Language Teaching Research Quarterly **EUROKD**

2025 (52), 1–26
https://doi.org/10.32038/ltrq.2025.52.01

Rethinking Assessment Practices in English Language Teaching in the Age of Generative Artificial Intelligence

Hossein Vafadar
Canadian Institute for Knowledge Development, Vancouver, Canada
Batuhan Selvi*
Yozgat Bozok University, Yozgat, Türkiye
Ferzan Atay
Yozgat Bozok University, Yozgat, Türkiye

Correspondence
Email: batuhanselvi@outlook.com

Abstract
The introduction of Generative Artificial Intelligence (GenAI) has raised several issues in English teaching and assessment such as authorship, academic honesty, and pedagogical appropriateness. Although these tools can be used for feedback, task design, and support, their integration may also result in a number of ethical and instructional challenges. In this sense, this study focuses on how in-service English teachers in Türkiye recognize and respond to GenAI in the context of language assessment. In the study, four frameworks, Teacher Cognition, Language Assessment Literacy, Technological Pedagogical Content Knowledge (TPACK), and Critical Pedagogy, were employed to investigate beliefs, attitudes, and experiences of the participants. Sixteen in-service teachers from different contexts took part in semi-structured interviews. The analysis identified four themes: ethical reasoning and teacher agency, changing conceptions of assessment literacy, technological competence and institutional uncertainty, and equity issues regarding access and student readiness. While participants reported that they used GenAI tools, they also raised some concerns. The participants also stressed lack of clear institutional guidance and training. In addition, the participants reported that they developed alternative assessment practices to reduce AI misuse. This study reveals a teacher-centered, theoretically grounded description of GenAI integration in under-researched educational settings.

ARTICLE HISTORY
Received: 02 June 2025
Revised: 25 October 2025
Accepted: 16 November 2025

KEYWORDS
Generative Artificial Intelligence, Language Assessment, Teacher Cognition, Critical Pedagogy, Technological Pedagogical Content Knowledge

How to cite this article (APA 7th Edition):
Vafadar, H., Selvi, B., & Atay, F. (2025). Rethinking assessment practices in English language teaching in the age of generative artificial intelligence. *Language Teaching Research Quarterly*, *52*, 1–26. https://doi.org/10.32038/ltrq.2025.52.01

Introduction

The introduction and promotion of Generative Artificial Intelligence (GenAI) technology has revolutionized the educational practices. In the context of English Language Teaching (ELT), it provides new opportunities for learners to improve their language proficiency, receive intant feedback, and practice language outside school environment. However, their ease of access and popularity have challenged well-established conventions such as academic honesty and authorship (Guo & Wang, 2025; Mizumoto & Eguchi, 2023). Such issues are especially evident in language assessment since teachers are now required to reassess both how and whom they evaluate. In this sense, recent studies have examined the effect of GenAI on student engagement (Yılmaz-Virlan & Tomak, 2025), teacher attitudes (Moorhouse & Kohnke, 2024), and the validity problems due to AI-generated texts (Casal & Kessler, 2023). However, most of the studies in the literature use a single-theoretical perspective, which limits their capacity to address the problem adequately and focus largely on learners or tools. In such an academic and educational setting, language teachers are expected to integrate GenAI into their assessment practices without sufficient institutional guidance, training, and ethical frameworks (Kohnke et al., 2023; Liu, 2025). This requirement results in a multifaceted and underexamined set of cognitive, ethical, and pedagogical issues.

In this context, this study tries to address this gap by examining how ELT teachers perceive, experience, and respond to GenAI in assessment contexts. Instead of considering GenAI as a neutral instructional tool, this study examines whether its use challenges beliefs, responsibilities, and practices of teachers. Therefore, we used a multi-framework approach to examine how language teachers integrate GenAI into their assessment practice:

1. Teacher Cognition Theory (Borg, 2003) examines teachers' beliefs, knowledge, and thought processes that shape their decisions about GenAI use in assessment contexts.
2. Assessment Literacy (Fulcher, 2012) analyzes teachers' knowledge and competencies in designing, implementing, and evaluating assessments in the GenAI era.
3. Technological Pedagogical Content Knowledge/TPACK (Mishra & Koehler, 2006) investigates how teachers integrate GenAI tools with pedagogical strategies and language content knowledge.
4. Critical Pedagogy (Freire, 1970) explores ethical dimensions of GenAI use in language assessment, including issues of equity, power, and academic integrity.

We used such an integrated approach since GenAI is not merely a new technological tool but has led to a fundamental shift in assessment practices of teachers. Effectively response to GenAI requires teachers to develop high levels of digital competence, ethical awareness, pedagogical reflexivity, and institutional support. (Celik, 2023; Zhao et al., 2024). By integrating these frameworks, we focus on the cognitive, professional, technological, and ethical dimensions of GenAI use in language assessment. In this sense, this study seeks to answer the following research questions:

RQ$_1$: How do in-service English language teachers perceive the role of GenAI in language assessment?
RQ$_2$: What ethical and pedagogical challenges do in-service English language teachers associate with the use of GenAI in assessment?
RQ$_3$: How do institutional contexts shape in-service English language teachers' adoption or resistance to GenAI in their assessment practices?
RQ$_4$: In what ways do issues of access and equity influence the implementation of GenAI-supported assessment in English language classrooms?

Literature Review

Teacher Cognition: Beliefs and Decisions about GenAI in Assessment

Teacher cognition refers to the beliefs, knowledge, and back ground which guide teachers in their instructional decisions (Borg, 2019). Therefore, it plays a significant role in how teachers interpret and apply new technologies in their everyday practice. According to Borg (2003), teacher cognition is a dynamic construct and includes teachers "think, know, and believe" in that it is shaped by personal, institutional, and sociocultural features. In the context of GenAI, teacher cognition is a more than just a tool adoption issue. In fact, it mediates ethical thinking, perceived validity, and the restructuring of agency between teacher, learner, and the tool. Recent studies show that GenAI have led to both interest and hesitation among teachers. For example, Bower et al. (2024) reported that GenAI paved the way for university instructors to rethink their assessment practices and to focus on critical thinking and more creativity tasks which are more prone to AI use. Similarly, Kizilcec et al. (2024) argued that students will have access to GenAI in the future and thus assessments practices should be redesigned and the teachers should act as implementers of GenAI instead of gatekeepers. Similarly, Barrot (2024), highlighted that GenAI tools such as ChatGPT can assist writing instruction at all levels. Nevertheless, although these studies demand new roles for teachers, they offer limited understanding regarding how teachers react to these changes with their existing values and limitations.

In addition, The potential of GenAI leads to significant ethical and practical challenges. For instance, Liu (2025) reported that GenAI is associated with plagiarism and authorship issues. Similarly, Farazouli et al. (2024) found that teachers tend to grade papers produced by ChatGPT higher than those written by students. These findings reveal a contradiction: teachers acknowledge that GenAI produces high-quality work while simultaneously questioning whether students should be allowed to use it. This paradox is even worsened due to a clear lack of guidance. In this context, Liu (2025) and Shahid et al. (2024) noted that when there is no formal institutional framework, teachers are required to establish their own ethical norms and pedagogical strategies. Yet little is known about how this "ethical improvisation" is experienced in different disciplines and institutions. For example, Kerneža and Zemljak (2023) found that science teachers were less motivated than social science teachers to integrate GenAI in assessment contexts.

Such a finding reveals that subject-specific norms may play a role in teachers' cognitive stances. However, the mechanisms behind these differences is still underexplored.

In Türkiye, this conflict is similarly profound. Yılmaz-Virlan and Tomak (2025) reported that although ELT instructors valued GenAI's potential for formative feedback, they were concerned that it may weaken learner autonomy and produce language beyond above their proficiency. Ulum (2024) also draws attention to age related differences. Whereas pre-service teachers were more open to use GenAI, in-service teachers were more cautious due to concerns regarding academic honesty. These findings are in line with Li's (2020) argument that cognition is socially negotiated across different contexts. Moreover, teacher cognition is also related to strategic behavior. For example, Arıkan and Peçenek (2024) identified "quiet adoption" patterns. They found that although teachers used GenAI tools in their own professional work, they did not declare it in public. This finding supports Tian and Wang's (2025) view that teachers' GenAI attitudes are shaped by the relationship of assessment literacy, digital competence, and institutional support. When they align, teachers typically use GenAI; otherwise, they tend to reject or use it minimally.

