International Journal of Multidisciplinary Research and Growth Evaluation.

EFL Educators' Perceptions of AI-Driven Assessments: A Systematic Review

Reza Sahmaniasl
School of Foreign Languages, University of Beykoz, Turkey

* Corresponding Author: **Reza Sahmaniasl**

Article Info

ISSN (online): 2582-7138
Volume: 06
Issue: 03
May-Jun 2025
Received: 05-03-2025
Accepted: 03-04-2025
Page No: 672-680

Abstract
As artificial intelligence gains prominence in education, its application in designing and generating language tests has generated renewed interest in assessment. This systematic review examines the perception of EFL educators regarding AI-generated exam questions and further investigates classroom judgments of AI-assisted assessment tools in terms of quality, relevance, and efficacy. Hence, through peer-reviewed studies published from 2023 to 2025, the review synthesizes empirical findings concentrating on the views of teachers concerning AI-generated questions. Along the lines of the PRISMA guidelines, the review incorporated studies found in ERIC, Scopus, Web of Science, and other academic databases. Some of the key themes explored are: trust in AI, alignment with language learning goals, perceived bias or limitations, and the changing role of the teacher in designing assessments. Results indicated mixed feelings regarding AI in that teachers feel AI can assist in improving efficiency and creativity in assessments but remain worried regarding aspects of accuracy, context appropriateness, and pedagogical value of AI-generated content. The review points out some recommendations, such as teacher training, tools that are more transparent, and further investigations to foster the responsible integration of AI in EFL education.

Keywords: EFL Education, AI-Driven Assessment, Teacher Perceptions, Language Learning Technology, Systematic Review

Introduction

The use of computers and technological tools in language teaching has created a new stage in language teaching and learning (Valizadeh, Sahmaniasl, 2023) [25]. In recent years, the integration of artificial intelligence (AI) into educational contexts has transformed various aspects of teaching and learning, including assessment design. AI technologies are now capable of generating exam content ranging from multiple-choice questions to essay prompts, often within seconds. These tools, powered by large language models and natural language processing, have gained popularity in English as a Foreign Language (EFL) classrooms for their potential to reduce teacher workload, increase efficiency, and provide scalable assessment solutions. However, while technological advancements are rapidly evolving, educators' perceptions and acceptance of AI-generated exam content remain a critical yet underexplored area of inquiry.

It is necessary to know the perspectives of EFL teachers toward AI assessment materials because their trust and expertise may either favor or hinder the introduction and use of these tools in the actual classroom. Quality issues often challenge the willingness of teachers to adopt AI-based solutions, encompassing respects of content, language appropriateness, cultural relevance, fairness, and adherence to curriculum objectives. Questions about academic integrity and standardization arise amidst the current shift in high-stakes testing contexts toward AI. In turn, such scenarios pose new questions about the teacher's role in creating assessments.

This systematic review aims to synthesize existing research on EFL educators' perceptions of AI-generated exam content. By examining their experiences, attitudes, and concerns, the study seeks to highlight the key factors influencing the adoption and evaluation of AI-assisted assessment tools and to identify gaps in the current literature.

Significance of the Review

Though AI is increasingly used in educational settings, there is still a new and under-investigated area in the application of AI-generated exams in EFL assessment. Although AI-powered tools like ChatGPT, Quizgecko, and other test generators promise to save time and customize learning, little empirical evidence is there on how EFL teachers view the reliability, fairness, and pedagogical benefits of AI-generated materials. These studies tend to revolve more around the technical abilities of AI systems rather than around the human factors that affect the ability of AI systems to serve as classroom tools.

Despite the growing use of AI in education, particularly in generating test materials, research on how teachers perceive and respond to these tools remains limited. This gap is especially important as AI begins to influence not just classroom exercises but also formal evaluations. Teachers' attitudes matter significantly — their trust in and understanding of AI can determine whether these tools are meaningfully incorporated into curricula, assessments, and professional development. If educators see them as unreliable or out of sync with their teaching aims, they are unlikely to adopt them, no matter how advanced the technology may be. This review brings together existing studies to explore how English language teachers interact with AI when designing exam materials. In doing so, it offers practical insights for education policymakers, technology developers, and teacher training providers. By outlining the benefits, challenges, and necessary supports for using AI in assessment, the study aims to guide more thoughtful and effective integration of these tools in language education.

Review of Related Literature

The Emergence of AI in Language Assessment

The integration of artificial intelligence (AI) into language assessment has introduced innovative tools capable of generating exam content, such as multiple-choice questions and essay prompts. These AI-driven tools, including platforms like ChatGPT and Quillionz, offer potential benefits in terms of efficiency and scalability in test creation. However, the adoption of these technologies in English as a Foreign Language (EFL) contexts necessitates an examination of their alignment with pedagogical objectives and assessment standards.

Recent research has begun to shed light on how EFL teachers perceive the use of AI in exam development. For instance, Nguyen *et al*. (2024) [18] examined educators' experiences with Quillionz, an AI platform used to generate reading comprehension tasks. While many teachers valued the time-saving aspect of the tool, they also raised concerns about the relevance and overall quality of the items it produced. In a similar study, Lee (2024) [16] focused on secondary school teachers in South Korea and their experiences with ChatGPT. The findings suggested a generally positive outlook toward adopting AI technologies, although educators also voiced reservations, particularly around the accuracy of content and the risk of students becoming overly dependent on these systems.

