Language Testing in Focus: An International Journal

EUROKD

2026 (13), 19–31
https://doi.org/10.32038/ltf.2026.13.02

Generative AI Technology and Teacher Language Assessment Literacy – Can Language Teachers Keep up?

Karin Vogt
University of Education Heidelberg, Germany

Correspondence
Email: vogt@ph-heidelberg.de

Abstract
Language Assessment Literacy (LAL) has long been central to teachers' professionalisation, yet the field has given relatively little attention to digitalisation. The rapid emergence of generative AI in language teaching and assessment now compels a reconceptualisation of teacher LAL to incorporate AI related knowledge, skills and ethical awareness. This paper maps key affordances—automated item generation, scalable scoring, and timely individualised feedback—alongside critical challenges including overreliance, academic integrity threats, bias, data protection and equity concerns. It argues that teachers require prompting skills, critical evaluation of AI output, understanding of automated scoring systems, and pedagogical strategies to redesign assessments, and that these competences must be embedded in initial and in service teacher education. Sustainable implementation demands coordinated action by policymakers, researchers, test developers, teacher educators and school leaders to provide localised frameworks, professional development, and time for practice. New conceptions of teacher LAL have to integrate a dynamic AI component, with AI literacy and related requirements representing a moving target and LAL levels being underdeveloped in many educational contexts.

ARTICLE HISTORY
Received: 22 October 2025
Revised: 25 March 2026
Accepted: 08 May 2026

KEYWORDS
Language Assessment Literacy, Generative Artificial Intelligence, Language Teachers, Language Assessment

How to cite this article (APA 7th Edition):
Vogt, K. (2026). Generative AI technology and teacher language assessment literacy – can language teachers keep up? *Language Testing in Focus: An International Journal, 13*, 19–31. https://doi.org/10.32038/ltf.2026.13.02

Introduction[1]

Language assessment has always been a vital part of language teaching for teachers, with feedback acting as a hinge between the two (Vogt, 2022). Language teachers

[1] This paper is part of a special issue (2026, 13) entitled: Language Assessment Literacy in the New Era: New Concerns, Same Solutions? edited by Hassan Mohebbi.

require the contextual principles, values and beliefs, knowledge and skills to assess learners' language proficiency and related competencies in such a way that the assessment reflects the learners' actual proficiency. This goal has been the driving force in Language Assessment Literacy (LAL) as a relatively young field of language testing and assessment. For teacher LAL in particular, it was also about closing the gap between teachers' current LAL levels and the LAL-related competence profile that would enable teachers or other stakeholders to achieve this objective.

Over the past decade, the conceptualisations of LAL have changed and have become more complex. From simple componential models (Davies, 2008), they have developed to become multidimensional with competence profiles (Pill & Harding, 2013) that cater for various stakeholder groups (Taylor, 2013). While the components of LAL have diversified in terms of types of competences (Bøhn & Tsagari, 2021), and stakeholders (e.g. Villa Larenas, 2020, Villa Larenas & Brunfaut, 2024), they have also started to be empirically validated (e.g. Bøhn & Tsagari, 2021; Kremmel & Harding, 2020).

Although digitalisation has been relevant for language assessment contexts, the digital aspect of LAL has been neglected, with a few exceptions (e.g., Nayernia & Mohebbi, 2023). However, the disruptions that the advent of accessible Artificial Intelligence (AI) tools, mostly generative AI chatbots based on Large Language Models (LLMs), has brought about have resulted in a keen research interest in the topic. Publications on (generative) AI and assessment have mushroomed (e.g., Hannah et al. (2023) on automated scoring; Ockey et al. (2023), on assessing interactional competence with AI; on AI-advanced language assessment in general, e.g., Xi, 2023). The perspective of test takers is taken into consideration (Jin & Fan, 2023; Zhang et al., 2023), and human and AI scorers are compared (Stewart et al., 2025). First reviews (e.g. O'Grady et al., 2026) and various special issues have been published or are in preparation (e.g. Language Assessment Quarterly, Studies in Language Assessment and Language Teaching Research Quarterly). Not only automated scoring, but also automated test item generation, adaptive algorithms or security measures like AI-based proctoring are used e.g., for standardised testing (Chuang & Yan, 2025). Despite the enormous interest and research activity in AI-related aspects of language assessment, deliberations on teacher LAL and the impact on AI applications on teacher LAL have remained scarce.

