and enhanced student engagement (Baidoo-Anu & Owusu Ansah, 2023).

Nonetheless, the survey results indicate that a substantial proportion of respondents (38.50%) perceived themselves as only "moderately knowledgeable" regarding the development and administration of language assessments, even prior to considering GenAI's implications. This suggests a prevalent need for technological pedagogical content knowledge (TPACK) development (Mishra & Koehler, 2006), whereby educators can harness GenAI's potential as a cognitive offloading tool (Dawson, 2020). As posited by Chen (2022), GenAI could facilitate the automation of specific assessment procedures, alleviating the "additional workload and time burden for educators" (Table 4) expressed in the interviews. However, the findings accentuate the necessity of circumventing an uncritical instrumentalist adoption of GenAI. As elucidated in the discussion, these technologies ought to be conceptualized as "catalysts for thoughtful and

creative academic deliberation" rather than repositories of epistemic authority. This premise aligns with Yildirim-Erbasali and Bulut's (2023) advocacy for "conversation-based assessment," wherein GenAI is leveraged to foster critical thinking and substantive engagement with course.

In essence, while the study's results foreground the risks and uncertainties surrounding GenAI's impact on EMI assessment practices, they simultaneously illuminate potential avenues for its judicious integration. A balanced approach, anchored in continuous professional development, ethical frameworks, and a reconceptualization of assessment paradigms, could harness GenAI's capabilities to enhance language teaching, learning, and evaluation. Ultimately, as articulated by the participants, the proliferation of GenAI necessitates a recalibration of higher education praxis to uphold core tenets of academic integrity while embracing technological innovation.

5.4. Limitations

Despite the potential headway made, there are several limitations to the study. Firstly, although the population sample includes a variety of participants from numerous different countries, it is somewhat limited in size and geographical representation, meaning that future studies ought to include a wider range of diverse respondents to improve generalisability of results. This point is particularly acute in the second phase of the study in which only 6.89% of respondents were able to participate principally owing to limited availability. A further potential limitation is that this study focused solely on online EMI HE which means that findings may vary amidst professionals in face-to-face settings.

In turn, perhaps, the greatest challenge for this study resides with the lack of established consensus as to what constitutes GenAI-specific LAL. It may also be noted that the study has conceptualized participant LAL as if it were a snapshot at a given point in time, which whilst useful is ultimately somewhat limited in scope, therefore in future a longitudinal methodological approach may in fact be of greater benefit. The findings here in this regard are therefore meant as a mere contribution to the ongoing discussion at this point as opposed to a cast-iron model for best practice.

5.5. Future Research

Drawing on these limitations and the previously highlighted lack of research in this area, there is a wealth of further work to be undertaken. Both empirical and conceptual research is suggested not only into further exploration of LAL per se in different online and face-to-face EMI HE settings on a larger scale, but also to determine the impact of GenAI on domain-specific assessment and how this is addressed. The rapidly evolving nature of GenAI means that the academic community’s understanding of this potential impact must draw on cross-disciplinary developments in scholarship.

The conceptual revision of established tools used to measure LAL is also recommended. Even though this study has drawn upon some of the more recent validated models and instruments, bespoke adaptations have been made to accommodate the research focus. Thus, considering the magnitude of GenAI impact, and the understandable lack of teacher preparedness found, it is recommended that further consideration be given to how such instruments and models may be adapted, or perhaps further offerings developed to inform the work of future investigation to potentially boost theoretical credibility.

6. Conclusion

This exploratory study is a timely novel contribution to scholarship which draws on some of the most recent developments in the field of GenAI to empirically examine the potential impact these may have on online EMI HE and educator LAL. Having used a sequential explanatory mixed-methods design, the findings illustrate a present international LAL educator landscape which is markedly underprepared in terms of knowledge, experience, and skills of the different dimensions of LAL with limited exceptions to this. Furthermore, whilst demonstrating awareness of the potential impact on established assessment praxis, the results show a drastic shift towards a notable decrease in self-assessed knowledge of LAL in relation to Gen AI. This may indeed be as a result of the lack of understanding locally and institutionally in terms of ways and means of addressing the linguistic and authorial challenges posed by GenAI tools to academic integrity online EMI HE academic integrity.