In summary, teacher cognition acts both as a lens and a filter through which GenAI is evaluated. It also functions as a form of agency and guides teachers on how to respond to the new technologies. While existing studies have identified general attitudes and policy gaps, few of them have focused on how teacher cognition may play a role in ethical conflicts, disciplinary values, and institutional quietness. By examining how Turkish ELT teachers regard GenAI's role in language assessment, this study contributes to an understanding of teacher cognition in the age of artificial intelligence.

Assessment Literacy: Adapting to Technology in Language Assessment

Language Assessment Literacy (LAL) broadly refers to knowledge, competence, and ethical understanding in assessment and evaluation design and implementation (Fulcher, 2012; Inbar-Lourie, 2008; Taylor, 2013). In the context of LAL, teachers have authority and control over assessment practices. However, GenAI has also challenged this framework. GenAI poses threat to the understanding of valid assessment in terms of technical and ethical issues as it is not clear whether teachers are assessing learners or machine-generated output. Therefore, this shift has paved the way for a reconsideration of pedagogical assumptions as well as GenAI tools. In this context, a number of studies emphasized the difficulty of assessing learner performance when learners receive AI assistance. For example, Suherman (2022) and Levi and Inbar-Lourie (2020) argued that although AI-produced output is linguistically polished, it may portray learners' proficiency and ability inaccurately. In addition, Mizumoto and Eguchi (2023) reported that that ChatGPT rated L2 writing tasks nearly same as human raters. This finding leads to the question whether the complex linguistic production of GenAI tools misguides teachers regarding language competence of learners. Koraishi (2024)'s finding that

ChatGPT's scoring of IELTS essays closely matched human raters, further demonstrates this worry and raises important concerns about whether teachers are evaluating students' actual abilities or the linguistic sophistication of AI tools.

These findings indicate a significant reconsideration of validity issues in assessment theory. Although most of the studies demand a shift in the understanding of validity in the age of GenAI, few have addressed how teachers react and respond to these conflicts in their everyday classroom assessment practices. To respond these problems, new models combining digital competence, ethical understanding, and flexible task design are required (Coombe et al., 2020; Banitalebi et al., 2025). However, most studies fail to investigate how teachers understand or implement these principles in the age of GenAI without clear institutional guides and technological infrastructure.

The studies in the literature have identified a gap between theoretical advances in LAL and teachers' readiness to apply them regarding GenAI. For instance, Moorhouse and Kohnke (2024) found that even teacher educators had limited confidence in ethical use of GenAI. Similarly, according to Yılmaz-Virlan and Tomak (2025), instructors in Türkiye questioned whether AI-supported student writing showed real learner competence. These findings imply a shift in assessment literacy, that is, we should now focus more on ethics, context, and pedagogical consistency in language assessment. However, most of these studies fail to address how teachers made decisions when there is no clear policy regarding GenAI use. This is even more problematic in settings where training and institutional support is not provided. In this sense, Liu (2025) reported that EFL teachers in China frequently resorted to informal practices or plagiarism detectors to respond to the challenges brought by GenAI use. In short, an only technical view of LAL is outdated in the age of AI. The field must provide solutions to pedagogical and ethical complexities issues and redefine assessment practices. Now students have widespread access to GenAI tools in their assignments and tasks and thus teachers are progressively more required to evaluate not only the product but also the process and authenticity of the tasks and assignments. This change brings issues of authorship, fairness, and agency into the agenda of teachers and teacher educators although they have been ignored thus far. In this context, this study contributes to this gap by examining how Turkish in-service ELT teachers experience and respond to the effects of GenAI in the context language assessment.

Technological Pedagogical Content Knowledge (TPACK) in the Age of GenAI

The Technological Pedagogical Content Knowledge (TPACK) (Mishra & Koehler, 2006) explains technology integration in teaching as a dynamic interaction between three spheres: technological knowledge (TK), pedagogical knowledge (PK), and content knowledge (CK). TPACK was originally developed to a framework for teachers to plan and implement their instructional practices in digitally rich environments. Over time it has become a key framework to illuminate how educators can integrate technological

innovation with the curriculum. However, the rise of GenAI has paved the way to several issues. For example, teachers are now required not only to use GenAI tools but to assess their appropriateness and ethical implications. In such an environment, TPACK needs to be redefined in order to provide solution to these issues. Recent studies have revealed a gap between teachers' awareness of GenAI and their ability to integrate it. For example, Arslan (2025) found that Turkish ELT instructors had moderate levels of confidence in using GenAI for assessment. However, studies in TPACK focus mostly on teachers' technological competence. This leads to another problem in understanding how they react to the pedagogical and ethical dimensions of GenAI. As a result, several professional development applications such as structured workshops (Koraishi, 2025) and design-based chatbot training (Choi et al., 2025) have been suggested to overcome this problem. However, such attempts are still local and unsystematic. These limitations are particularly evident in under-resourced contexts. The studies in the literature have revealed that GenAI tools are still used in a limited way or at the surface level in classroom practice. Seo (2024), for instance, found that students primarily utilized ChatGPT for superficial language corrections in GenAI-assisted narrative writing tasks rather than cultivating more profound critical thinking or narrative structure, indicating a pedagogical depth gap. Similarly, Ong and Annamalai (2024) found that pre-service ESL teachers mostly used AI to produce basic teaching materials such as slide shows or grammar activities instead of using it to design student centered or higher level activities. This raises another question. When teachers are not provided training and support, they may use GenAI in a way that maintain traditional practices instead of integrating it effectively. Therefore, as the educational use of GenAI becomes more and more widespread, TPACK should also focus on issues such as ethical awareness, pedagogical adaptability, and contextual responsiveness. In other words, teachers should evaluate the pedagogical suitability and anticipate consequences of GenAI use, and tailor the tools to address student needs.

Ethics and Critical Pedagogy in ELT Assessment with GenAI

GenAI's rapid integration into education forces English language teachers to rethink their ethical assumptions about assessment. Following Freire's (1970) view of education as liberating rather than controlling, critical pedagogy encourages educators to question not only what they assess but how and why they assess it. This questioning becomes urgent in GenAI-enhanced classrooms, where traditional categories—such as human versus machine authorship and original versus plagiarized work—are no longer clear-cut. When students use tools like ChatGPT or Quillbot to collaborate on their assignments, the boundaries between these categories blur, challenging conventional notions of academic integrity and authorship. In this context, assessment practices that focus on detecting and punishing risk are becoming ethically inadequate and pedagogically outdated. Instead, a critical pedagogical perspective urges educators to foreground ethical literacy, promoting open dialogue about the use of AI, authorship, and academic integrity. This shift—from surveillance to reflection—echoes Freirean commitments to dialogue,

learner agency, and critical consciousness. Lynch (2001) and Wiseman (2024) emphasize the importance of such reflective practices in technology-enhanced learning environments, where notions of power, access, and authorship are continually renegotiated. However, while ethical debates are growing in the literature, relatively few studies investigate how teachers operationalize these ideals in day-to-day assessment practices, particularly in under-resourced or policy-ambiguous contexts.

In Türkiye, these tensions are increasingly visible. Celik (2023) notes that many ELT instructors report confusion among students regarding the ethical boundaries of GenAI use, exacerbated by the lack of formal institutional guidelines. Beyond the classroom, GenAI also highlights structural inequities. The digital divide—between rural and urban learners or across socioeconomic groups—creates disparities in students' access to high-quality AI tools and internet infrastructure (Tajeddin et al., 2022). Therefore, ethical use of AI tools necessitates careful design to avoid reinforcing biases or excluding students with limited access to technology (Klimova & Chen,2024). To provide solution to such possibilities, more inclusive and ethics-oriented strategies for GenAI integration have been proposed. For example, Wiseman (2024) argues that more accessible designs where free tools are used instead of paid services should be preferred and the use of GenAI should be extended to classroom discussion. However, these practices are often limited to individual attempts. Therefore, more inclusive integration of GenAI is susceptible to teacher motivation and competence due to lack of institutional obligation.

The use of GenAI in assessment contexts implies a shift from traditional understandings of academic honesty to a more ethically oriented and socially responsive one. Now, the issues such as plagiarism, authorship, fairness, and access to GenAI are basically pedagogical and ideological instead of just technical or policy-based. In the context of critical pedagogy, ELT teachers are considered not just as practitioners of technology but as responsible actors influencing how language assessment changes in a progressively more AI-oriented context. These brief overview show that GenAI integration cannot be limited to a technical or policy matter. It requires ethical, contextual, and pedagogical awareness. Broader reviews (e.g., Al-khresheh, 2024; Al-Kadi & Ali, 2024) similarly argue that while GenAI tools offer learning opportunities, their responsible use depends on teacher competence, institutional support, and inclusive strategies.