The purpose of this reflective paper is to identify the affordances and challenges that language teachers are facing with regard to their assessment practices in an era of (generative) AI and to discuss their potential impact on key aspects that are crucial to teachers' LAL development, for teacher LAL that is digital, AI-enhanced and future ready.

Affordances and Challenges for Language Teachers in the New Era of AI

Language teachers have to navigate the changes in educational contexts that have been induced by AI and are still continuing to occur. Wu et al. (2025) designate the advent of generative AI tools as a new era for language education. Teachers in particular are affected by the transformative impact of AI technology both in their instructional decisions and in their assessment-related activities, as they design and manage learning environments for their learners. They are often involved in all phases of test development, from test design and administration to rating and scoring.

Depending on the educational context, teachers' expertise and LAL are required for classroom-based language assessment as well as high-states exams, often reflected in local assessment on a regional or national level (Dimova et al., 2022). Therefore, the focus of the current paper is on language teachers and the of audiences and challenges they are facing with regard to language assessment practices in a new era of generative AI although a variety of stakeholders is affected, such as educational authorities (Funa and Gabay, 2025), learners (Dilikitaş et al., 2024) and parents (e.g., Otermans et al., 2024).

By AI, Moorhouse (2024, p. 378 f.) understands "computer systems that use algorithms to imitate human intellectual abilities and skills and are capable of performing tasks that historically required human intelligence, e.g. recognizing speech, making decisions, and identifying patterns" (cf. also Ali et al., 2024). AI in education encompasses a broad range of applications such as intelligent tutoring systems (Zerkouk et al., 2025, automated writing evaluation (Vogt, 2025), AI chatbots and conversational agents (e.g. Kasneci et al., 2023) and other specialised tools like AI-assisted speech recognition (speech-to-text) tools that can function as assistive technology (Edyburn, 2020). AI literacy has also become a fuzzy term, so for now I would like to follow Chiu's (2025, p. 3225) conceptualisation that goes beyond mere technical knowledge, but that "refers to a foundational conceptual understanding of AI [that] (...) focuses on knowledge, critical thinking, and ethical awareness rather than technical skill". This is particularly important in the context of language education and touches on the competences involved in dealing with AI technology.

Generative AI as a subfield of AI is understood as "(...) a class of AI systems designed to produce content – such as text, images, audio – by learning patterns and structures from large datasets" (Chuang & Yan, 2025, p. 2). Lee et al. (2025, p. 85) observe that genAI functions as an umbrella term, "encompassing pre-trained AI technologies including image, music and text generation tools".

In the context of this paper, classroom-based testing and assessment and local tests are relevant. Classroom-based language assessment, in this connection, is conceptualised as the type of information that does not only obtain information on learners' attainment, but also learner progress (Lan & Fan, 2019). In order for language teachers to do this, LAL is necessary for teachers. A more recent definition

of LAL that has attempted to synthesise previous definitions is offered by Vogt et al. (2024, p. 1) as “the skills, knowledge, methods, techniques and principles needed by various stakeholders in language assessment to design and carry out effective assessment tasks and to make informed decisions based on assessment data”.

While the empirical knowledge base on the use of genAI technology in education is being built from various perspectives such as learners, teachers and institutions and in various contexts of language education, the focus of this paper is on language teachers in secondary and post-secondary education who have to navigate the impact of genAI in their professional contexts, gauging affordances and challenges of genAI on their teaching and assessment practices. Again, while the diversity of various AI applications is rapidly increasing, I would like to focus on genAI tools, mostly based on Neurolinguistic Programming, a form of AI use that is typically found in chatbots. The chatbots used by teachers are often free versions of chatbots like ChatGPT, GoogleGemini, Microsoft Copilot, DeepSeek Anthropic’s Claude or Mistral,that use less powerful models, depending on their availability.

The keen interest in genAI in education has brought about a rapidly growing body of literature, with conceptual pieces and empirical studies, on gen-AI-based practices in the language classroom (e.g., content creation, task design, feedback; Moorhouse, 2024), perceptions of language teachers (e.g., Khabbazbashi et al., 2024) and an evaluation of the opportunities and challenges of genAI in general and language education (e.g., Kiryakova & Angelova, 2023; Ngyuen & Goto, 2024) in terms of understanding its impact. The intensive attention for genAI in (language) education can be gauged by the numerous special issues (e.g., Jang & Sawaki, 2025; Yan & Huang, 2025) in major journals on language teaching and assessment, as well as by an increasing number of scoping or systematic reviews (e.g., Ali et al., 2024; Huang et al., 2022; O’Grady et al., 2026; Wang & Zou, 2025; Wang et al., 2024).