As an ever-increasing number of GenAI tools continue to emerge, the international academic community is faced with the challenge of establishing expert consensus on GenAI tool usage in HE in scholarship and sharpening HE

institutional response in practice. Nevertheless, it is of the utmost importance that considered action is prioritized over hasty or impulsive responses which fail to address the complexity of the issue in full. As further inroads are made in GenAI tool capability, sight should not be lost of the potential benefits too, as they may be exploited by key stakeholders in assessment, for instance, the automation of certain assessment procedures as a means of cognitive off-loading.

From a pedagogical standpoint, the findings of this study have significant implications for the design and implementation of professional development programs for EMI educators. The results illustrate general tendencies of low levels of LAL among the participants, necessitating comprehensive initiatives aimed at enhancing educators' LAL and competence in GenAI technology usage. As highlighted by the qualitative data, there is a clear lack of understanding and unease towards GenAI's potential impact on academic integrity: "This is so much more than ChatGPT and we just can't keep up with what they can all do. Academic integrity for us in the field has never been more vulnerable and at times I feel like we are defenseless - I just don't know enough" (Interview 1128). Such initiatives should not only address the technical aspects of GenAI but also emphasize ethical considerations and strategies for maintaining academic integrity in the face of this emerging technology.

Furthermore, these findings underscore the importance of curriculum developers reevaluating assessment strategies in light of GenAI's capabilities. The quantitative results reveal that traditional assessment artifacts like essays are perceived as highly vulnerable to GenAI-assisted academic misconduct, with 66.09% of respondents rating themselves as only "very knowledgeable" in developing and administering language assessments. This aligns with concerns raised in the interviews about the lack of effective local detection tools and procedures: "I naively thought that it was game-over when I found a couple of sites that claimed to detect AI-generated text...But in practice, looking into it myself, they aren't as good as I thought. If a student really wants to, they could cheat these systems quite easily" (Interview 356). These findings add weight to calls for a shift towards alternative assessment approaches that leverage the strengths of GenAI to foster critical thinking and creativity among students. This could involve more dynamic, project-based assessments where the process of learning and the application of knowledge are evaluated in addition to the final output. Such assessments would not only mitigate the potential for academic dishonesty but also better align with the demands of a rapidly evolving technological landscape.

Educators may however eventually have to come to terms with the irreparable disruption to assessment established praxis, such as the latent susceptibility of certain traditional assessment artefacts, like the essay, to GenAI-assisted e-cheating. Nevertheless, it is imperative now and more than ever that HEIs do not acquiesce in the anthropic prioritization of present and future learner and educator needs. Only by upholding the core underlying principles which inform both assessment and more broadly academic culture, through the continuing adaptation of praxis and policy, resides the possibility of withstanding the test of GenAI.

Conflict of Interest

The author declares that he has no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Ethics Board Approval Statements

The research was carried out having obtained ethics committee approval from the UNIR Ethics Committee with reference PI020/2023.

Funding

This research has been financed by the Project of Analysis and Development for the Optimization of Assessment and Regulation of Generative Artificial Intelligence (PANDORA) Research Project with reference PP-2023-02 granted in the 2023 UNIR Research Projects Call of Universidad Internacional de La Rioja.