As AI tools become more integrated into assessment practices, a number of challenges have surfaced. Teachers have questioned whether the content produced by AI is culturally appropriate or contextually meaningful for their students. Another area of concern involves academic integrity, as some studies have shown that AI-generated responses can bypass traditional plagiarism checks, leading to doubts about the effectiveness of existing assessment strategies. These issues highlight the importance of carefully assessing the role of AI in educational settings to ensure it aligns with both ethical expectations and pedagogical goals.

The integration of AI in assessment necessitates targeted professional development for educators. Studies emphasize the importance of enhancing teachers' AI literacy, enabling them to effectively utilize AI tools while maintaining pedagogical integrity.

EFL Educators' Perceptions of AI-Generated Exam Content

The use of artificial intelligence in language teaching has become a growing area of interest, particularly in relation to how EFL instructors view and apply these tools in designing exam materials. Nguyen *et al*. (2024) [18] conducted a study involving 48 EFL teachers from a vocational college in Hanoi, examining their engagement with Quillionz—an AI platform that creates reading comprehension quizzes. Over a four-week period, researchers used pre- and post-intervention surveys to track changes in teacher attitudes. Results showed a noticeable improvement in perceptions, with many educators recognizing the tool's ability to streamline quiz preparation. Still, questions were raised about the depth and contextual fit of the AI-generated items, reinforcing the importance of teacher involvement in refining such content.

In a similar vein, Lee (2024) [16] investigated the perspectives of secondary school teachers in South Korea regarding the use of ChatGPT in classroom settings. The study surveyed 41 teachers from 40 different schools, revealing that nearly 70% had already used ChatGPT for tasks such as creating instructional materials and offering immediate feedback to students. Teachers praised its usefulness in saving time and adapting to individual learning needs. At the same time, they voiced concerns about students becoming too dependent on AI, the risk of academic misconduct, and the reliability of AI-generated content. The findings highlighted the need for well-designed teacher training programs to ensure AI is used thoughtfully and effectively within educational frameworks.

Further research by Dilzhan (2024) [9] examined EFL teachers' perceptions and use of ChatGPT. The study highlighted that while educators recognized the benefits of AI in providing diverse language inputs and facilitating student engagement, there were prevalent concerns about the ethical implications, including plagiarism and the authenticity of student work. Teachers emphasized the importance of establishing clear guidelines and ethical frameworks when incorporating AI tools into language assessment.

Collectively, these studies suggest that EFL educators are cautiously optimistic about the integration of AI tools in exam content creation. While acknowledging the efficiency and innovative potential of AI, educators emphasize the indispensable role of human judgment in ensuring the quality, relevance, and ethical standards of assessments. The findings advocate for ongoing professional development and the establishment of robust frameworks to guide the ethical and effective use of AI in language education.

Challenges and Concerns in AI-Driven Assessment

The integration of AI-generated exam content in educational settings has introduced a myriad of challenges that educators and institutions must navigate. One significant concern revolves around the contextual relevance and cultural

appropriateness of AI-generated questions. AI tools, trained predominantly on datasets reflecting certain cultural norms, may produce content that lacks sensitivity to diverse cultural contexts, potentially leading to misunderstandings or offense. This limitation underscores the necessity for human oversight to ensure that assessment materials are culturally inclusive and pedagogically sound.

One of the most urgent challenges emerging from AI-assisted assessment is the issue of academic integrity. A study at the University of Reading revealed that responses generated by AI were not identified as such in the vast majority of cases—94%, to be exact—and were even awarded higher marks on average than those written by students. This finding underscores the growing difficulty teachers face in distinguishing between authentic student work and machine-generated content, raising important questions about whether current assessment practices are still effective in maintaining academic standards.

The situation is further complicated by the limited accuracy of AI detection tools. Existing systems often misidentify content, either flagging student work incorrectly or failing to catch AI-produced material altogether. These inconsistencies pose a serious risk: on one hand, students may be wrongly accused of dishonesty, and on the other, unoriginal work may pass unchecked. Such challenges highlight the need to improve detection technologies and rethink how assessments are designed—shifting away from formats that can be easily manipulated by AI.

Ethical concerns also loom large in the discussion around AI in education. Because AI systems are trained on vast datasets, they may unknowingly reflect the biases embedded in those data, potentially reinforcing inequalities in the classroom. At the same time, the increasing use of AI-generated material in academic settings challenges conventional ideas of originality and authorship. This shift calls for a reassessment of institutional policies and honour codes to reflect the new realities of teaching and learning.

Implications for Teacher Training and Professional Development

The integration of AI in assessment contexts calls for urgent and targeted teacher training initiatives. As AI-generated exam content becomes increasingly accessible, EFL educators are expected not only to use these tools but also to critically evaluate their output for validity, fairness, and alignment with curriculum standards. However, research shows that many teachers feel unprepared for this role due to limited exposure to AI during their initial training or professional development (Zawacki-Richter *et al*., 2019) [29].