For language education, research centres around the capabilities of genAI which are often perceived as opportunities for language teaching and learning in various ways. Scholars have identified the potential for personalised language instruction, due to the generation of individual tasks and materials (Perez Nuñez, 2023), the ability to provide detailed feedback at scale and in a short time (e.g., Muniandy & Selvanathan, 2024) and the potential role of genAI-based chatbots as conversational and learning partners or as scaffold providers (Dikilitaş et al., 2024). The latter creates opportunities for individualised, flexible and on-demand practice opportunities in and exposure to the target language, which is usually English (e.g., Wang & Zou, 2025).

The new opportunities for AI-generated individualised materials designed at scale can be extended to assessment. In the literature, three affordances are highlighted in relation to assessment, namely producing assessment materials and generating test items at scale, automated scoring and providing timely and individualised feedback (e.g., Mizumoto & Eguchi, 2023). All three are relevant for teachers, potentially

facilitating assessment in all phases of test development. A precondition is that teachers are empowered to harness genAI for these purposes, which involves a solid amount of LAL but also the ability to use genAI for these purposes. We will discuss this aspect in the following section.

The challenges of genAI for language teachers brought up in the literature revolve around the same issues, namely

- overreliance on genAI tools, both on the part of the learners and possibly on the teachers
- academic integrity as a key concern
- bias, equity and resulting problems for quality criteria of assessment like validity or reliability
- data protection
- copyright issues

Many of the above aspects are related to ethics in generative AI, and can be seen as interrelated.

Overreliance on genAI tools can be considered on various levels. On the level of learners uncritically taking over genAI-created content, the danger of reduced creativity and lack of critical thinking is implied, and the learning progress and competence development may be inhibited. For the language learning context, for example, Yan and Huang (2025) emphasise the danger of failing to interrogate AI suggestions for writing, with learners reducing self-editing and revision, thus jeopardising their competence in this area. Overreliance and subsequent uncritical practices in language learning, teaching and assessment are seen as an ethical challenge.

Overreliance on genAI content is connected to hallucinations as inaccurate content. While hallucinations are being reduced due to technological progress, learners still have to be guided by teachers in developing an awareness around the trustworthiness of genAI content (Galaczi & Luckin, 2024), but language teachers need to develop a critical approach themselves, in particular when it comes to developing assessment items for exams in a co-assessment procedure, i.e. co-constructing the item with genAI technology. Adopting genAI feedback as whole texts gives rise to authorship concerns. The question whether a large amount of AI-generated text in an assignment can still be considered a learner's own writing and can be assessed as such, touches on questions of validity and reliability.

This question is connected to the issue of academic integrity as part of AI ethics. The language teachers surveyed by Khabbazbashi et al. (2024) mentioned cheating and plagiarism as their primary challenge. Similarly, Liu (2025) sees gen AI-involved plagiarism as a threat to educational assessment because it undermines the reliability

and validity of assessments. The findings of Liu's interview study with seventeen higher education EFL teachers in China highlighted the negative impact of genAI that teachers perceived, related to increased genAI induced plagiarism and the difficulty of detecting and dealing with it. They consequently saw assessment validity and fairness as potentially compromised (p. 719). As a coping strategy, they identified clear guidelines on AI use by educational institutions to provide guidance and support to teachers. Additionally, despite its limitations, AI detection software was considered to be a useful part of a variety of approaches against AI-involved plagiarism although it is often false positive and can never be fully accurate as technology advances more quickly than the detection software. Another coping strategy that Liu identified in the data was redesigning assessments, either in a way that limits the use of AI or in a fashion that would involve assessment of multiple practical skills like projects or alternative, formative-leaning assessment methods. The former might have repercussions on assessment authenticity as a quality criterion of assessment, but it might also deprive students of opportunities of meaningful learning with AI. The latter requires LAL on the part of the teachers in an area that has been found to be underdeveloped (e.g., Vogt et al., 2020).

Bias as part of AI ethics has received increasing attention as a potential challenge of genAI and language teaching and assessment on various levels. Training data for AI applications may be biased in terms of gender, new cultural or linguistic stereotypes, and risk perpetuating them in biased output and mainstream norms that reinforce societal inequities and undermine social and/or linguistic justice (Lee et al., 2025). For language education, AI models privileging certain linguistic varieties over others constitute a problem e.g., for acknowledging and valuing diverse varieties of English and for multilingual approaches to language teaching and assessment.