References

Adams, W. C. (2015). Conducting semi-structured interviews. In K. E. Newcomer, H. P. Hatry, & J. S. Wholey (Eds.), *Handbook of practical program evaluation* (pp. 492-505). Wiley. https://doi.org/10.1002/9781119171386.ch19

Afshar, H. S., & Ranjbar, N. (2021). EAP teachers' assessment literacy: From theory to practice. *Studies in Educational Evaluation*, *70*, 101042. https://doi.org/10.1016/j.stueduc.2021.101042

Amani, S., White, L., Balart, T., Arora, L., Shryock, K. J., Brumbelow, K., & Watson, K. L. (2023). *Generative AI perceptions: A survey to measure perceptions of faculty, staff, and students on generative AI tools in academia*. arXiv. https://doi.org/10.48550/arXiv.2304.14415

Andujar, A., & Spratt, M. (2023). Using AI to support CLIL teacher language. *Journal of Research in Applied Linguistics*, *14*(2), 7-19. https://doi.org/10.22055/rals.2023.45267.3177

Baidoo-Anu, D., & Owusu Ansah, L. (2023). Education in the era of generative artificial intelligence (AI): Understanding the potential benefits of ChatGPT in promoting teaching and learning. *SSRN*. http://dx.doi.org/10.2139/ssrn.4337484

Bannister, P., Santamaría Urbieta, A., & Alcalde Peñalver, E. (2023a). A Delphi study on generative AI and English medium instruction assessment: Implications for social justice. *Iranian Journal of Language Teaching Research*, *11*(3), 53-90. https://doi.org/10.30466/ijltr.2023.121406

Bannister, P., Santamaría Urbieta, A., & Alcalde Peñalver, E. (2023b). A systematic review of generative AI and (English medium instruction) higher education. *Aula Abierta*, *52*(4), 401-409. https://doi.org/10.17811/rifie.52.4.2023.401-409

Bañeres, D., Rodríguez-González, M. E., Guerrero-Roldán, A.-E., & Cortadas, P. (2023). An early warning system to identify and intervene online dropout learners. *International Journal of Educational Technology in Higher Education*, *20*(1), 1-25. https://doi.org/10.1186/s41239-022-00371-5

Battersby, M. (2019). Appeals to authority: Sources and experts. In J. A. Blair (Ed.), *Studies in critical thinking. Windsor in argumentation vol. 8* (pp. 289-305). Open Library.

Bearman, M. & Ajjawi, R. (2023). Learning to work with the black box: Pedagogy for a world with artificial intelligence. *British Journal of Educational Technology*, *54*(5), 1160-1173. https://doi.org/10.1111/bjet.13337

Bearman, M., Ryan, J., & Ajjawi, R. (2022). Discourses of artificial intelligence in higher education: a critical literature review. *Higher Education*, *86*, 369-385. https://doi.org/10.1007/s10734-022-00937-2

Birhane, A., Kasirzadeh, A., Leslie, D., & Wachter, S. (2023). Science in the age of large language models. *Nature Reviews Physics*, *5*(5), 277–280. https://doi.org/10.1038/s42254-023-00581-4

Boden, M. A. (2004). *The creative mind myths and mechanisms*. Routledge.

Bozkurt, A., & Sharma, R. C. (2023). Challenging the status quo and exploring the new boundaries in the age of algorithms: Reimagining the role of generative AI in distance education and online learning. *Asian Journal of Distance Education*, *18*(1), i-vii. https://doi.org/10.5281/zenodo.7755273

Bozkurt, A., Xiao, J., Lambert, S., Pazurek, A., Crompton, H., Koseoglu, S., Farrow, R., Bond, M., Nerantzi, C., Honeychurch, S., Bali, M., Dron, J., Mir, K., Stewart, B., Costello, E., Mason, J., Stracke, C., Romero-Hall, E., Koutropoulos, A., . . . Jandrić, P. (2023). Speculative futures on ChatGPT and generative artificial intelligence (AI): A collective reflection from the educational landscape. *Asian Journal of Distance Education*, *18*(1), 53-130. https://doi.org/10.5281/zenodo.7636568