A key issue highlighted in recent literature is the need for greater AI literacy among educators—specifically, their understanding of how AI functions, what its limitations are, and how to apply it thoughtfully in educational contexts. Huang *et al*. (2023) [12] point out that when teachers are unfamiliar with how AI-generated test questions are created, they may rely too heavily on tools like ChatGPT or Quizgecko without critically evaluating the results. This can lead to the use of materials that are poorly contextualized or culturally insensitive, ultimately failing to meet learners' actual needs. Professional development should therefore go beyond the technical use of AI tools and include strategies for reviewing and adapting content, as well as aligning it with pedagogical goals.

Another important dimension is the ethical use of AI in the classroom. As Borenstein *et al*. (2022) [6] explain, content produced by AI can unintentionally reflect the biases embedded in its training data or contain factual errors that misguide both teachers and students. This highlights the importance of embedding ethics into teacher training. Educators should be encouraged to question the fairness and inclusivity of AI-generated materials and to evaluate whether such content supports equitable learning for all students. Issues such as data privacy, transparency, and bias mitigation need to be part of every AI training framework.

Evidence also suggests that hands-on training opportunities can significantly improve teachers' confidence and skill in working with AI. In a pilot project in Singapore, Teo *et al*. (2023) [23] offered workshops to help EFL teachers tailor AI-generated questions to suit various proficiency levels. Not only did the training lead to more frequent use of AI in classrooms, but it also helped shift teachers' attitudes—from scepticism to curiosity and cautious acceptance. This shift demonstrates that when educators are provided with practical, relevant support, they are more willing to engage with new technologies. Ultimately, as AI tools become more embedded in education, the role of the teacher is not diminished but redefined. Teachers are not expected to be passive users of these technologies but rather thoughtful evaluators who can adapt AI-generated materials to fit the needs of their students.

Methodology

Research Design

This study follows a systematic review design, adhering to the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) framework. The objective is to identify, synthesize, and critically evaluate existing empirical studies that explore EFL educators' perceptions of AI tools and their performance on exam content, including both qualitative and quantitative research. The process was designed to ensure transparency, reproducibility, and rigor throughout all stages of the review.

To maintain both quality and relevance, the literature review began with a focused search across major academic databases, including Scopus, Web of Science, ERIC, and ScienceDirect. The search targeted peer-reviewed articles published between 2018 and 2025 to ensure the review reflected the most recent discussions and findings in the field. Preference was given to research featured in well-regarded journals specializing in applied linguistics, language testing, and educational technology. Each article that met these criteria was examined in detail, with attention paid to the abstract, introduction, research design, data analysis, and key findings. This thorough reading process enabled a critical evaluation of the studies' strengths, their methodological approaches, and the relevance of their contributions to understanding how EFL teachers perceive and engage with AI-generated exam materials.

Research Questions

Following the careful selection and analysis of relevant literature, the review was guided by a set of focused research questions. These questions were designed to synthesize existing knowledge on the topic, highlight key trends and challenges, and identify areas for future investigation regarding the use of AI-generated exam content in EFL contexts. The central aim was to explore educators' perspectives on the effectiveness, reliability, and pedagogical

implications of such tools.

What are the perceptions of EFL educators regarding the use of AI-generated exam content in language assessment?

1. What benefits and challenges do educators identify when using AI tools for generating test materials?
2. How do educators evaluate the pedagogical validity and reliability of AI-generated assessment tasks?
3. What implications do these perceptions have for classroom practice, teacher training, and policy development?

To ensure the relevance, credibility, and scholarly quality of the studies selected for this review, clearly defined inclusion and exclusion criteria were applied during the screening process. Only peer-reviewed articles published between 2022 and 2025 were considered, reflecting the most recent developments in AI-assisted assessment. Studies were included if they were conducted within EFL or ESL teaching contexts and specifically focused on teachers' or educators' perceptions of AI-generated, AI-generated exam content, such as automated assessment tools, AI-powered test generation platforms, or related technologies. Additionally, only publications written in English were reviewed to maintain consistency in language and interpretability. Conversely, sources such as opinion pieces, blog posts, and other non-peer-reviewed content were excluded to preserve academic integrity. Studies that focused exclusively on student interaction with AI, without addressing educator perspectives, were also omitted. Furthermore, articles that did not engage with themes of assessment or language education were excluded from the final dataset.

Selection Process

The selection process for this systematic review was conducted in three distinct phases to ensure the inclusion of studies most relevant to the research focus. First, titles and abstracts were screened to eliminate clearly irrelevant or off-topic publications. In the second phase, the full texts of the remaining articles were carefully reviewed to evaluate their relevance and methodological soundness. Finally, the predefined inclusion and exclusion criteria were applied to determine the final list of studies to be included in the review.

Data Extraction and Analysis

A systematic approach was taken to analyse the selected studies using a clearly defined coding strategy. This allowed for consistent extraction of essential details from each article, including the year it was published, the journal in which it appeared, the research methods used, and the size and context of the sample. Particular focus was placed on identifying insights into how EFL educators perceive the use of AI-generated exam content—both in terms of its advantages and the concerns it raises. Once this information was gathered, a thematic analysis was carried out to organize the findings into major recurring topics. These included educators' trust in AI systems, how well these tools align with teaching objectives, ethical considerations in their use, and the implications for teacher training. These themes provided a structured basis for interpreting the data and drawing meaningful conclusions about current trends in the field.