Bias in AI-generated assessment content is an equity, diversity and inclusion issue and thus seriously impacts assessment quality. In addition, transparency concerns might be raised by different AI model architectures and training procedures that involve deep learning techniques known as black box (Chuang & Yan, 2025). Although explainable AI has met with increasing interest to mitigate the transparency issue, quality control processes concerning, for example, generalising assessment items or automatic scoring cannot be fully made transparent and might represent an ethical issue.

Lastly, differential socio-economic factors determine access to genAI tools in general and thus widen the digital divide or they create inequalities when some learners have access to free genAI tools only and others to more powerful, paid versions, resulting in an unfair advantage of better equipped learners. This does not only represent an equity issue, but also a validity issue which might compromise assessment (Muniandy & Selvanathan, 2024) due to construct irrelevance.

Copyright issues, in particular concerning output and training data use, and data protection are further challenges to genAI use. Particularly in language assessment,

uploading potentially sensitive learner data for automated scoring might break data protection regulation, especially if free AI versions are used that use the data for further training. To mitigate this issue, APIs have been implemented as solutions, but in the classroom setting, this requires technological knowledge on the part of the teachers.

The requirements that teachers are faced with assessment in the era of AI and the question whether teachers can cope will be broached in the next section.

Requirements for Teachers and the Need for AI-Integrate Teacher LAL

New opportunities regarding genAI -related assessment, generating assessment items at scale, automated scoring and automated, detailed and timely feedback on student work, leads to new assessment-related competences. Language teachers have to acquire these competences if they want to navigate AI-driven educational landscapes, educational language assessment included. Being "AI-ready" is crucial in order to optimise the use of genAI in their work (Galaczi & Luckin, 2024, p. 9) and language assessment. Being "AI-ready"requires

- professional knowledge and teacher agency,
- the human (teacher) in the loop during the entire process of assessment development that includes being able to generate assessment -related content and assessment items with specific prompts,
- a critical evaluation of genAI created output and based on the output,
- instructional decisions that embed output in teaching or assessment in line with curricular requirements.

The requirements for gen AI-related competences in the literature can be classified into specific and general requirements, which are interrelated. In terms of specific requirements, developing prompting skills for language teachers is key, both for teaching and assessment (Moorhouse, 2024). For integrating AI generated feedback into teacher feedback, Chuang and Yan (2025) have identified relevant preconditions, namely language educators providing relevant instructions such as feedback criteria and written texts to the gen AI tool, carefully reviewing the AI-generated output and adapting and modifying the feedback in accordance with learners' needs. For automated scoring, language teachers ideally understand the data sources behind scoring models and monitor the AI scoring outcomes for errors, construct-irrelevant variants, potential bias and data protection as well as privacy issues (e.g., Sawaki et al., 2025). By extension, they have to gauge the quality of genAI-generated language assessment items and assessment types in terms of reliability, validity and fairness as quality criteria. In order to safeguard academic integrity, language teachers have to develop coping strategies such as redesigning assessments (Liu, 2025). They consequently require assessment competences to select and evaluate assessment types, gauge the potential of affordances and challenges with regard to academic integrity. In addition, language teachers need to recognise and mitigate bias in AI-generated output to be used for language assessment. On a more general level, language teachers face the need to integrate AI applications into language learning,

teaching and assessment and thus have to position AI as a part of their expertise (Wu et al., 2025). With regard to language assessment, they have to be empowered to integrate gen AI -assisted assessment tasks with learning on the basis of the critical thinking skills and their capacity to critically evaluate the trustworthiness of AI output. Including an AI component in teacher education programmes, is seen as a logical consequence (Wu et al., 2025). The findings of the language teacher survey by Khabbazbashi et al. (2024) concerning teachers' self-reported training needs revolve around similar aspects, namely general training on teaching learners how to use AI, technology-related training, ensuring academic integrity, and effective prompting strategies.

The aforementioned aspects are grouped in the literature under AI literacy although the concept of AI literacy differs with various authors. Liu (2025) proposes a skills-based approach with AI literacy comprising a) basic knowledge of the working mechanism of AI tools, the functions, limitations and AI ethics and b) skills and effectively using the AI functions in different learning scenarios. Wu et al. (2025) take a more holistic approach and integrate ethics in ethical AI literacy. Ethical AI literacy comprises technical proficiency, critical thinking, ethical awareness as well as identity negotiation. Zhu and Wang (2025) add aspects such as staying informed about latest developments as a requirement for teachers in the generative AI era, although it is questionable how realistic this is for teachers who usually have heavy workloads. Pérez-Paredos et al. (2025) suggest critical AI literacy for students of applied linguistics, future language teachers included.