Braun, V., & Clarke, V. (2023). Toward good practice in thematic analysis: Avoiding common problems and be(com)ing a knowing researcher. *International Journal of Transgender Health*, *24*(1), 1-6. https://doi.org/10.1080/26895269.2022.2129597

Braun, V., Clarke, V., Hayfield, N. & Terry, G. (2019). Thematic analysis. In P. Liamputtong (Ed.), *Handbook of research methods in social sciences* (pp. 843-860). Springer. https://doi.org/10.1007/978-981-10-5251-4_103

Brew, M., Taylor, S., Lam, R., Havemann, L., & Nerantzi, C. (2023). Towards developing AI literacy: Three student provocations on AI in higher education. *Asian Journal of Distance Education*, *18*(2), 1-11. https://doi.org/10.5281/zenodo.8032387

Carrigan, M. (2023, April 27). Are universities too slow to cope with Generative AI? *Impact of Social Sciences*. https://blogs.lse.ac.uk/impactofsocialsciences/2023/04/27/are-universities-to-slow-to-cope-with-generative-ai/

Chan, C. K. Y., & Lee, K. K. W. (2023). The AI generation gap: Are Gen Z students more interested in adopting generative AI such as ChatGPT in teaching and learning than their Gen X and Millennial Generation teachers? *Smart Learning Environments*, *10*(1), 1-23. https://doi.org/10.48550/arXiv.2305.02878

Chen, Y., Jensen, S., Albert, L. J., Gupta, S., & Lee, T. (2023). Artificial intelligence (AI) student assistants in the classroom: Designing chatbots to support student success. *Information Systems Frontiers*, *25*(1), 161-182. https://doi.org/10.1007/s10796-022-10291-4

Chen, Z. (2022). Artificial intelligence evaluation for mathematics teaching in colleges under the guidance of wireless network. *Mobile Information Systems*, *2022*, 1-9. https://doi.org/10.1155/2022/3201004

Creswell, J. W., & Plano Clark, V. L. (2011). *Designing and conducting mixed methods research*. SAGE Publications Ltd.

Dalalah, D., & Dalalah, O. M. A. (2023). The false positives and false negatives of generative AI detection tools in education and academic research: The case of ChatGPT. *The International Journal of Management Education*, *21*(2), 100822. https://doi.org/10.1016/j.ijme.2023.100822

Dawson, P. (2021). *Defending assessment security in a digital world: Preventing e-cheating and supporting academic integrity in higher education*. Routledge.

Dawson, P. (2020). Cognitive offloading and assessment. In M. Bearman, P. Dawson, R. Ajjawi, J. Tai, & D. Boud (Eds.), *Re-imagining university assessment in a digital world* (pp. 37-48). Springer. https://doi.org/10.1007/978-3-030-41956-1_4

Dearden, J. (2014). *English as a medium of instruction —a growing global phenomenon*. British Council. https://www.britishcouncil.org/sites/default/files/e484_emi_cover_option_3_final_web.pdf

Deng, C., Ji, X., Rainey, C., Zhang, J., & Lu, W. (2020). Integrating Machine Learning with Human Knowledge. *IScience*, *23*(11), 101656. https://doi.org/10.1016/j.isci.2020.101656

Derakhshan, A., & Shakki, F. (2024). Opportunities and challenges of implementing online English courses in Iranian public and private schools. *Journal of Research in Applied Linguistics*, *15*(1), 17-31. https://doi.org/10.22055/rals.2023.44418.3111

Dwivedi, Y. K., Kshetri, N., Hughes, L., Slade, E. L., Jeyaraj, A., Kar, A. K., Baabdullah, A. M., Koohang, A., Raghavan, V., Ahuja, M., Albanna, H., Albashrawi, M. A., Al-Busaidi, A. S., Balakrishnan, J., Barlette, Y., Basu, S., Bose, I., Brooks, L., Buhalis, D., & Carter, L. (2023). "So what if ChatGPT wrote it?" Multidisciplinary perspectives on opportunities, challenges and implications of generative conversational AI for research, practice and policy. *International Journal of Information Management*, *71*, 102642. https://doi.org/10.1016/j.ijinfomgt.2023.102642

Eager, B., & Brunton, R. (2023). Prompting higher education towards AI-augmented teaching and learning practice. *Journal of University Teaching & Learning Practice, 20*(5), 1-19. https://doi.org/10.53761/1.20.5.02

Eaton, S. E. (2021). *Plagiarism in higher education: Tackling tough topics in academic integrity*. Bloomsbury Publishing.