Method

Research Design

A qualitative exploratory design was used in this study to examine Turkish in-service ELT teachers' perceptions and responses to the integration of GenAI in language assessment. A qualitative design was employed since it allowed us to obtain in-depth and contextually grounded insgihts (Creswell & Poth, 2016). Due to the ever-changing nature of GenAI and the absence of clear institutional guidelines, an exploratory approach also allowed us to get a more comprehensive perspective. In addition, four connected theoretical

frameworks were used in the study. Teacher Cognition Theory (Borg, 2003) helped us examine the effects of teachers' beliefs, experiences, and values regarding GenAI on the assessment contexts. Assessment Literacy (Fulcher, 2012) was used to examine teachers' understandings of validity, fairness and authorship in AI-assisted contexts. The TPACK model (Mishra & Koehler, 2006) provided us a perspective to investigate the integration GenAI in classroom practices with its opportunities and limitations. Finally, Critical Pedagogy (Freire, 1970) provided a sociopolitical and ethical perspective to examine learner agency, access equity and the ideological implications of GenAI use. Taken together, they allowed use to conduct a multi-dimensional analysis, that is cognitive, pedagogical, technological, and ethical, to address the research questions.

Participants

This study involved 16 in-service English language teachers. They were selected using purposeful maximum variation sampling to ensure rich representation across teaching levels, school types, regional settings, and levels of GenAI use. We deliberately chose a varied group of English teachers in Türkiye. Teachers needed to be currently working at middle school, high school, or university level and have at least basic knowledge of GenAI. This diverse sample allowed us to both describe different experiences and compare patterns across different types of teachers, which fits with their use of four different theoretical frameworks. While extensive GenAI experience was not required, participants needed basic familiarity to reflect meaningfully on its perceived affordances, risks, and ethical implications in assessment. Teachers with no exposure to GenAI or not actively teaching during the study period were excluded. The final sample size of 16 was determined using the principle of information power (Malterud et al., 2016), which suggests that smaller samples are sufficient when the aim is conceptually rich, context-specific insight. In this study, the accounts yielded analytically dense data on themes such as ethical reasoning, assessment redesign, institutional policy, and digital equity. Vasileiou et al. (2018) similarly note that saturation in interview-based research is commonly achieved with 12–24 participants, depending on topic specificity and sample heterogeneity. This sampling strategy was not intended for statistical generalization but for analytical transferability—to uncover patterns of professional reasoning, dilemmas, and adaptation among ELT instructors navigating GenAI-supported assessment. All participants provided informed consent, and pseudonyms were used to ensure anonymity in reporting. Table 1 presents demographic and professional characteristics of the participants, reflecting the diversity that informed the cross-case analysis.

Table 1

Demographics of the Participants

Participant ID	Gender	Teaching Experience	Institution Type	Region
P1	Female	10+ years	Public	Urban
P2	Male	5 years	Private	Urban
P3	Female	12 years	Public	Rural
P4	Male	3 years	Private	Urban
P5	Female	15 years	Public	Urban
P6	Male	8 years	Public	Rural
P7	Female	20+ years	Private	Urban
P8	Male	2 years	Public	Rural
P9	Female	6 years	Public	Urban
P10	Male	9 years	Private	Urban
P11	Female	7 years	Public	Rural
P12	Female	4 years	Private	Urban
P13	Male	10+ years	Public	Urban
P14	Female	3 years	Public	Rural
P15	Male	11 years	Private	Urban
P16	Female	5 years	Public	Urban

Data Collection Tool

In the study, semi-structured interviews were used to examine the participants' beliefs, practices, and ethical considerations regarding GenAI use in language assessment. Semi structured interview was chosen for its ability to obtain narrative, reflective, and contextually grounded responses while maintaining thematic consistency across participants (Kvale & Brinkmann, 2015). The interview guide was developed through based on our multi-dimensional framework. First, key conceptual issues, that are assessment design and validity, ethical reasoning, digital integration, and access equity, were identified through a detailed review of recent literature (e.g., Mizumoto & Eguchi, 2023; Liu, 2025; Yılmaz-Virlan & Tomak, 2025; Chapman et al., 2024). Then, each domain was associated with an equivalent framework. For instance, digital integration was associated with TPACK, and access equity with critical pedagogy. The first version of the interview guide consisted of 16 open-ended questions and was reviewed by two experts, one in language assessment and one in ELT teacher education, for conceptual alignment, clarity, and contextual appropriateness. Based on their feedback, the interview questions were revised. The interview guide was then piloted with two in-service teachers not included in the main sample. After the piloting, minor modifications such as simplified terminology and reordered questions were applied and the interview guide was finalized.

Data Collection and Analysis

Interviews were conducted via Google Meet between April and May 2025. To protect participant privacy, only audio was recorded, and informed consent was obtained prior to each session. Interviews lasted 30 to 60 minutes and were scheduled at participants' convenience. Each session began with brief demographic questions, followed by semi-structured prompts exploring the use of GenAI in assessment, its perceived benefits and

risks, pedagogical strategies, and institutional factors—framed by the study's four theoretical lenses. All interviews were transcribed verbatim, anonymized, and participants were assured of confidentiality and the right to withdraw at any stage. Thematic analysis was conducted following Braun and Clarke's (2006) six-phase framework. Due to licensing constraints, coding was performed manually in two cycles: an inductive phase to capture emergent insights, followed by a deductive phase guided by the conceptual framework. This dual approach enabled both data-driven and theory-informed interpretation. Codes were refined into four overarching themes: (1) Teacher Cognition and Ethical Reasoning, (2) Assessment Literacy and Pedagogical Practice, (3) Technological Readiness and Institutional Capacity, and (4) Digital Equity and Structural Inclusion. Illustrative quotes were selected to foreground participants' voices. To enhance trustworthiness, the primary researcher maintained analytic memos and coding logs (Miles et al., 2014). Peer debriefing with a second qualitative researcher ensured credibility through external review of selected transcripts and codes (Lincoln & Guba, 1985). While formal member checking was not implemented, participants were invited to clarify or expand on responses during interviews, supporting in-interview validation (Carlson, 2010). Researcher reflexivity was maintained throughout, with attention to how professional background and familiarity with GenAI could shape interpretation (Berger, 2015). Ethical approval was granted by Yozgat Bozok University's Ethics Committee. All identifying details were removed from transcripts and analysis files, and pseudonyms were used in all reporting to ensure anonymity and respectful representation of participant perspectives.

Results

Teacher Cognition and Ethical Reasoning

This theme explores participants' ethical reflections on GenAI use in assessment. The investigation revealed that teachers responded to GenAI use in the contexts of their responsibilities, learner agency, and academic honesty. As summarized in Table 2, participants described drawing moral boundaries between productive support and inappropriate substitution, reflecting principles from teacher cognition theory and critical pedagogical ethics.

It was found that the participants drew a clear line between using GenAI to support learner work and using it to replace learner contributions. They told that even though the use of GenAI for brainstorming and error correction is acceptable, content generation should be prohibited. For example, P1 said that "*If they use ChatGPT to do all of the assignment it is plagiarism*." In addition, P6 expressed that "*Students should think first and then use AI not vice versa.*" These statements show that the participants had three main concerns: learner effort, learner production, and accountability.

The participants also questioned the issue academic honesty. In this sense, P4 stated that "*Students are not aware of plagiarism while using AI tools*." Similarly, P9 stated that "*Many*

students think that they can use AI in assignments." These statements reveal a need to reconsider ethical literacy in AI-assisted learning environments.

Table 2

Teacher Cognition and Ethical Reasoning

Theme	Sub-Theme	Code	Participants	Description
Teacher Cognition and Ethical Reasoning	Beliefs about Student Use of GenAI	Perceived Ethical Boundaries	P1, P2, P5, P6, P11	Participants set clear ethical lines between appropriate and inappropriate uses of GenAI.
		Instrumental vs Substitutive Use	P3, P6, P8	Distinction made between AI as a supportive tool vs a full replacement for student effort.
		Academic Integrity Risk	P1, P3, P4, P9	Concerns raised about students submitting AI-generated work as their own.
		Shared Authorship Perception	P5, P7, P12	Participants note the ambiguity of student authorship when using GenAI tools.
	Teacher Moral Reasoning	Ethical Stewardship	P3, P5, P10	Participants view themselves as moral guides in modeling responsible technology use.
		Promotion of Critical Engagement	P1, P4, P6, P9	Emphasis on helping students think critically rather than rely passively on AI.
		Learner Accountability	P2, P8	Participants expect students to take responsibility for their learning outputs.
		Individual Ethical Frameworks	P3, P4	Each teacher brings personal values that influence their stance on GenAI use.
	Vision for AI in Education	Perceived Inevitability of GenAI	P2, P11, P12	GenAI is seen as an inevitable component of future teaching and assessment.
		Conditional Optimism	P5, P8, P11	Optimism expressed about GenAI's potential, if guided and regulated well.
		Cautious Adoption Stance	P4, P7, P10	Participants emphasize the need for selective and thoughtful AI integration.