Focusing on language assessment, the need for integrating AI literacy into the concept of LAL is apparent. Existing LAL competence profiles need to be enhanced by knowledge, practices and skills, principles and attitudes related to gen AI and language assessment. First ideas to conceptually combine LAL and AI literacy have been suggested e.g. by Xu (2024), whose knowledge-based concept of AI assessment literacy could be developed further. Xu's model of AI assessment literacy consists of conceptual knowledge, practical knowledge and technical knowledge. Conceptual knowledge refers to an understanding of assessment principles and theories intersecting with AI theory; practical knowledge concerns hands-on experience in using AI; and technical knowledge involves understanding AI models and how genAI applications work from a technical perspective. This model represents an appropriate first step because it effectively merges LAL and AI literacy but could be differentiated further.

In summary, new conceptions of teacher LAL have to integrate a dynamic AI component, with AI literacy and related requirements representing a moving target and LAL levels being underdeveloped in many educational contexts (Vogt et al., 2020). In this regard, the question whether teachers can keep up with the rapid development of gen AI permeating language teaching and, in particular, assessment, has to be raised. In light of the dynamic change and disruption that gen AI technology is causing,

language teachers cannot be left alone to navigate AI-driven educational landscapes. They can only keep up if they experience a joint effort of policymakers, researchers, test developers, teacher educators and teachers themselves. These stakeholdersneed need to pool their expertise to provide a support system for language teachers concerning AI-integrated LAL development.

Policymakers on various levels as well as educational authorities have to guide teachers on AI in language assessment by providing clear policies on AI use and language assessment. AI literacy components relevant to language teachers in the respective contexts have to be included in teacher education programs and professional development initiatives. Educational authorities and leadership bodies at schools have to provide opportunities and grant teachers time to develop the AI-integrated LAL that capitalises on teacher agency. Professional development initiatives are to empower language teachers by providing them with flexible upscaling options that take into consideration the dynamic development of the field and that do not only target technical developments, but also pedagogical implementation, including broader questions of equity, diversity and inclusion. In addition, policymakers have to take responsibility for tailoring and localising global frameworks like the UNESCO AI competency framework for teachers (2024) in order to be meaningful and relevant to language teachers, thus increasing the chance of being implemented.

Researchers are called on to continue to build the empirical knowledge base related to gen AI involved language assessment, including validity and reliability and fairness related questions of gen AI-enhanced language assessment. Additionally, conceptual work is needed to identify the tenets of AI literacy as part of LAL that is relevant to language teachers and integrate them into assessment and learning tasks, striving for an alignment of learning, teaching and assessment. Researchers should also approach teacher educators as another key stakeholder group and share their insights, for example at (teacher) conferences and organise events such as SIG meetings, explicitly addressing this group.

Test developing companies have heavily invested in AI-based assessment systems and associated quality control systems (Chuang & Yan, 2025). Proprietary research by test developers is not always publicly accessible. In line with the initiative of open science in language assessment (Kremmel & Isbell, 2024), it would be beneficial if relevant insights for language teachers could be shared, thus increasing accessibility of the empirical knowledge base related to quality control questions in large-scale assessment that could be transferred or adapted to CBLA contexts.

And finally, teachers themselves can be invited to share experiences, discuss critical issues and co-construct AI-involved assessment. Online communities of practice allow flexible and dynamic exchange among teachers that are independent of time and place

(Göktürk Sağlam & Sevgi-Sölé, 2023), provided that teachers obtain the time resources necessary for this type of informal professional development.

The overall goal is to provide language teachers with solid support structures that help them to meet the new requirements induced by AI-related disruptions in language assessment, on the basis of a solid LAL basis.

Conclusion and Way Forward

In the present article, I have discussed the affordances and challenges of AI-involved language assessment that language teachers are confronted with in a situation of dynamic change in language assessment landscapes. The various requirements made on language teachers' professionalism have been presented and related to teachers' AI literacy on the one hand and LAL on the other hand. Against the background of the manifold requirements, the need for AI-integrated teacher LAL development was highlighted. In order to avoid teachers bearing the full brunt of the new LAL/AI professional development needs, the tenets of a multi-professional support system for language teachers were suggested to help teachers keep up.