Table 1: Demographic of Reviewed Articles

	Article	Journal	Publication Year	Methodology	Participants
1	Examining the perceptions of EFL instructors on AI integration in an English preparatory school	Thesis	2024	Thematic analysis method	10 EFL Instructors
2	Perceptions, Integration, and Learning Needs of Chatgpt Among EFL Teachers	Korean TESOL Journal	2024	Mixed Methods	41 Teachers
3	Chatgpt in foreign language teaching and assessment: exploring efl instructors' experience	Information Technologies and Learning Tools	2024	Mixed Methods	36 Instructors
4	English-as-a-foreign language university instructors' perceptions of integrating artificial intelligence: A Turkish perspective	SYSTEM	2025	Mixed Methods	230 EFL instructors
5	EFL teachers' perceptions of the use of an AI grading tool (cograder) in English writing assessment at Saudi universities: an Activity Theory Perspective	Cogent Education	2024	Mixed Methods	10 EFL Teachers
6	Generative AI in Student Essays: English Teachers' Perspectives on Effective Assessment Methods	Xlinguae	2024	Mixed Methods	50 Educators
7	Teachers' Perspectives on Integrating chatgpt into EFL Writing Instruction	TESOL Communications	2024	Phenomenological Approach	22 Higher Education Teachers
8	Exploring Teachers' Perceptions Toward the Integration of AI Tools in the Language Classroom	NIDA Journal of Language and Communication	2024	Mixed-Method	208 English Instructors
9	Turkish EFL Teachers' Awareness and Perspectives on Artificial Intelligence Incorporation into Language Instruction	Thesis	2024	Mixed-Method	443
10	The Future of Language Education: Teachers' Perceptions About the Surge of Large Language Models like chatgpt	Technology in Language Teaching & Learning	2024	Mixed-Method	100 Instructors
11	Can instructors detect AI-generated papers? Postsecondary writing instructor knowledge and perceptions of AI	Journal of Applied Learning & Teaching	2024	Mixed-Method	20 Instructors
12	Moroccan EFL Teachers' Perceptions of AI-Generated Content: Impact, Effectiveness, and Challenges in Language Learning	Frontiers in English Language and Linguistics	2024	Qualitative Approach	40 EFL Teachers
13	Pre-service teachers' attitudes towards artificial intelligence	Journal of Language and Cultural Education	2023	Quantitative Methods	137 pre-service EL Teachers

	and its integration into EFL teaching and learning				
14	Investigating the relationship between teachers' attitudes toward artificial intelligence and their artificial intelligence literacy	Journal of Educational Technology & Online Learning	2024	Mixed Methods	361 teachers
15	English Teachers' Attitudes and Opinions Towards Artificial Intelligence	International Journal of Research in Teacher Education	2023	Case Study	35 Teachers
16	Perceptions of EFL teachers on integrating ai tools for foundation-level English language instruction	International Journal of Education and Practice	2025	Descriptive Research Design	21 EFL Teachers
17	Investigating the tripartite interaction among teachers, students, and generative AI in EFL education: A mixed-methods study	Computers and Education: Artificial Intelligence	2025	Mixed Methods	66
18	The role of artificial intelligence in language teacher education: EFL learners' views on the concept of "chatgpt" through metaphorical analysis	Social Sciences & Humanities Open	2025	Qualitative Research	105
19	Teaching English and Artificial Intelligence: EFL Teachers' Perceptions and Use of chatgpt	Thesis	2024	Qualitative Research	11 Teachers
20	Teachers' Perspectives on AI-Driven Quillionz for Generating EFL Reading Comprehension Quizzes	Proceedings of the asiacall International Conference	2024	Mixed Methods	48 English Teachers
21	Artificial intelligence in education and EFL teachers' attitudes: scale development and validation	Innovation in Language Learning and Teaching	2025	Quantitative Research	524 efl Teachers
22	Investigating EFL teachers' use of generative AI to develop reading materials: A practice and perception study	Language, Teaching Research	2024	Content Analysis	3 EFL Teachers

Thematic Findings

To systematically synthesize the evidence gathered from the selected studies, the findings were organized into four major thematic categories reflecting recurring patterns in the literature. These themes emerged through close reading of the abstracts and methodological details of each article, as guided by the review's research questions. The analysis reveals not only educators' general attitudes toward AI in assessment contexts but also the ways in which AI tools are being integrated, the concerns they raise, and the implications for teacher development. Each of the following subsections presents a detailed synthesis of the reviewed literature under a distinct thematic focus: General Perceptions and Awareness, Integration Practices and Pedagogical Uses, Concerns and Ethical Considerations, and Professional Development and Training Needs.

General Perceptions and Awareness

Across the reviewed studies, a clear pattern emerges suggesting that EFL educators hold a cautiously optimistic view of artificial intelligence (AI) in educational contexts, particularly when it comes to its role in exam content generation. While teachers do not unanimously endorse AI-generated tools, a majority recognize their potential to enhance efficiency, support learner engagement, and facilitate content development — particularly in areas like reading comprehension, writing prompts, and formative assessment. For example, several studies (Articles 1, 2, 5, 7, 12) document teachers' acknowledgment of AI's ability to save time in preparing exam materials, automate repetitive tasks, and assist with immediate feedback provision. These perceived benefits are particularly prominent among educators who have had direct experience using platforms like ChatGPT, CoGrader, and Quillionz.