Teachers also expressed their own ethical responsibilities. Some of them regarded themselves as gatekeepers of responsible GenAI use. For example, P10 stated that "*It is my job to teach them what they can do with AI*". Similarly, P1 emphasized critical engagement: "*I use AI to get ideas not to think for me*." This change from implementation of the rules to value-based guidance indicates the impact of critical reflection on instructional practice.

Although they were mainly careful about GenAI, the participants also expressed some kind of positivity. In this sense, they considered GenAI as a tool which should be integrated considerately. For example, P12 said that "*AI will not disappear. We should learn to use it responsibly.*" In another vein, some participants such as P7 stressed the danger of overreliance: "*The students should be able to write and think without using AI tools.*"

In sum, it was found that the participants did not passively respond to GenAI. Instead, they involved in ethical reasoning associated with their professional identities. Their statements show that they are trying to adjust innovation with basic values such as authenticity, justice, and critical thinking.

Assessment Literacy and Pedagogical Practice

This theme focuses on participants' use of GenAI in assessment design and evaluation. As shown in Table 3, the findings indicate a shift in assessment literacy. It was found that teachers are tailoring their practices to respond to the opportunities and challenges of GenAI.

Table 3

Assessment Literacy and Pedagogical Practices

Theme	Sub-Theme	Code	Participants	Description
Assessment Literacy and Pedagogical Practice	GenAI Integration into Assessment Tasks	AI-Assisted Item Generation	P2, P5, P6	Participants use GenAI to generate quizzes, comprehension questions, or test banks.
		AI-Enhanced Feedback Practices	P3, P7, P9, P12	GenAI supports formative feedback, especially in writing and language practice.
		Efficiency-Oriented Use of GenAI	P1, P4, P11	Some rely on GenAI to reduce prep time or automate repetitive assessment tasks.
	Shifts in Assessment Principles	Redefining Validity and Reliability	P2, P5, P10	Participants express that traditional notions of validity are being rethought.
		Formative vs Summative Rebalancing	P4, P6, P11	Participants shift focus from product-based to process-oriented assessment due to AI.
		Authenticity in AI Contexts	P1, P3, P9	Authenticity is emphasized more to ensure students produce original work.
	Assessment Design Challenges	AI-Resilient Task Design	P7, P8, P12	Participants explore task formats that minimize AI misuse e.g., oral or live writing.
		Concerns Over Assessment Transparency	P6, P10	Concerns arise about whether students understand how they're being evaluated with AI.
		Evaluation Uncertainty	P3, P4, P7	Uncertainty remains about how to judge AI-influenced work fairly and consistently.

Many participants reported using GenAI to develop assessment materials and accelerate feedback, particularly in writing tasks. These applications were praised for saving time and expanding pedagogical creativity. P6 noted, "*I often use ChatGPT to create multiple-choice and short-answer questions. It gives me more variety.*" Likewise, P5 highlighted GenAI's ideation potential: "*Sometimes I ask GenAI to suggest vocabulary quiz items. It gives me ideas I would not have thought of.*" In terms of feedback, GenAI served as a useful assistant for several teachers. In this sense, P9 stated that "*AI helped me give quick feedback on writing assignments. I manually check it later but it is saves time.*" Similarly,

P12 mentioned using AI to make it more efficient: "*I paste student writing and ask AI to highlight errors. Then I write comments. It quickens the feedback process.*" These statements shows that GenAI functions more than a mere automatic assessment tool. It has a formative potential with an efficient workflow and quick feedback.

However, the use of GenAI also required participants to reevaluate key issues such as validity and authenticity. In this sense, some of the participants questioned the legitimacy of traditional understanding of test validity in the age of GenAI. For example, P2 stated that "*What we consider 'valid' is changing. Students are getting high scores using GenAI tools. We need to reconsider our assessment ways.*" Similarly, P10 expressed: "*A good grade does not always require original thinking. We should revise our standards.*" In this vein, it was also found that some participants had already revised their assessment practices and focused more on drafts and in-class activities. For example, P4 expressed that "*I have changed the exams due to AI. I now take account of in-class activities more.*" Similarly, P11 reported that s/he had implemented continuous assessment by stating that "*I now prefer ongoing in-class assignments to eliminate threats of AI.*" Such practices show an attempt to adopt a more process-oriented and AI-resistant assessment methods.

The participants frequently mentioned authenticity as an important protective measure. For example, in-class writing was preferred by the participants. P3 express it as: "*I require in-class writing. Only in this way, I can be sure that they do it themselves.*" In addition, some participants used multimodal assessment methods such as oral exams and live writing tasks. In this sense, P8 stated that "*I have started to use oral exams and live writing tasks. AI cannot perform this kind of tasks.*" Similarly, P7 reported that "*I use in-class assignments and ask oral follow-ups. Students cannot use AI in spoken tasks.*" However, although the participants individually responded to GenAI use, they criticized the lack of clear GenAI use policies. For example, P6 stated that "*I think students are not aware of what is AI support or cheating. We all need clear instructions.*" In addition, P10 said that "*Students sometimes ask: 'Can we use AI here' This shows that we did not make our expectations clear*." Finally, participants mentioned the difficulty of assessing student production co-authored with AI. In this sense, P7 said that "*Sometimes I am unsure how much students use AI in their assignments. It is difficult to make a fair evaluation.*" P4 mentioned the same issue by saying that "*Even when the assignment is good, I am not sure how much belongs to students.*"

In summary, although the participants used GenAI to facilitate their assessment practices and provide feedback, they were also required to reconsider the philosophical foundations of their practices. The findings reveal both adaptive innovation and institutional gaps. Teachers are actively negotiating tensions between pedagogical authenticity, assessment fairness, and technological affordances—all without clear policy frameworks to guide them. In this regard, their practices reflect not only emerging forms

of assessment literacy but also the ethical and epistemological demands of teaching in a GenAI-integrated world.

Technological Readiness and Institutional Capacity

This theme examines participants' digital confidence, professional learning pathways, and the broader institutional landscape that shapes GenAI integration into assessment. As presented in Table 4, the findings reveal disparities in self-efficacy, a reliance on peer-based and exploratory learning, and a notable absence of formal training and institutional policy.

Table 4

Technological Readiness and Institutional Capacity

Theme	Sub-Theme	Code (Refined Name)	Participants	Description
Technological Readiness and Institutional Capacity	Teacher Self-Efficacy and Confidence	Digital Fluency and Familiarity	P2, P5, P11	Participants with stronger tech backgrounds expressed greater ease in using GenAI tools.
		Confidence Gaps and Hesitancy	P1, P4, P9	Some participants voiced insecurity or lack of confidence in their AI-related abilities.
		Exploratory Practice	P3, P6, P8	Participants described learning GenAI tools through experimentation and informal testing.
	Professional Development and Informal Learning	Peer Learning and Collaboration	P5, P6, P7	Collaboration with colleagues was cited as a primary mode of GenAI skill development.
		Lack of Structured Training	P4, P8, P10	Participants emphasized the lack of formal training in AI integration for education.
		Self-Directed Skill Development	P3, P12	Many participants taught themselves prompting, editing, and adapting GenAI output.
	Institutional Policy and Support Structures	Policy Absence and Uncertainty	P1, P4, P7	Lack of institutional AI policies leads to uncertainty and inconsistent practices.
		Call for Clear Guidelines	P2, P9, P10	Participants expressed a strong need for formal school/university guidelines on GenAI.
		Administrative Attitudes and Institutional Readiness	P6, P11, P12	Perceived administrative support varied, influencing adoption and experimentation.

Digital fluency was mentioned as a key issue influencing confidence in implementing GenAI tools. Participants who had experience with technological tools expressed reported that they are more comfortable and open to use GenAI. For instance, P2 stated that "*I am used to digital tools, so using GenAI was not a big deal for me*". In contrast, participants without prior experience exhibited unwillingness and uncertainty. For instance, P4 stated that "*I still hesitate. I do not feel confident using GenAI except for some basic tasks*".