Eke, D. O. (2023). ChatGPT and the rise of generative AI: Threat to academic integrity? *Journal of Responsible Technology*, *13*, 100060. https://doi.org/10.1016/j.jrt.2023.100060

Elkhatat, A. M. (2023). Evaluating the authenticity of ChatGPT responses: A study on text-matching capabilities. *International Journal for Educational Integrity*, *19*(15), 1-23. https://doi.org/10.1007/s40979-023-00137-0

Farazouli, A., Cerratto-Pargman, T., Bolander-Laksov, K., & McGrath, C. (2023). Hello GPT! Goodbye home examination? An exploratory study of AI chatbots impact on university teachers' assessment practices. *Assessment & Evaluation in Higher Education*, *49*(3), 363-375. https://doi.org/10.1080/02602938.2023.2241676

Fergus, S., Botha, M., & Ostovar, M. (2023). Evaluating academic answers generated using ChatGPT. *Journal of Chemical Education*, *8*, 423-438. https://doi.org/10.1021/acs.jchemed.3c00087

Fido, D., & Wallace, L. (2023). The unique role of ChatGPT in closing the awarding gap. *The Interdisciplinary Journal of Student Success*. https://cdspress.ca/wp-content/uploads/2023/02/IJSS_FEB_2023_8_Final.pdf

Fulcher, G. (2012). Assessment literacy for the language classroom. *Assessment Quarterly*, *9*, 113-132. https://doi.org/10.1080/15434303.2011.642041

Gan, L., & Lam, R. A review on language assessment literacy: Trends, foci and contributions. *Language Assessment Quarterly*, *19*(5), 503-525. https://doi.org/10.1080/15434303.2022.2128802

Harding, L., Kremmel, B., & Erberharter, K. (2022). Language assessment literacy in second spoken language assessment contexts. In T. Huag, W. Mann, & U. Knoch (Eds.), *The handbook of language assessment across modalities* (pp. 373-382). Oxford University Press. https://doi.org/10.1093/oso/9780190885052.003.0032

Hultgren, A. K., Owen, N., Shrestha, P., Kuteeva, M., & Mežek, S. (2022). Assessment and English as a medium of instruction. Challenges and opportunities. *Journal of English Medium Instruction*, *1*(1), 105-123. https://doi.org/10.1075/jemi.21019.hul

Johinke, R., Cummings, R., & Di Lauro, F. (2023). Reclaiming the technology of higher education for teaching digital writing in a post—pandemic world. *Journal of University Teaching and Learning Practice*, *20*(2), 1-16. https://doi.org/10.53761/1.20.02.01

Koszowy, M., & Walton, D. (2019). Epistemic and deontic authority in the *argumentum ad verecundiam*. *Pragmatics and Society*, *10*(2), 287-315. https://doi.org/10.1075/ps.16051.kos

Kremmel, B., & Harding, L. (2020). Towards a comprehensive, empirical model of language assessment literacy across stakeholder groups: Developing the language assessment literacy survey. *Language Assessment Quarterly*, *17*(1), 100-120. https://doi.org/10.1080/15434303.2019.1674855

Lasagabaster, D. (2022). Teacher preparedness for English-medium instruction. *Journal of English-Medium Instruction*, *1*(1), 48-64. https://doi.org/10.1075/jemi.21011.las