Further conceptual work on LAL heuristics that includes AI technology and its implementation in teacher-led language assessment is required as a research direction that would integrate the two concepts. Xu's (2024) model of AI assessment literacy represents a step towards AI-integrated LAL heuristics that serve the needs of language teachers in the era of AI.

ORCID

https://orcid.org/0000-0001-6019-2655

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

The project has received no particular funding.

CRediT Authorship Contribution Statement

Karin Vogt: Conceptualisation, Methodology, Writing – Original Draft, Writing – Review and Editing

Generative AI Use Disclosure Statement

Generative AI was not used for this paper.

Ethics Declarations

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

World Medical Association (WMA) Declaration of Helsinki – Ethical Principles for Medical Research Involving Human Participants are not applicable for this reflective paper.

Competing Interests

There are no competing interests.

Data Availability

This paper is not based on empirical data that could be shared.

References

Ali, O., Murray, P.A., Momin, M., Dwivedi, Y.K., & Malik, T. (2024). The effects of artificial intelligence applications in educational settings: Challenges and strategies. *Technological Forecasting and Social Change*, *199*, Article 123076. https://doi.org/10.1016/j.techfore.2023.123076

Bøhn, H., & Tsagari, D. (2021). Teacher educators' conceptions of language assessment literacy in Norway. *Journal of Language Teaching and Research, 12*(2), 222–233. https://doi.org/10.17507/jltr.1202.02

Chiu, T. (2025). AI literacy and competency: definitions, frameworks, development and future research directions. *Interactive Learning Environments, 33*(5), 3225-3229, https://doi.org/10.1080/10494820.2025.2514372

Chuang, P., & Yan, X. (2025). Language assessment in the era of generative artificial intelligence: Opportunities, challenges, and future directions. *System, 134*, Article 103846. https://doi.org/10.1016/j.system.2025.103846

Davies, A. (2008). Textbook trends in teaching language testing. *Language Testing*, *25*(3), 327–347. https://doi.org/10.1177/0265532208090156

Dikilitaş, K., Furenes Klippen, M.I., & Keles, S. (2024). A systematic rapid review of empirical research on students' use of ChatGPT in higher education. *Nordic Journal of Systematic Reviews in Education, 2*, 103–125. https://doi.org/10.23865/njsre.v2.6227

Dimova, S., Yan, X., & Ginther, A. (2022). Local tests, local contexts. *Language Testing, 39*(3), 341–354. https://doi.org/10.1177/02655322221092392

Edyburn, D. L. (2020). *Rapid literature review on assistive technology in education: Research report*. Knowledge by Design. https://www.knowledge-by-design.com/ukat/final_report.pdf

Funa, A. A., & Gabay, R. A. E. (2025). Policy guidelines and recommendations on AI use in teaching and learning: A meta-synthesis study. *Social Sciences & Humanities Open, 11*, Article 101221. https://doi.org/10.1016/j.ssaho.2024.101221

Galaczi, E., & Luckin, R. (2024). *Generative AI and language education: Opportunities, challenges and the need for critical perspectives. Cambridge Papers in English Language Education*. Cambridge University Press & Assessment.

Göktürk Sağlam, A. L., & Sevgi-Sölé, E. (2023). Participant evaluation of an online course on language assessment literacy. In A. L. Göktürk Sağlam & E. Sevgi-Sölé (Eds.), *Emerging practices for online language assessment, exams, evaluation, and feedback* (pp. 142–161). IGI Global.

Hannah, L., Jang, E. E., Shah, M., & Gupta, V. (2023). Validity arguments for automated essay scoring of young students' writing traits. *Language Assessment Quarterly, 20*(4–5), 399–420. https://doi.org/10.1080/15434303.2023.2288253

Huang, W., Hew, K. F., & Fryer, L. K. (2022). Chatbots for language learning—Are they really useful? A systematic review of chatbot-supported language learning. *Journal of Computer Assisted Learning*, *38*(1), 237–257. https://doi.org/10.1111/jcal.12610

Jang, E. E., & Sawaki, Y. (2025). Advancing language assessment for teaching and learning in the era of the artificial intelligence (AI) revolution: Promises and challenges. *Language Testing*, *42*(4), 361–368. https://doi.org/10.1177/02655322251348685

Jin, Y., & Fan, J. (2023). Test-taker engagement in AI technology-mediated language assessment. *Language Assessment Quarterly*, *20*(4–5), 488–500. https://doi.org/10.1080/15434303.2023.2291731

Kasneci, E., Sessler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Günnemann, S., Hüllermeier, E., Krusche, S., Kutyniok, G., Michaeli, T., Nerdel, C., Pfeffer, J., Poquet, O., Sailer, M., Schmidt, A., Seidel, T., ... Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences*, *103*, Article 102274. https://doi.org/10.1016/j.lindif.2023.102274

Khabbazbashi, N., Galaczi, E., Allen, H., Lopes, S., Nakatsuhara, F., & Halley, K. (2024). *Exploring the impact of generative AI on language education: Insights from teachers. Cambridge Papers in English Language Education*. Cambridge University Press & Assessment.