Nevertheless, it seems that exploratory practice may play a role in bridging these gaps. Some participants stated that they learned how to use GenAI through trial and error. For example, P3 staetd that "*I started by giving prompts to AI tool and discover what happens*". Such a statement show that teachers used GenAI experimentally even though a systematic training was not provided. This experimental alignment was even shared among the colleagues. P5 explain this by saying "*We share what we learn in the teachers' room*.". Nevertheless, this social support was not enough for some participants who required systemic support. In this sense, P10 said that "*We have not had a single training or workshop on how to use GenAI. It is all self-learning*.". In addition, P4 said that "*We are dealing with this on our own and blindly without a support or training.*" In such settings, participants resorted to YouTube videos, peer explanations, and even AI-generated feedback to improve their digital literacy.

Institutional support, or its lack in our context, appeared as a significant limitation. Many participants reported that lack of institutional guidance in GenAI led to inconsistent standards and uncertainty. For example, P1 said that "*No one told us what we can or cannot do with AI.*". In addition, P7 stated that "*Everyone does something different.*" The participants expressed a crucial need for clear guidelines and support. In this sense, P2 expressed that "*We need clear policies. We need to know what is allowed and what is not*". Similarly, P9 stated that *"... a handbook or training document from the ministry will help us a lot*."

Some participant reported that they had received some kind of support even though it was representational or restricted. For example, P6 stated that "*Our principal encourages the use of GenAI but has not provided concrete support.*". Similarly, P12 expressed that "*School administration is aware of the issue but prefer to be silent.*" Such statements reveal a critical gap. Although teachers are expected to respond to the use GenAI in the classrooms, school administrations wait for higher level instructions, which in turn leads to both opportunities and problems for teachers.

In sum, it was found that many participants had high levels of personal agency and eager to use GenAI. However, they were not provided support, training and clear guidelines. This, in turn, paved the way for different integration patterns and variations in ethical and pedagogical attitudes. The findings imply that GenAI readiness is not just relate to teacher competence, but is quite linked to strategic policy and clear guidance.

Digital Equity and Structural Inclusion

This theme focuses on concerns about the systemic inequities regarding GenAI use in language assessment. As shown in Table 5, differences in infrastructure, socioeconomic status, and language proficiency were identified as key limitations. Participants emphasized that if these structural issues were not resolved, the integration of GenAI may increase existing inequities.

Table 5

Digital Equity and Structural Inclusion

Theme	Sub-Theme	Code (Refined Name)	Participants	Description
Digital Equity and Structural Inclusion	Access Disparities and Infrastructure Gaps	Device and Connectivity Limitations	P3, P6, P10	Limited access to devices and stable internet hinders GenAI use in low-resource settings.
	Access Disparities and Infrastructure Gaps	Institutional Infrastructure Gaps	P1, P5, P9	Schools often lack sufficient digital infrastructure to support equitable AI integration.
	Access Disparities and Infrastructure Gaps	Language and Literacy Barriers to GenAI Use	P2, P4, P11	Language limitations (e.g., low English proficiency) affect how effectively students can use GenAI tools.
	Socioeconomic and Regional Inequities	Rural-Urban Divide in Digital Access	P2, P4, P8	Participants report substantial differences between urban and rural students' access to GenAI tools.
	Socioeconomic and Regional Inequities	Financial Barriers to GenAI Use	P3, P7, P11	Economic challenges prevent some students from using premium or consistent GenAI tools.
	Ethical and Pedagogical Concerns	GenAI Privilege and Educational Fairness	P6, P10, P12	Concerns about GenAI widening existing privilege gaps and creating new educational injustices.
	Ethical and Pedagogical Concerns	Inclusive Pedagogy and Tool Selection	P4, P5, P9	Educators emphasize the need for inclusive design and fair access when integrating GenAI.

Participants widely reported that students' access to GenAI is shaped by technological infrastructure. In low-resource contexts, limited device availability and internet instability pose major hurdles. As P10 put it, "*Many students do not even have regular internet access at home.*" P3 added that some students "*… only use mobile phones, and AI tools don't work well on them.*" These infrastructural gaps extend to teachers as well, with P1 stating, "*Even I have to bring my own device to school.*" Participants also mentioned linguistic proficiency as a limitation. For example, P2 said that *"…. students with low English proficiency have problems using ChatGPT*." In addition, P11 stated that, "*Even when responses are translated, students with limited English have problems*". These statements indicate that linguistic ability functions as a gatekeeper to GenAI efficacy. In the study, geographic and economic inequalities were also mentioned. P4 stated that "*Students in the urban areas can use AI tools with no trouble. We do not have access stable internet here in the rural.*" In addition, P3 emphasized that *"… some students cannot afford premium tools.*"

The participants also raised ethical concerns. They stated that GenAI has the potential to worsen inequality as it may offer an advantage to the students who had better resources and competence. In this sense, P6 used, "*Students having more resources can perform better with using AI.*" Similarly, P12 said that "*It is like giving a calculator to someone who already knows math.*" In response, some participants reported that they tried to adopt

inclusive choices. For example, P9 said that "*I use free tools that every student can access.*". Similarly, P5 emphasized that "*We need tools for everyone, not just for those having fast Wi-Fi and good devices.*". However, the attempt to make use of inclusive GenAI implementation was limited to individual teachers since there was no institutional framework. Although they tried to be inclusive in GenAI integration, the participants lacked systemic support to ensure widespread change.

Discussion

Ethical Gatekeeping and Teacher Cognition in AI-Supported Assessment

It was found that the participants' perceptions regarding GenAI were influenced by their beliefs about pedagogy, ethics, and learner autonomy instead of institutional guidelines. This finding is line with Borg's (2003) teacher cognition framework. In teacher cognition, instructional decisions are closely connected with teachers' thought, information, and value. Participants also made a clear distinction between using GenAI as a supportive tool and using it as a replacement for learner work. This finding indicates that the participants appreciated student agency. This contrast, however, highlights a more profound conflict: teachers praised GenAI's capacity to improve learning, but they were also concerned that it might undermine the very learner autonomy they were trying to safeguard. It was also revealed that the participants acted as "ethical stewards" (Wiseman, 2024) in that they tried to promote responsible GenAI use by students. Since there was not a formal policy or guideline on the use of GenAI, participants used their own personal ethical frameworks. This finding is supported by Tajeddin et al. (2022) who found that teachers often establish their own behavior codes. Although flexibility was made possible by this autonomy, it also meant that different classrooms made different decisions regarding authorship, fairness, and acceptable AI use, which could have resulted in inconsistent practices and confused students.

In the study, the participants also stressed the importance of academic ownership and critical thinking. This finding aligns with Shohamy (2001) and Canagarajah (2005), who promoted fostering learner autonomy. It was also found that instead of being against the use of AI, the participants were cautious and tried to balance its use between educational potential and ethical consequences. Such a contradiction also reported by Bower et al. (2024) who found that although teachers appreciated GenAI's innovative promises, they required assessment methods promoting learner agency. This contradictory approach reflects a form of pedagogical ambivalence in which worries about the long-term development of students temper excitement for innovation. In this context, most of our participants expressed optimism. Although they supported pedagogical use of GenAI in the classroom, they paid attention to the threat of overdependence. Such an attitude also exists in the literature. For example, Yılmaz-Virlan and Tomak (2025) and Ong and Annamalai (2024) reported that teachers are interested in using GenAI in their classroom practices but they also fear that students' own abilities may decline if they excessively use it. This ongoing worry highlights an important problem: how can teachers use GenAI

to enhance learning without inadvertently encouraging cognitive passivity or lessening the demand for in-depth engagement? Since there were no clear policy or guidelines, the participants used their own decision making codes in their classroom practice. This supports Li's (2020) notion that teacher cognition is influenced by relationships and the context in which they work. Neverheless, the dependence on individual judgment also exposes a structural weakness: when institutions say nothing, teachers are left to shoulder the ethical and pedagogical burden of integrating GenAI, which may not be a long-term viable strategy.