However, the level of awareness and conceptual understanding of AI among teachers is uneven. While some instructors reported regular use of AI-based tools for exam preparation and classroom instruction (Articles 2, 5, 19), others were less confident in their understanding of how AI systems function, what their capabilities and limitations are, or how to evaluate their outputs critically (Articles 3, 4, 14). This variability often correlates with teachers' prior exposure to digital technologies and institutional support. For instance, studies conducted in technologically advanced educational environments (e.g., Korea, Kazakhstan, and Turkey) tended to report higher levels of AI familiarity among participants, as opposed to contexts where digital infrastructure and AI-specific professional development remain underdeveloped.

Moreover, teacher perceptions were often shaped by their level of AI literacy — a construct involving both technical knowledge and pedagogical application skills. In articles where surveys or mixed-methods designs were used (Articles 6, 9, 14), it became evident that teachers with higher academic qualifications, longer teaching experience, or previous engagement in technology-enhanced learning environments were more likely to perceive AI positively. They were also more open to integrating AI tools in assessment settings, provided there were safeguards for accuracy and ethical considerations. In contrast, educators unfamiliar with AI expressed scepticism, not necessarily rooted in technophobia, but rather in concerns about quality, contextual relevance, and the potential erosion of pedagogical autonomy.

Interestingly, some articles (e.g., Articles 8, 13) emphasized the generational dimension of AI adoption. Younger teachers and digital natives showed more enthusiasm toward experimenting with AI, whereas older or more traditionally trained educators expressed reservations, citing uncertainty about AI's long-term implications for academic integrity, student independence, and fairness in assessment design. This points to a generational learning gap that may need to be addressed through tailored training initiatives.

Finally, it is important to note that while general perceptions were mostly positive or neutral, few studies found unequivocal support for full integration of AI into high-stakes testing. Instead, educators called for a balanced, human-centered approach where AI serves as a supplementary tool, not a replacement for teacher judgment or subject matter expertise. Thus, while awareness is growing, and acceptance is gradually increasing, there remains a strong emphasis on the need for critical engagement with AI rather than blind adoption.

Integration Practices and Pedagogical Uses

Many of the studies reviewed show that EFL teachers are beginning to explore how AI can support their day-to-day teaching, especially when it comes to assessments. Teachers aren't just using AI for simple planning tasks anymore; they're experimenting with it to design quizzes, exam questions, and even writing prompts. Tools like ChatGPT, Quillionz, and CoGrader are being used not as passive assistants, but as creative partners that help generate content more efficiently.

One of the most common uses is creating test items, particularly for reading and vocabulary exercises. Teachers found AI helpful for producing quick question sets, saving time—especially when managing large classes. Although these tools often needed tweaking to better fit student levels or curriculum goals, many educators found them to be a useful starting point. Writing assessments were another area of interest. Some teachers used ChatGPT to offer students feedback or used tools like CoGrader to score short essays. While this freed up time for teachers to focus on higher-level writing issues, there was also concern that students might become too dependent on these tools instead of building their own language skills.

Even with growing interest, most teachers stopped short of fully automating their exams. Many expressed doubts about the quality and relevance of AI-generated questions. Some reported that the content lacked nuance, contained cultural mismatches, or just wasn't deep enough for formal assessments. This often meant teachers had to review and revise AI-created content before using it. The message across studies was clear: AI can support assessment design, but it still requires careful human oversight.

In some classrooms, teachers adapted AI-generated content to suit different student levels, using it for personalized learning or tiered activities. But across the board, they emphasized the need to align AI outputs with real curriculum expectations. A number of studies also pointed to a gap between informal use and institutional practice. While teachers were willing to experiment with AI in low-stakes settings, many avoided using it for official assessments due to a lack of policy guidance, unclear standards, and concerns about fairness. Overall, AI is being used as a support tool—not to replace the teacher's role, but to extend it—when implemented thoughtfully.

Concerns and Ethical Considerations

While the reviewed literature demonstrates a growing openness toward integrating AI in exam content creation, it equally underscores a wide range of concerns, many of which stem from ethical, cultural, and pedagogical considerations. These concerns were not marginal or peripheral; rather, they were central to educators' decision-making processes about whether, when, and how to use AI-generated exam content.

One of the most commonly cited concerns involves the accuracy and appropriateness of AI-generated material. Multiple studies (e.g., Articles 4, 5, 7, 10, 18) report that teachers found AI-generated questions to be occasionally misleading, vague, or factually incorrect. Errors in grammar, misinterpretation of texts, and irrelevant multiple-choice distractors were frequently noted. In some cases, teachers expressed worry that over-reliance on AI tools could compromise the quality and validity of assessments, especially in formal exam contexts where precision and fairness are paramount. These findings highlight a broader issue: while AI can generate content rapidly, it often lacks the contextual sensitivity and subject-specific discernment that educators bring to the task.