Liang, W., Yusekgonul, M., Mao, Y., Wu, E., & Zou, J. (2023). GPT detectors are biased against non-native English writers. *Patterns*, *4*, 100779. https://doi.org/10.48550/arXiv.2304.02819

Liu, J. E., Lo Y. Y., & Xin, J. J. (2023). CLIL teacher assessment literacy: A scoping review. *Teaching and Teacher Education*, *129*, 104150. https://doi.org/10.1016/j.tate.2023.104150

Macaro, E. (2022). English medium instruction: What do we know so far and what do we still need to find out? *Language Teaching*, *55*, 533-546. https://doi.org/10.1017/S0261444822000052

Macaro, E. (2018). *English medium instruction: Content and language in policy and practice*. Oxford University Press.

Mancho-Barés, G., Khan, S., & Aguilar-Pérez, M. (2022). An EMI lecturer's assessment practices with engineering laboratory reports. *Journal of English-Medium Instruction*, *1*(2), 232-254. https://doi.org/10.1075/jemi.21008.man

McKinley, J., & Rose, H. (2022). English language teaching and English-medium instruction. *Journal of English-Medium Instruction*, *1*(1), 85–104. https://doi.org/10.1075/jemi.21026.mck

Nieminen, J. H., & Carless, D. (2023). Feedback literacy: A critical review of an emerging concept. *Higher Education*, *85*, 1381-1400. https://doi.org/10.1007/s10734-022-00895-9

Nikolic, S., Daniel, S., Haque, R., Belkina, M., Hassan, G. M., Grundy, S., Lyden, S., Neal, P., & Sandison, C. (2023). CHATGPT versus engineering education assessment: A multidisciplinary and multi-institutional benchmarking and analysis of this generative artificial intelligence tool to investigate assessment integrity. *European Journal of Engineering Education*, *48*(4), 559–614. https://doi.org/10.1080/03043797.2023.2213169

Pack, A., & Maloney, J. (2023). Using generative artificial intelligence for language education research: Insights from using OpenAI's ChatGPT. *TESOL Quarterly*, *57*(4), 1571-1582. https://doi.org/10.1002/tesq.3253

Rudolph, J., Tan, S., & Tan, S. (2023). ChatGPT: Bullshit spewer or the end of traditional assessments in higher education? *Journal of Applied Learning & Teaching*, *6*(1), 342-363. https://doi.org/10.37074/jalt.2023.6.1.9

Ruslin, R., Mashuri, S., Rasak, M. S. A., Alhabsyi, F., & Syam, H. (2022). Semi-structured interview: A methodological reflection on the development of a qualitative research instrument in educational studies. *Journal of Research & Method in Education*, *12*(1), 22-29. https://doi.org/10.9790/7388-1201052229

Russell, S. (2019). *Artificial intelligence and the problem of human control*. Viking.

Sadasivan, V. S., Kumar, A., Balasubramanian, S., Wang, W., & Feizi, S. (2023). *Can AI-generated text be reliably detected?* arXiv. https://doi.org/10.48550/arXiv.2303.11156

Sah, P. K. (2022). A research agenda for English-medium instruction. Conversations with scholars at the research fronts. *Journal of English-Medium Instruction*, *1*(1), 124-136. https://doi.org/10.1075/jemi.21022.sah

Scarino, A. (2013). Language assessment literacy as self-awareness: Understanding the role of interpretation in assessment and teacher learning. *Language Testing*, *30*(3), 309–327. https://doi.org/10.1177/0265532213480128

Shahzadi, A., & Ducasse, A. M. (2022). Language assessment literacy of teachers in an English medium of instruction university: Implications for ELT training in Pakistan. *Studies in Language Assessment*, *11*(1), 92-118. https://doi.org/10.58379/BZWF5085

Sharples, M. (2023). Towards social generative AI for education: Theory, practices and ethics. *Learning: Research and Practice*, *9*(2), 159-167. https://doi.org/10.48550/arXiv.2306.10063

Sharples. M. (2022). Automated essay writing: An AIED opinion. *International Journal of Artificial Intelligence in Education*, *32*, 1119-1126. https://doi.org/10.1007/s40593-022-00300-7

Sharples. M. & Pérez y Pérez, R. (2022). *Story machines: How computers have become creative writers*. Routledge.