Reconsideration of Assessment Literacy in AI Contexts

It was found that the participants did not just integrate GenAI in their classroom practices but also reconsidered the foundational assessment doctrines such as validity, authenticity, and fairness. In other words, the participants adopted a reflective position on language assessment literacy (Inbar-Lourie, 2017; Fulcher, 2012). The participants use of GenAI showed that they acted as technology managers and communicators at the same time, which is what Banitalebi et al. (2025) call technologist-communicators in their TALiDE model. However, the participants also expressed concerns as to if AI-generated texts or answers precisely show what the students can do. Similar concerns were also reported by Suherman (2022) and Mizumoto and Eguchi (2023). They expressed that students' use of GenAI may provide a wrong impression of high proficiency and hide their actual level. A fundamental validity question is brought up by this ambiguity is whether teachers evaluate the output quality of AI tools or student performance. In these situations, the distinction between learner competency and machine assistance becomes more blurred. In order to eliminate this, our participants used a number techniques such as using in-class writing and oral exams, which they believed to be more prone to AI. The use of such tasks is also supported by Taylor (2013) and Farooqui et al. (2024) since while conducting them students themselves are required provide the key ideas and language rather than simply using AI tools. This method might increase authenticity, but it also runs the risk of favoring some expressions such as handwriting or spontaneous speech over others that might still show language proficiency. In addition, since there was no available standardized rubrics, teachers had to develop and use their own rubrics. This situation indicates that the participant may a lack ethical assessment literacy skills (Coombe et al., 2020). While self-made rubrics demonstrate flexibility, they also expose educators to inconsistent grading and moral judgment, particularly in situations where institutional policies are unclear or nonexistent. Similarly, since there is no clear guidelines or policy, the participants developed their own guidelines to decide how GenAI should and shouldn't be used. This behavior is described as informal boundary-setting by Chapelle et al. (2024). Although such informal regulation is short-term necessary, it raises long-term questions about fairness and accountability. The experiences and outcomes of students can differ greatly between classrooms if there are no common institutional standards. The participants also expressed that they increasingly focused on continuous, developmental feedback, which reflects Shohamy's (2001) critical assessment paradigm.

This suggests a shift toward learner-centered, formative approaches. But teachers also need to put in more time, effort, and be digitally literate—resources that not all teachers may have equal access to. In sum, the findings reveal that language LAL frameworks should be revised to address the issues brought about by GenAI. It is more crucial that these frameworks go beyond theoretical proficiency in order to take into consideration the contextual, logistical, and ethical realities that educators encounter in classrooms mediated by AI.

From TPACK to Institutional Readiness: The Role of Teacher Agency

The findings showed that GenAI integration was heavily bottom-up (teacher-led and peer-supported) rather than top-down (institution-led). This findings points to a significant limitation in the TPACK framework. Even though it is good at linking technology, teaching methods, and subject content in theory, it often ignores structural and policy-level issues (Chapelle et al., 2024; Mishra et al., 2023). Such a discrepancy highlights a fundamental conflict in the way TPACK is currently used. This highlights the value of teacher agency while ignoring the ways in which an excessive dependence on teacher-led innovation may be a sign of institutional neglect. It was also found that participants mainly developed their confidence in using GenAI on their own for example by trying independently or through social learning from colleagues. Similarly, Ning et al. (2024) reported that TPACK knowledge is often developed through trial and error rather than planned instruction. Celik (2023) also reports that Turkish ELT teachers make use of self-developed strategies and peer networks since there is no clear policy support. Although this type of grassroots learning shows resiliency, it also brings up issues with sustainability, quality assurance, and the strain that comes with teachers having to handle complicated technologies alone. Although they tried to adopt, the participants also reported dissatisfaction with the lack of training, resources, and clear guidelines. These concerns were also reported in the literature. For example, Ong and Annamalai (2024), Koraishi (2025), and Moorhouse and Kohnke (2024) showed that teachers feel insecure or uncertain since they lack clear policies, training, or institutional guidance. This indicates not only a lack of resources but also a more profound mismatch between institutional preparedness and teacher motivation, which could impede long-term innovation or result in burnout. Although some participants mentioned that they benefited from peer collaboration, such attempts were not systematic and sufficient for an effective GenAI integration. This requests the important question of whether unofficial support systems can actually take the place of formal institutional training or if they run the risk of sustaining disparities in access to knowledge about GenAI. It was also found in the study that institutions lacked clear policy for GenAI use. Teachers were allowed to use and experiment with AI tools, but they did not receive any support on how to use it responsibly in classroom practice. As Chapelle et al. (2024) put it, this lack of policy places the responsibility of innovation totally on teachers. This circumstance reflects a larger problem with the adoption of educational technology: teachers are overworked and lack empowerment when institutional silence reigns. In this situation, autonomy becomes

essential rather than optional. Such a situation indicates a flaw in the TPACK framework. Although it takes into accout teachers' development in technology and pedagogy, but it ignores institutional factors. In this sense, Banitalebi et al. (2025) suggest broadening the TPACK model to cover institutional readiness and policy alignment. Without this extension, TPACK runs the risk of endorsing a highly customized approach to tech integration that ignores shared responsibilities and systemic injustices. However, insufficient institutional and policy support will not ensure the effective integration of GenAI in language assessment contexts. Thus, it is necessary to move away from frameworks that emphasize individual competence and toward ones that acknowledge the structural and collective aspects of technological change.

Justice-Oriented Integration of GenAI

The findings indicate that GenAI integration is happening in unequal educational contexts. Socioeconomic status, infrastructure inequalities, regional differences, and digital literacy levels play an important role in students' effective and successful use of GenAI. These issues are in line with critical pedagogy's emphasis on equity and structural power in learning (Norton & Toohey, 2004). Participants mentioned infrastructure problems such as unstable internet and limited devices as significant limitations. This finding is supported by Shohamy (2001) who argued that assessment performance is often influenced more by access than ability. This brings up a significant issue: tests run the risk of rewarding privilege over skill when access to technology becomes a requirement for success. Similar issues are also reported in Celik's (2023) study, where Turkish ELT teachers tried to implement classroom practices to preserve equity in under-developed contexts. In this study, our participants preferred free tools over paid ones and designed assignments that even students with fewer resources can easily perform. These teacher-led strategies illustrate what Wiseman (2024) calls an "ethics of care which is making choices based on justice, inclusion, and responsibility. These initiatives are admirable, but they also highlight a structural weakness: people are bearing more of the ethical burdens instead of institutions. Educational equity becomes erratic and precarious when fairness is dependent on teacher judgment rather than the creation of policies. Similarly, according to Chapman et al.'s (2024) AARDVARC framework, when GenAI is used in assessment, two key principles, that are responsibility and access, should be decisive in classroom practice. Nevertheless, as this study demonstrates, these values are frequently respected despite the lack of institutional mandates rather than because of them. This discrepancy highlights a more significant paradox: a dedication to digital justice does not always coincide with the drive for digital innovation in education. Consequently, critical pedagogy cannot function in a vacuum, even though it inspires educators to oppose injustices and adopt inclusive practices. Teacher efforts will remain isolated acts of resistance rather than steps toward structural change if GenAI is to be a truly transformative tool in education. Systemic reforms must go hand in hand with individual initiatives.

Implications for Practice

This study offers several implications regarding the ethical and sustainable integration of GenAI in language assessment contexts. Although the participants revealed professional agency and ethical sensitivity, their practices were often limited as schools or institutions did not provide clear policies, training, or resources. This shows that broader and system-level measures should be taken for an effective integration of GenAI in assessment contexts. Teachers need flexible and equitable ways to use GenAI without harming assessment validity. In addition, GenAI should be considered as a support tool instead of a content generator replacing students. Teachers should use assessment methods that focus on the learning process such as in-class assignments, oral examis, and process portfolios so that GenAI cannot be misused to give false results. At the same time, training should be provided to students regarding responsible use and ethics of GenAI. The lack of institutional policy was identified as a significant problem. Therefore, schools should provide guidelines that clearly define what is acceptable use of GenAI and what is not. In addition, institutions should provide both the physical resources such as devices and internet connection, the time for teachers to effectively collaborate and integrate GenAI in the assessment. Participants' dependence on trial-and-error implies the need for planned professional development. Already existing peer learning and collaboration should be facilitated by offering official workshops and ongoing institutional support. Last but not least, GenAI may deepen existing digital inequalities if teachers and policymakers do not focus on equality and inclusion. Therefore, both teachers and schools are required to address inequalities in student access especially in rural or under-resourced contexts.

Conclusion

This study examined the beliefs and responses of Turkish in-service English language teachers regarding using GenAI in language assessment. It was revealed teachers did not accept or reject GenAI carelessly. Instead, they examined it carefully and decide how and when to use it based on their expertise, ethical principles, and awareness. From a teacher cognition perspective, they demonstrated reflective decision-making. They both accepted GenAI's potential and raised concerns regarding academic honesty and learner autonomy. In order to respond to these concerns, participants redesigned their assessment practices and used in-class writing, oral tasks, and process-oriented assignments in order to ensure student voice and production. The findings also revealed that schools were not prepared to support GenAI use. Since there were no training or institutional guidelines, teachers had to teach themselves and rely on informal help from their colleagues. This highlighted the limitations of existing frameworks such as TPACK when institutional support is lacking. Equity issues made GenAI use even more problematic. Inequalities in students' access to technology, internet, and language proficiency ifluenced both the effectiveness and fairness of GenAI-supported assessment. Though teachers made efforts to mitigate these barriers, such adaptations remained informal and context-bound. By offering a context-specific and theoretically informed

analysis, this study adds to the evolving conversation on GenAI in education. It underscores the central role of teacher agency in developing ethical and pedagogically grounded responses to innovation, while also revealing the systemic challenges posed by policy gaps and digital inequities. Several limitations must be noted. The study draws from a relatively small sample within one national context and does not incorporate the perspectives of students, administrators, or policymakers. It also relies on self-reported data without observational or longitudinal validation. Future research should expand to include multi-stakeholder and mixed-method approaches to assess how GenAI impacts learning, assessment, and institutional structures over time. Ultimately, GenAI integration in assessment demands more than technical solutions—it calls for sustained professional development, supportive policies, and a commitment to pedagogical equity and justice.