Another significant area of concern relates to cultural relevance and bias. Several educators, particularly those working in non-Western or multilingual environments (e.g., Articles 3, 12, 14), noted that AI-generated content was sometimes based on cultural assumptions that did not align with students' backgrounds. For example, scenarios used in comprehension passages or speaking prompts occasionally reflected Western-centric norms, unfamiliar idioms, or culturally insensitive topics. These issues raise red flags about equity and inclusivity, especially in diverse EFL settings where cultural appropriateness is a core principle of effective pedagogy.

Concerns over academic integrity also featured prominently. In studies like Articles 8, 9, and 13, teachers described AI as a "double-edged sword." While it could assist students in revising and understanding test content, it also made it easier for learners to bypass independent work by relying on AI-generated answers. The increasing use of generative AI like ChatGPT has blurred the lines between support and substitution. Some educators observed that AI tools could be used by students to compose essays, answer reading questions, or even simulate oral responses — leading to potential breaches of authenticity in student performance.

Closely linked to this is the concern about plagiarism and authorship. Teachers reported that students occasionally submitted AI-generated content as their own, prompting questions about how to uphold academic honesty in an AI-rich environment. As noted in Article 9, even experienced educators expressed difficulty in distinguishing human-written from AI-written responses. This signals the need for more sophisticated plagiarism detection tools and revised assessment strategies that account for the presence of generative AI in students' learning ecosystems.

From a broader ethical perspective, some educators voiced discomfort with AI decision-making in high-stakes contexts. In Article 15, for instance, teachers questioned whether it was appropriate to allow an AI system — no matter how advanced — to determine the level of difficulty, scoring rubric, or appropriateness of a test item without human review. This echoes concerns raised in Article 20, where participants emphasized the need for human oversight and transparency in how AI-generated materials are produced and evaluated.

Finally, several studies (Articles 6, 13, 21) emphasized the psychological and professional discomfort some teachers felt in using AI tools. Fears of being replaced, loss of professional autonomy, and the devaluation of human judgment were cited as underlying anxieties. While many educators embraced AI as a support mechanism, others felt it introduced a level of surveillance or automation that threatened their traditional roles in assessment design.

In summary, the reviewed literature presents a nuanced view: EFL educators are not opposed to AI, but their engagement is filtered through a strong ethical lens. Concerns about accuracy, fairness, cultural bias, academic honesty, and pedagogical integrity remain dominant. For AI to be successfully integrated into EFL assessment, these concerns must be addressed through transparent design, teacher agency, and the development of robust guidelines that ensure AI serves as an aid — not a threat — to responsible assessment practices.

Professional Development and Training Needs

A clear theme across the reviewed studies is that while many EFL teachers are open to using AI in assessment, they often feel unprepared to do so effectively. Most educators are learning to work with AI tools like ChatGPT or Quillionz on their own, often through trial and error or informal exchanges with colleagues. Very few have received structured training or institutional guidance on how to use these technologies in ways that align with sound pedagogy (Articles 3, 6, 9, 14, 17).

One of the main obstacles is limited AI and digital literacy. While some younger or tech-savvy teachers feel confident experimenting with AI tools, others—especially those who trained before digital tools became widespread—admit they don't fully understand how the technology works or how to evaluate its reliability. Article 14, for instance, notes that many teachers want to use AI for generating exam content but hesitate due to a lack of technical skills and confidence. There's a clear need for foundational training that not only explains how to use the tools but also teaches the underlying principles, including data ethics and how AI makes decisions.

Several studies stress that professional development needs to go beyond basic tool demonstrations. Teachers want support in answering real pedagogical questions: How do we adapt AI content for different learner levels? How do we prevent students from becoming overly dependent on AI? And how can we maintain academic integrity when AI is easily accessible to students? These are practical and ethical challenges that require collaborative, context-specific training and space for reflection (Articles 8, 12, 16).

Importantly, many educators also pointed out the lack of clear institutional policies to guide AI use in assessment. Without clear rules or expectations, teachers feel they're working in a grey area—unsure of what is acceptable or how to use AI responsibly (Article 13). The literature calls for a broader shift in teacher education: training future teachers to work in AI-rich classrooms should no longer be optional. Effective AI integration requires more than tools—it requires a support system. Studies recommend long-term strategies like peer mentoring, expert consultation, and professional learning communities, all of which help educators move from experimenting with AI to using it in thoughtful, pedagogically sound ways (Articles 10, 15, 18, 21).

Findings

The reviewed studies collectively reveal a landscape of growing engagement with artificial intelligence in EFL assessment, marked by both curiosity and caution among educators. Teachers increasingly recognize that AI tools—particularly platforms like ChatGPT, CoGrader, and Quillionz—can support assessment-related tasks by reducing preparation time, generating question formats, and offering personalized feedback. Yet, this potential is filtered through varying degrees of familiarity, institutional support, and pedagogical readiness.

A central finding is the diversity in teacher experience and confidence. While some educators actively use AI to draft or refine exam materials, others approach it with hesitation, often due to limited training or lack of clarity about how to critically evaluate AI-generated content. This divide is shaped by multiple factors, including age, technological exposure, and institutional context. Teachers in more digitally mature environments tend to show greater willingness to experiment with AI, though even among this group, few fully automate assessment tasks without manual oversight.