Kiryakova, G., & Angelova, N. (2023). ChatGPT—A Challenging Tool for the University Professors in Their Teaching Practice. *Education Sciences*, *13*(10), Article 1056. https://doi.org/10.3390/educsci13101056

Kremmel, B., & Isbell, D. (2024). Open Science practices in language assessment: Introducing the special issue. *Language Testing*, *41*(4), 697-702. https://doi.org/10.1177/02655322241264092

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

Lan, C., & Fan, S. (2019). Developing classroom-based language assessment literacy for in-service EFL teachers: The gaps. *Studies in Educational Evaluation*, *61*, 112–122. https://doi.org/10.1016/j.stueduc.2019.03.003

Lee, S., Jeon, J., McKinley, J., & Rose, H. (2025). Generative AI and English language teaching: A global Englishes perspective. *Annual Review of Applied Linguistics*, *45*, 85–108. https://doi.org/10.1017/S0267190525100184

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

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. (2024). Generative artificial intelligence and ELT. *ELT Journal*, *78*(4), 378–392. https://doi.org/10.1093/elt/ccae032

Muniandy, J., & Selvanathan, M. (2024). ChatGPT, a partnering tool to improve ESL learners' speaking skills: A case study in a public university in Malaysia. *Teaching Public Administration*, *43*(1), 4–20. https://doi.org/10.1177/01447394241230152

Nayernia, A., & Mohebbi, H. (2023). Teachers' online language assessment literacy. In Z. Tajeddin & C. Griffiths (Eds.), *Language education programs: Perspectives on policies and practices* (pp. 121–140). Springer. https://doi.org/10.1007/978-3-031-38754-8_8

Nguyen, H. M., & Goto, D. (2024). Unmasking academic cheating behavior in the artificial intelligence era: Evidence from Vietnamese undergraduates. *Education and Information Technologies*, *29*, 15999–16025. https://doi.org/10.1007/s10639-024-12495-4

Ockey, G. J., Chukharev-Hudilainen, E., & Hirch, R. R. (2023). Assessing interactional competence: ICE versus a human partner. *Language Assessment Quarterly*, *20*(4–5), 377–398. https://doi.org/10.1080/15434303.2023.2237486

O'Grady, S., Göktürk, N., & Rossi, O. (2026). Artificial intelligence in language assessment research: A scoping review. *Research Synthesis in Applied Linguistics*. Advance online publication. https://doi.org/10.1080/29984475.2026.2633389

Otermans, P. C. J., Baines, S., Livingstone, C., & BouĆ, S. B. (2024). Unlocking the Dynamics of Online Team Based Learning: A Comparative Analysis of Student Satisfaction and Engagement across Psychology Modules. *International Journal of Technology in Education and Science*, *8*(3), 481-490.

Pérez-Paredes, P., Neil, C., & Ordoñana-Guillamón, C. (2025). Critical AI literacy for applied linguistics and language education students. *Journal of China Computer-Assisted Language Learning*, *5*(2), 175–214. https://doi.org/10.1515/jccall-2025-0005

Pérez-Núñez, A. (2023). Exploring the potential of generative AI (ChatGPT) for foreign language instruction: Applications and challenges. *Hispania*, *106*, 355–362. https://doi.org/10.1353/hpn.2023.a906568

Pill, J., & Harding, L. (2013). Defining the language assessment literacy gap: Evidence from a parliamentary inquiry. *Language Testing*, *30*(3), 381–402. https://doi.org/10.1177/0265532213480337

Sawaki, Y., Ishii, Y., Yamada, H., & Tokunaga, T. (2025). Examining the consistency of instructor vs. large language model ratings on summary content: Toward checklist-based feedback provision with second language writers. *Language Testing, 42*(4), 447–475. https://doi.org/10.1177/02655322251349217