ORCID

https://orcid.org/0000-0003-4103-4912

https://orcid.org/0000-0002-4755-3361

https://orcid.org/0000-0001-7412-8646

Publisher's Note

The claims, arguments, and counter-arguments made in this article are exclusively those of the contributing authors. Hence, they do not necessarily represent the viewpoints of the authors' affiliated institutions, or EUROKD as the publisher, the editors and the reviewers of the article.

Acknowledgements

Not applicable.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

CRediT Authorship Contribution Statement

Hossein Vafadar: Conceptualization, Writing – Review & Editing, Supervision

Batuhan Selvi: Conceptualization, Methodology, Formal Analysis, Writing – Original Draft

Ferzan Atay: Conceptualization, Resources, Writing – Original Draft

Generative AI Use Disclosure Statement

During the revision and preparation of this manuscript, the authors used Claude to improve readability. All content and intellectual contributions were generated by the authors. No generative AI tools were used for data analysis, interpretation, or autonomous content creation. The authors bear full responsibility for the integrity and accuracy of the final manuscript.

Ethics Declarations

World Medical Association (WMA) Declaration of Helsinki–Ethical Principles for Medical Research Involving Human Participants

This study was conducted in accordance with the Declaration of Helsinki and approved by Yozgat Bozok University Social and Human Sciences Ethics Committee. All participation was voluntary, and written informed consent was obtained from all participants prior to data collection. Participants were informed about the study's purpose, procedures, and their right to withdraw at any time without consequences.

Competing Interests

The author(s) declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this article.

Data Availability

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

References

Al-Kadi, A., & Ali, J. K. M. (2024). A holistic approach to ChatGPT, Gemini, and Copilot in English learning and teaching. *Language Teaching Research Quarterly*, *43*, 155–166. https://doi.org/10.32038/ltrq.2024.43.09

Al-khresheh, M. H. (2024). The future of artificial intelligence in English language teaching: Pros and cons of ChatGPT implementation through a systematic review. *Language Teaching Research Quarterly*, *43*, 54–80. https://doi.org/10.32038/ltrq.2024.43.04

Arıkan, A., & Peçenek, D. (2024). The effects of English language teaching enriched with digital applications on teacher cognition and student beliefs. *Dil Dergisi, 175*(2), 27–70. https://doi.org/10.33690/dilder.1493666

Arslan, S. (2025). English-as-a-foreign language university instructors' perceptions of integrating artificial intelligence: A Turkish perspective. *System, 131*, Article 103680. https://doi.org/10.1016/j.system.2025.103680

Banitalebi, Z., Estaji, M., & Brown, G. T. L. (2025). Measuring teacher assessment literacy in digital environments: Development and validation of a scenario-based instrument. *Educational Technology & Society*, *28*(2), 169–215. https://doi.org/10.30191/ETS.202504_28(2).RP10

Barrot, J. S. (2024). Leveraging ChatGPT in the writing classrooms: Theoretical and practical insights. *Language Teaching Research Quarterly*, *43*, 43–53. https://doi.org/10.32038/ltrq.2024.43.03

Berger, R. (2015). Now I see it, now I don't: Researcher's position and reflexivity in qualitative research. *Qualitative Research*, 15(2), 219–234. https://doi.org/10.1177/1468794112468475

Borg, S. (2003). Teacher cognition in language teaching: A review of research on what language teachers think, know, believe, and do. *Language Teaching*, *36*(2), 81–109. https://doi.org/10.1017/S0261444803001903

Borg, S. (2019). Language teacher cognition: Perspectives and debates. In X. Gao (Ed.), *Second handbook of English language teaching* (pp. 1149–1170). Springer. https://doi.org/10.1007/978-3-030-02899-2_59

Bower, M., Torrington, J., Lai, J., Petocz, P., & Alfano, M. (2024). How should we change teaching and assessment in response to increasingly powerful generative artificial intelligence? Outcomes of the ChatGPT teacher survey. *Education and Information Technologies, 29*, 15403–15439. https://doi.org/10.1007/s10639-023-12405-0

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology, 3*(2), 77–101. https://doi.org/10.1191/1478088706qp063oa

Canagarajah, S. (2005). Critical pedagogy in L2 learning and teaching. In E. Hinkel (Ed.), *Handbook of research in second language teaching and learning* (pp. 931–949). Routledge. https://doi.org/10.4324/9781410612700

Carlson, J. A. (2010). Avoiding traps in member checking. *The Qualitative Report*, *15*(5), 1102–1113. https://doi.org/10.46743/2160-3715/2010.1332

Casal, J. E., & Kessler, M. (2023). Can linguists distinguish between ChatGPT/AI and human writing?: A study of research ethics and academic publishing. *Research Methods in Applied Linguistics*, *2*(3), Article 100068. https://doi.org/10.1016/j.rmal.2023.100068

Celik, I. (2023). Towards Intelligent-TPACK: An empirical study on teachers' professional knowledge to ethically integrate artificial intelligence (AI)-based tools into education. *Computers in Human Behavior, 138*, Article 107468. https://doi.org/10.1016/j.chb.2022.107468

Chapelle, C.A., Baghestani, S. Beck, J., Kurt, S., Kusumaningrum, W., Lim, H. J., & Yin, S. (2024). Exploring Open AI in language teacher professional development. In C. A. Chapelle, G. H. Beckett, & J. Ranalli (Eds.), *Exploring artificial intelligence in applied linguistics* (pp. 217-242). Iowa State University Digital Press. https://doi.org/10.31274/isudp.2024.154.13

Chapman, E., Zhao, J., & Sabet, P. G. P. (2024). Generative artificial intelligence and assessment task design: Getting back to basics through the lens of the AARDVARC model. *Education Research and Perspectives, 51*, 1–36. https://doi.org/10.70953/ERPv51.2412001.

Choi, L. J., Kim, R., & Chung, S. J. (2025). Integrating chatbot technology in language teacher education: A TPACK-based analysis of pre-service teachers' professional development. *Language Learning & Technology*, *29*(1), 1–15. https://hdl.handle.net/10125/73621

Chong, S. W., & Reinders, H. (2023). Introduction: Learning-oriented language assessment—insights for evidence-based practices. In S. W. Chong & H. Reinders (Eds.), *Innovation in learning-oriented language assessment. New language learning and teaching environments* (pp. 1–11). Palgrave Macmillan, Cham. https://doi.org/10.1007/978-3-031-18950-0_1

Coombe, C., Vafadar, H., & Mohebbi, H**.** (2020). Language assessment literacy: What do we need to learn, unlearn, and relearn? *Language Testing in Asia, 10*(3). https://doi.org/10.1186/s40468-020-00101-6

Creswell, J. W., & Poth, C. N. (2016). *Qualitative inquiry and research design: Choosing among five approaches*. Sage publications.

Farazouli, A., Cerratto-Pargman, T., Bolander-Laksov, K., & McGrath, C. (2024). 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

Farooqui, M. O., Siddiquei, M. I., & Kathpal, S. (2024). Framing assessment questions in the age of artificial intelligence: Evidence from ChatGPT 3.5. *Emerging Science Journal*, *8*(3), 948–956. https://doi.org/10.28991/ESJ-2024-08-03-09

Freire, P. 1970. *Pedagogy of the oppressed*. Seabury

Fulcher, G. (2012). Assessment literacy for the language classroom. *Language Assessment Quarterly*, *9*(2), 113–132. https://doi.org/10.1080/15434303.2011.642041.