Stobart, G. (2008). *Testing times: The uses and abuses of assessment*. Routledge.

Su, J., & Yang, W. (2023). Unlocking the power of ChatGPT: A framework for applying generative AI in education. *ECNU Review of Education*, *6*(3), 355-503. https://doi.org/10.1177/20965311231168423

Sudajit-apa, M. (in press). Dismantling the discursive representation of women in AI-generated life-changing narratives: A critical discourse analysis. *Journal of Research in Applied Linguistics*. https://doi.org/10.22055/rals.2024.46100.3232

Sullivan, M., Kelly, A., & McLaughlan, P. (2023). ChatGPT in higher education: Considerations for academic integrity and student learning. *Journal of Applied Learning & Teaching*, *6*(1), 1-10. https://doi.org/10.37074/jalt.2023.6.1.17

Taylor, L. (2013). Communicating the theory, practice and principles of language testing to test stakeholders: Some reflections. *Language Testing*, *30*, 403–412. https://doi.org/10.1177/0265532213480338

Tzirides, A. O., Saini, A., Zapata, G. Searsmith, D., Cope, B., Kalantzis, M., Castro, V., Kourkoulou, T., Jones, J., Abrantes da Silva, R., Whiting, J., & Polyxeni Kastania, N. (2023). Generative AI: Implications and applications for education. *arXiv*. https://doi.org/10.48550/arXiv.2305.07605

Tsagari, D. (2020). Language assessment literacy. Concepts, challenges and prospects. In S. Hidri (Ed.), *Perspectives on language assessment literacy: Challenges for improved student learning* (pp. 13-32). Taylor & Francis Books. https://doi.org/10.4324/9781003016083-2

Tsagari, D., & Vogt, K. (2017). Assessment literacy of language teachers around Europe: Research, challenges, and future prospects. *Papers in Language Testing and Assessment*, *6*(1), 41-63. https://doi.org/10.58379/UHIX9883

Vogt, K., & Tsagari, D. (2014). Assessment literacy of foreign language teachers: Findings of a European study. *Language Assessment Quarterly*, *11*(4), 374-402. https://doi.org/10.1080/15434303.2014.960046

Wang, T., Lund, B. D., Marengo, A., Pagano, A., Mannuru, N. R., Teel, Z. A., & Pange, J. (2023). Exploring the potential impact of artificial intelligence (AI) on international students in higher education: Generative AI, chatbots, analytics, and international student success. *Applied Sciences*, *13*(11), 6716. https://doi.org/10.3390/app13116716

Yildirim-Erbasli, S. N., & Bulut, O. (2023). Conversation-based assessment: A novel approach to boosting test-taking effort in digital formative assessment. *Computers and Education: Artificial Intelligence*, *4*, 100135. https://doi.org/10.1016/j.caeai.2023.100135

Yu, H., & Guo, Y. (2023). Generative artificial intelligence empowers educational reform: Current status, issues, and prospects. *Frontiers in Education*, *8*, 1183162. https://doi.org/10.3389/feduc.2023.1183162

Yufik, Y. M., Friston, K. J., & Moran, R. J. (2022). Editorial: Understanding in the human and the machine. *Frontiers in Systems Neuroscience*, *16*, 10811112. https://doi.org/10.3389/fnsys.2022.1081112

Zajko, M. (2022). Artificial Intelligence, algorithms, and social inequality: Sociological contributions to contemporary debates. *Sociology Compass*, *16*(3), e12962. https://doi.org/10.1111/soc4.12962