Educators consistently identify both benefits and boundaries in their use of AI. On the one hand, AI offers efficiency and a source of ideas, especially for routine or low-stakes tasks. On the other hand, concerns arise around content accuracy, cultural relevance, and academic integrity. Teachers noted that AI-generated questions sometimes contain superficial language, factual errors, or culturally mismatched references, particularly when applied across diverse learner groups. These issues reinforce the importance of teacher judgment, as most educators continue to act as editors and evaluators of AI content rather than passive adopters.

Perhaps most telling is the underlying tension between innovation and uncertainty. Teachers value AI's support but do not feel fully equipped or guided in its use. Few institutions appear to provide formal frameworks for ethical or pedagogical use, and this absence leaves educators vulnerable to guesswork. As a result, AI use often remains informal—helpful in planning and formative feedback but rarely adopted in high-stakes summative assessments. Without policies, training, or safeguards in place, many teachers choose caution over integration.

In addition, the studies show that professional development is widely seen as both necessary and insufficient. Teachers express a strong desire for context-sensitive training that addresses more than the technical operation of AI tools. They want to explore how AI intersects with real pedagogical challenges: how to differentiate tasks, uphold fairness, and maintain student engagement in AI-supported learning environments. While some teachers gain confidence through peer learning or self-study, sustainable and responsible AI integration will likely depend on structured support networks, ongoing mentoring, and the inclusion of AI literacy in pre-service education.

Taken together, the findings suggest that EFL educators are not resistant to AI, but they are navigating a complex transition. They are optimistic about AI's role in easing assessment burdens, but cautious about relinquishing their professional agency. Their experiences point to a clear conclusion: AI can support quality assessment in EFL contexts, but only when human expertise, ethical oversight, and pedagogical alignment remain central to the process.

Conclusion and Implications

This systematic review explored EFL educators' perceptions of AI-generated exam content by analyzing 22 open-access studies from a range of educational contexts. The findings show a profession in transition — one that is increasingly aware of the affordances AI brings to assessment but also alert to its limitations. Teachers see the practical benefits of AI tools for saving time, generating assessment items, and offering scalable feedback. Yet, their willingness to engage with these tools is closely tied to how well they understand them, how much support they receive, and how clearly their use aligns with pedagogical and ethical standards.

What emerges most strongly is that teachers are not rejecting AI; rather, they are cautiously adapting it to fit within their professional frameworks. AI is not being used to replace educator expertise but to assist it — and even then, its outputs are carefully reviewed and reworked. Teachers remain gatekeepers of quality, adapting AI-generated content to meet their students' needs, cultural contexts, and curriculum goals. However, this gatekeeping role also brings strain, especially in the absence of clear policies or adequate training. Many

educators are left navigating the challenges of AI implementation without institutional guidance, often relying on informal experimentation or peer collaboration.

These realities hold important implications for both educational practice and policy. First, institutions must move beyond simply promoting the use of AI and instead provide structured professional development that addresses both technical and pedagogical dimensions. Teachers need to know not only how to use AI tools, but also when and why — and how to evaluate their outputs critically. Training programs should help educators identify cultural bias, uphold academic integrity, and differentiate AI content for varied learner profiles.

Second, ethical frameworks and institutional guidelines are urgently needed to support fair, transparent, and responsible AI use in assessment. Educators require clarity on issues like plagiarism detection, student authorship, and the role of AI in summative evaluation. Without such frameworks, the risk is not only inconsistency in classroom practices, but also a loss of trust in AI's role within the broader learning process.

Finally, the review calls attention to future directions for research and teacher preparation. There is a pressing need for longitudinal studies to understand how AI affects assessment quality over time, as well as investigations into how students experience and perceive AI-generated exams. Equally, pre-service teacher education programs must begin to embed AI literacy as a core competency, preparing the next generation of educators to enter classrooms where AI will be an expected — if not essential — part of instructional design and assessment.

In short, the integration of AI into EFL assessment is not a technological issue alone; it is a pedagogical, ethical, and institutional one. Its success depends not on replacing the teacher, but on empowering the teacher to use AI wisely, critically, and creatively.

Despite the growing interest in AI's role in education, research on its use in EFL assessment remains in its early stages. The findings of this review point to several key areas for future scholarly inquiry:

1. Longitudinal studies are needed to evaluate the sustained impact of AI integration on assessment quality, teacher agency, and student outcomes over time.
2. Comparative studies could explore cross-cultural differences in how EFL educators perceive and adapt AI-generated materials, helping to identify best practices across diverse educational systems.
3. More empirical research is required to evaluate the effectiveness of professional development programs aimed at building teachers' AI literacy.
3. There is also a need to investigate student perceptions and responses to AI-generated exams, especially regarding motivation, trust, and fairness.
4. Finally, researchers should explore how pre-service teacher education programs are equipping future educators to navigate AI-enhanced assessment environments.