Guan, L., Zhang, E. Y., & Gu, M. M. (2024). Examining generative AI–mediated informal digital learning of English practices with social cognitive theory: a mixed-methods study. *ReCALL*, 1–17. https://doi:10.1017/S0958344024000259

Guo, Y., & Wang, Y. (2025). Exploring the effects of artificial intelligence application on EFL students' academic engagement and emotional experiences: A mixed-methods study. *European Journal of Education*, *60*(1), Article e12812. https://doi.org/10.1111/ejed.12812

Inbar-Lourie, O. (2008). Constructing a language assessment knowledge base: A focus on language assessment courses. *Language Testing*, *25*(3), 385–402. https://doi.org/10.1177/0265532208090158

Inbar-Lourie, O. (2017). Language Assessment Literacy. In E. Or. I. Shohamy & S. May (Eds.), *Language testing and assessment. Encyclopedia of language and education* (pp. 257-270). Springer, Cham. https://doi.org/10.1007/978-3-319-02261-1_19

Kerneža, M., & Zemljak, D. (2023). Science teachers' approach to contemporary assessment with a reading literacy emphasis. *Journal of Baltic Science Education*, *22*(5), 851–864. https://doi.org/10.33225/jbse/23.22.851

Kizilcec, R. F., Huber, E., Papanastasiou, E. C., Cram, A., Makridis, C. A., Smolansky, A., Zeivots, S., & Raduescu, C. (2024). Perceived impact of generative AI on assessments: Comparing educator and student perspectives in Australia, Cyprus, and the United States. *Computers and Education: Artificial Intelligence, 7*, Article 100269. https://doi.org/10.1016/j.caeai.2024.100269

Klimova, B., & Chen, J. H. (2024). The impact of AI on enhancing students' intercultural communication competence at the university level: A review study. *Language Teaching Research Quarterly*, *43*, 102–120. https://doi.org/10.32038/ltrq.2024.43.06

Kohnke, L., Moorhouse, B. L., & Zou, D. (2023). Exploring generative artificial intelligence preparedness among university language instructors: A case study. *Computers and Education: Artificial Intelligence*, *5*, Article 100156. https://doi.org/10.1016/j.caeai.2023.100156

Koraishi, O. (2024). The intersection of AI and language assessment: A study on the reliability of ChatGPT in grading IELTS writing task 2. *Language Teaching Research Quarterly*, *43*, 22–42. https://doi.org/10.32038/ltrq.2024.43.02

Koraishi, O. (2025). *Cultivating AI literacy among educators through an AI-enhanced language teaching workshop* [Master's thesis, Bursa Uludağ University, Institute of Educational Sciences]. https://hdl.handle.net/11452/50796

Kvale, S., & Brinkmann, S. (2015). *Interviews*. Sage.

Levi, T., & Inbar-Lourie, O. (2020). Assessment literacy or language assessment literacy: Learning from the Teachers. *Language Assessment Quarterly*, *17*(2), 168–182. https://doi.org/10.1080/15434303.2019.1692347

Li, L. (2020). *Language teacher cognition: A sociocultural perspective*. Palgrave Macmillan UK. https://doi.org/10.1057/978-1-137-51134-8

Lincoln, Y. S., & Guba, E. G. (1985). *Naturalistic inquiry*. Sage.

Liu, X. (2025). Navigating uncharted waters: Teachers' perceptions of and reactions to AI-induced challenges to assessment. *Asia-Pacific Education Researcher, 34*, 711–722. https://doi.org/10.1007/s40299-024-00890-x

Lynch, B. K. (2001). Rethinking assessment from a critical perspective. *Language Testing, 18*(4), 351-372. https://doi.org/10.1177/026553220101800403.

Malterud, K., Siersma, V. D., & Guassora, A. D. (2016). Sample size in qualitative interview studies: Guided by information power. *Qualitative Health Research, 26*(13), 1753–1760. https://doi.org/10.1177/1049732315617444

Mavroudi, A., & Tsagari, D. (2016, July). Language assessment literacy enhancement: Any room for technology-enhanced language learning? In *2016 IEEE 16th International Conference on Advanced Learning Technologies (ICALT)* (pp. 342–346). IEEE. https://doi.org/10.1109/ICALT.2016.87.

Miles, M. B., Huberman, A. M., & Saldaña, J. (2014). *Qualitative data analysis: A methods sourcebook* (3rd ed.). Sage.

Mishra, P., & Koehler, M. J. (2006). Technological pedagogical content knowledge: A framework for teacher knowledge. *Teachers College Record, 108*(6), 1017–1054. https://doi.org/10.1111/j.1467-9620.2006.00684.x

Mishra, P., Warr, M., & Islam, R. (2023). TPACK in the age of ChatGPT and generative AI. *Journal of Digital Learning in Teacher Education, 39*(4), 235–251. https://doi.org/10.1080/21532974.2023.2247480

Mizumoto, A., & Eguchi, M. (2023). Exploring the potential of using an AI language model for automated essay scoring. *Research Methods in Applied Linguistics, 2*(2), Article 100050. https://doi.org/10.1016/j.rmal.2023.100050

Moorhouse, B. L., & Kohnke, L. (2024). The effects of generative AI on initial language teacher education: The perceptions of teacher educators. *System, 122*, Article 103290. https://doi.org/10.1016/j.system.2024.103290.

Ning, Y., Zhang, C., Xu, B., Zhou, Y., & Wijaya, T. T. (2024). Teachers' AI-TPACK: Exploring the relationship between knowledge Elements. *Sustainability*, *16*(3), 978. https://doi.org/10.3390/su16030978

Norton, B., & Toohey, K. (2004). Critical pedagogies and language learning: an introduction. In B. Norton & K. Toohey (Eds.), *Critical pedagogies and language learning* (pp. 1-17). Cambridge University Press

Ong, Q. K. L., & Annamalai, N. (2024). Technological pedagogical content knowledge for twenty-first century learning skills: The game changer for teachers of industrial revolution 5.0. *Education and Information Technologies, 29*, 1939–1980. https://doi.org/10.1007/s10639-023-11852-z

Seo, J. Y. (2024). Exploring the educational potential of ChatGPT: AI-assisted narrative writing for EFL college students. *Language Teaching Research Quarterly*, *43*, 1–21. https://doi.org/10.32038/ltrq.2024.43.01

Shahid, M. K., Zia, T., Liu, B., Iqbal, Z., & Ahmad, F. (2024). Exploring the relationship of psychological factors and adoption readiness in determining university teachers' attitude on AI-based assessment systems. *The International Journal of Management Education, 22*(2), Article 100967. https://doi.org/10.1016/j.ijme.2024.100967

Shohamy, E. (2001). *The power of tests: A critical perspective on the uses of language tests*. Pearson Education. (Reprinted 2013 by Routledge). https://doi.org/10.4324/9781003062318

Suherman. A. (2022). Exploring language assessment literacy of EFL teachers in the context of Indonesian higher education. *English Review: Journal of English Education, 10*(3), 783–792. https://doi.org/10.25134/erjee.v10i3.6525

Tajeddin, Z., Khatib, M., & Mahdavi, M. (2022). Critical language assessment literacy of EFL teachers: Scale construction and validation. *Language Testing, 39*(4), 649–678. https://doi.org/10.1177/02655322211057040

Taylor, L. (2013). Communicating the theory, practice and principles of language testing to test stakeholders: Some reflections. *Language Testing*, *30*(3), 403–412. https://doi.org/10.1177/0265532213480338

Tian, Z., & Wang, C. (Eds.). (2025). *Rethinking language education in the age of generative AI* (1st ed.). Routledge. https://doi.org/10.4324/9781003426929

Ulum, Ö. G. (2024). Unveiling the layers: Analyzing ChatGPT Implementations in Turkish State Universities. *Base for Electronic Educational Sciences*, *5*(1), 114–134.

Vasileiou, K., Barnett, J., Thorpe, S., & Young, T. (2018). Characterising and justifying sample size sufficiency in interview-based studies: Systematic analysis of qualitative health research over a 15-year period. *BMC Medical Research Methodology, 18*(1), 148. https://doi.org/10.1186/s12874-018-0594-7

Wiseman, C**.** (2024). *Critical moments: Exploring critical pedagogy in an adult ESL classroom* [Doctoral dissertation, York University]. https://hdl.handle.net/10315/42413

Yılmaz Virlan, A., & Tomak, B. (2025). AI tools for writing: A Q-method study with Turkish instructors of English. *Education and Information Technologies, 30,* 16997–17021. https://doi.org/10.1007/s10639-025-13455-2

Yu–Ting, K. (2023). Empowering preservice English teachers with language assessment literacy concepts and practices: Application of Vygotskian concept-based language instruction. *The Modern Language Journal*, *107*, 68–94. https://doi.org/10.1111/modl.12818.

Zhao, J., Chapman, E., & Sabet, P. G. P. (2024). Generative AI and educational assessments: A systematic review. *Education Research and Perspectives*, *51*, 124–155. https://doi.org/10.70953/ERPv51.2412006