Stewart, J., Anthony, L., Batty, A. O., Nakamura, K., Nicklin, C., McLean, S., & Tomaru, K. (2025). Can we reliably score meaning recall vocabulary tests using AI? A comparison of human vs. AI scoring. *Computer Assisted Language Learning*, 1–23. https://doi.org/10.1080/09588221.2025.2481397

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

UNESCO. (2024). *AI competency framework for teachers*. UNESCO. https://unesdoc.unesco.org/ark:/48223/pf0000391104

Villa Larenas, S. (2020). *An investigation of the language assessment literacy of teacher educators in Chile: knowledge, practices, learning, beliefs, and context*. [Doctoral Thesis, Lancaster University]. Lancaster University. https://doi.org/10.17635/lancaster/thesis/1012

Villa Larenas, S., & Brunfaut, T. (2024). How do teacher educators learn about and teach language assessment? In B. Baker & L. Taylor (Eds.), *Language assessment literacy and competence: Volume 1: Research and reflections from the field* (pp. 66–87). Cambridge University Press & Assessment. Vogt, K., Tsagari, D., Csépes, I., Green, A., & Sifakis, N. (2020). Linking learners' perspectives on language assessment practices to Teachers' Assessment Literacy Enhancement (TALE): Insights from four European countries. *Language Assessment Quarterly, 17*(4), 410–433. https://doi.org/10.1080/15434303.2020.1776714

Vogt, K., Bøhn, H., & Tsagari, D. (2024). Language assessment literacy. *Language Teaching, 57*(3), 325–340. https://doi.org/10.1017/S0261444824000090

Vogt, K. (2022). Assessment feedback in der unterrichtsbezogenen Leistungsbeurteilung des Fremdsprachenunterrichts [Assessment feedback in classroom-based language assessment]. In E. Burwitz-Melzer, C. Riemer, & L. Schmelter (Eds.), *Feedback beim Lehren und Lernen von Fremdsprachen: Arbeitspapiere der 42. Frühjahrskonferenz zur Erforschung des Fremdsprachenunterrichts* (pp. 215–224). Narr.

Vogt, K. (2025). Generative artificial intelligence applications in foreign language teaching and assessment. In A. Bast-Schneider, C. Eckhardt-Kamps, K.-H. Dammer, & H. Schön (Eds.), *KI und ChatGPT in Bildungskontexten* [*AI and ChatGPT in educational contexts*] (pp. 35–50). Dr. Mattes.

Wang, C., & Zou, B. (2025). AI versus teachers and peers: A comparative study on speaking performance, willingness to communicate, self-perceived communicative competence and foreign language anxiety. *International Review of Applied Linguistics in Language Teaching*. Advance online publication. https://doi.org/10.1515/iral-2024-0310

Wang, S., Wang, F., Zhu, Z., Wang, J., Tran, T., & Du, Z. (2024). Artificial intelligence in education: A systematic literature review. *Expert Systems with Applications, 252,* Article 124167. https://doi.org/10.1016/j.eswa.2024.124167

Wu, J. G., Lee, S., Zhang, D., & Szabo, C. (2025). Cultivating AI literacy in language education: Theoretical discussions and practical applications. *Journal of China Computer-Assisted Language Learning, 5*(2), 169–174. https://doi.org/10.1515/jccall-2025-0023

Yan, X., & Huang, B. H. (2025). Generative AI for the teaching, learning, and assessment of productive skills: An evidence-based approach to understanding its real impact. *TESOL Quarterly, 59*, S5–S18. https://doi.org/10.1002/tesq.70043

Xu, H. (2024). EFL teachers' AI assessment literacy: An exploratory study. In C. A. Chapelle, G. H. Beckett, & J. Ranalli (Eds.), *Exploring artificial intelligence in applied linguistics* (pp. 137–150). Iowa State University Digital Press. https://doi.org/10.31274/isudp.2024.154.09

Zerkouk, M., Mihoubi, M., & Chikhaoui, B. (2025). A comprehensive review of AI-based intelligent tutoring systems: Applications and challenges. Journals of Computers in Education, *arXiv preprint arXiv:2507.18882*. https://doi.org/10.48550/arXiv.2507.18882

Zhang, D., Hoang, T., Pan, S., Hu, Y., Xing, Z., Staples, M., Xu, X., Lu, Q., & Quigley, A. (2023). Test-takers have a say: Understanding the implications of the use of AI in language tests [Preprint]. arXiv. https://doi.org/10.48550/arXiv.2307.09885