References

1. Ahmadi Fatalaki J, Saeedi Z, Sheikhzadeh Marand E, Nasiri Firuz AR. Artificial intelligence in education and EFL teachers' attitudes: scale development and validation. Innov Lang Learn Teach. 2025;1-17.
2. Alsalem MS. EFL teachers' perceptions of the use of an AI grading tool (CoGrader) in English writing assessment at Saudi universities: an Activity Theory Perspective. Cogent Educ. 2024;11(1):2430865.
3. Arslan S. English-as-a-foreign language university instructors' perceptions of integrating artificial intelligence: A Turkish perspective. System. 2025;131:103680.
4. Azap S. The role of artificial intelligence in language teacher education: EFL learners' views on the concept of "ChatGPT" through metaphorical analysis. Soc Sci Humanit Open. 2025;11:101394.
5. Boğaz N. Examining the perceptions of EFL instructor on AI integration in an English preparatory school [Master's thesis]. Middle East Technical University; 2024.
6. Borenstein J, Jasso H, Howard A. The ethical implications of AI in education. AI Soc. 2022. https://doi.org/10.1007/s00146-022-01413-3
7. Çayak S. Investigating the relationship between teachers' attitudes toward artificial intelligence and their artificial intelligence literacy. J Educ Technol Online Learn. 2024;7(4-ICETOL 2024 Special Issue):367-383.
8. Dahia I. Moroccan EFL Teachers' Perceptions of AI-Generated Content: Impact, Effectiveness, and Challenges in Language Learning. Front Engl Lang Linguist. 2024;1(1):22-34.
9. Dilzhan B. Teaching English and artificial intelligence: EFL teachers' perceptions and use of ChatGPT [Dissertation]. SDU University; 2024.
10. Guan L, Lee JCK, Zhang Y, Gu MM. Investigating the tripartite interaction among teachers, students, and generative AI in EFL education: A mixed-methods study. Comput Educ Artif Intell. 2025;8:100384.
11. Hossain MK, Al Younus MA. Teachers' perspectives on integrating ChatGPT into EFL writing instruction. TESOL Commun. 2025;4(1):41-60.
12. Huang H, Yang C, Hung S. Teachers' Perceptions of AI in EFL Classrooms: Challenges and Opportunities. TESOL Q. 2023.
13. Kalra R. Exploring Teachers' Perceptions Toward the Integration of AI Tools in the Language Classroom. J Lang Commun. 2024;29(45):21-36.
14. Khattak ZI, Ghorbanpoor M, Mathew BP. Perceptions of EFL teachers on integrating AI tools for foundation-level English language instruction. Int J Educ. 2025;13(2):653-668.
15. Kornieva VLZ. Generative AI in Student Essays: English Teachers' Perspectives on Effective Assessment Methods. XLinguae. 2024;17(4):235-250.
16. Lee H. Perceptions, integration, and learning needs of ChatGPT among EFL teachers. Korea TESOL J. 2024;20(1):3-19.
17. Murray N, Tersigni E. Can instructors detect AI-generated papers? Postsecondary writing instructor knowledge and perceptions of AI. J Appl Learn Teach. 2024;7(2).
18. Nguyen TL, Nguyen HL, Le DH. Teachers' Perspectives on AI-Driven Quillionz for Generating EFL Reading Comprehension Quizzes. In: Proceedings of the AsiaCALL International Conference. 2024;6:20-34.
19. Pokrivcakova S. Pre-service teachers' attitudes towards artificial intelligence and its integration into EFL teaching and learning. J Lang Cult Educ. 2023;11(3):100-114.
20. Sahmaniasl R. AI integration in teaching ELT for Iranian

intermediate and higher levels learners. Multidiscip Res Growth Eval. 2024;5(3):390-393.
21. Sahmaniasl R. Utilizing artificial intelligence for creating educational content, a case study of Iranian ELT teachers. Multidiscip Res Growth Eval. 2024;5(3):394-399.
22. Sütçü SS, Sütçü E. English Teachers' Attitudes and Opinions Towards Artificial Intelligence. Int J Res Teach Educ. 2023;14(3).
23. Teo T, Lim KM, Lee CB. AI-assisted language instruction: A case study from Singapore. Comput Educ. 2023;194:104690.
24. Üretmen S. Turkish EFL teachers' awareness and perspectives on artificial intelligence incorporation into language instruction [Dissertation]. Necmettin Erbakan University, Institute of Educational Sciences; 2024.
25. Valizadeh S, Sahmaniasl R. The impact of using Grammarly in teaching writing skills of Iranian advanced EFL students. All Multidiscip Res Growth Eval. 2023;2(5):24-26.
26. Valizadeh S, Valizadeh G. Digital Storytelling for Intermediate EFL Learners: Enhancing Narrative Skills and Cultural Awareness. All Multidiscip Res Growth Eval. 2025;6(1):455-460.
27. Xin JJ. Investigating EFL teachers' use of generative AI to develop reading materials: A practice and perception study. Lang Teach Res. 2024;13621688241303321.
28. Zaiarna I, Zhyhadlo O, Dunaievska O. ChatGPT in foreign language teaching and assessment: exploring EFL instructors' experience. Inf Technol Learn Tools. 2024;4(102):176-191.
29. Zawacki-Richter O, Marín VI, Bond M, Gouverneur F. Systematic review of research on artificial intelligence applications in higher education – where are the educators? Int J Educ Technol High Educ. 2019;16(1):39.
30. Zimotti G, Frances C, Whitaker L. The Future of Language Education: Teachers' Perceptions about the Surge of Large Language Models Like ChatGPT. Technol Lang Teach Learn. 2024;6(2).