Language Testing in Focus: An International Journal

EUROKD

2026 (13), 77–88
https://doi.org/10.32038/ltf.2026.13.05

DeepThink-Mediated Language Assessment Literacy: EFL Teachers' Practices and Insights from China and Cambodia

Wenwen Tian*
School of Foreign Studies, Northwestern Polytechnical University, China
Stephen Louw
Pañña sastra University of Cambodia, Siem Reap, Cambodia

Correspondence
Email: wenwen.tian@nwpu.edu.cn

Abstract
Inspired by Coombe et al.'s (2020) conceptualization of Language Assessment Literacy (LAL) as a recursive process of learning, unlearning, and relearning, we reflect on how teachers' LAL is socially and technologically mediated in using DeepSeek's DeepThink function. Through reflective analysis of Wenwen's supervisory interactions with her MA student in China, and Steve's in-service teacher education practices in Cambodia, this reflective paper contributes to the discussions about LAL in the age of AI. We present our successes and failures with promoting the use of AI as a tool for the development of LAL in teachers and learners. Our experiences with LAL and AI are framed in Coombe et al.'s (2020) recursive cycle of LAL development.

ARTICLE HISTORY
Received: 15 October 2025
Revised: 17 March 2026
Accepted: 25 May 2026

KEYWORDS
DeepThink-Mediated Language Assessment Literacy, Language Assessment Literacy, Language Assessment Literacy in China and Cambodia

How to cite this article (APA 7th Edition):
Tian, W., & Louw, S. (2026). DeepThink-mediated language assessment literacy: EFL teachers' practices and insights from China and Cambodia. *Language Testing in Focus: An International Journal, 13*, 77–88. https://doi.org/10.32038/ltf.2026.13.05

Introduction[1]

Although a growing body of research has examined the pedagogical affordances of AI in higher education (Davoodifard & Eskin, 2024; Sidorkin, 2026), its implications for language assessment literacy (LAL) remain underexplored, particularly in relation to teacher interpretive agency and ethical decision-making. Recent scholarship has moved beyond viewing LAL as a set of technical competencies, instead conceptualizing

[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.

it as a context-sensitive, interpretive, and value-laden practice (Coombe et al., 2020; Crusan et al., 2023). From this perspective, LAL entails not only knowing how to design and administer tests, but also understanding how assessment decisions are constructed, justified, and enacted within specific educational contexts, especially in the age of AI. As Nayernia and Mohebbi (2023) argue, the emergence of online and AI-supported assessment systems requires teachers to develop expanded forms of LAL that integrate technological, pedagogical, and ethical dimensions. These developments shift LAL from a predominantly human-centered activity to a distributed process involving both human and machine reasoning.

Drawing on Coombe et al.'s (2020) framework of LAL as a recursive process of learning, unlearning, and relearning, we explored how DeepSeek's DeepThink function mediates this process. Coombe et al. (2020) frames LAL development in terms of *learning*, in which teachers acquire foundational assessment knowledge, *unlearning*, which entails shedding inappropriate beliefs and practices about LAL, and *relearning* assessment paradigms that are sensitive to local contexts. This framing is relevant to the need for critical self-examination to meet the demands of an AI-mediated context where teachers must continuously calibrate their own epistemic assumptions. Extending this line of reasoning into digital environments, Nayernia and Mohebbi (2023) demonstrate that online LAL requires teachers to navigate additional layers of technological mediation, platform affordance, and interpretive uncertainty, competencies that the current study argues are intensified and rendered newly visible through DeepThink's externalised reasoning chains. We argue that DeepThink operates not merely as a tool, but as a reflective and epistemic scaffold that externalizes reasoning and supports dialogic engagement with assessment practices. In doing so, the paper contributes to emerging discussions on AI and LAL by proposing a human–AI co-mediation model, through which assessment expertise is understood as co-constructed across human cognition, technological mediation, and sociocultural context.

GenAI systems such as ChatGPT and DeepSeek introduce both opportunities and tensions. While these tools can support adaptive assessment and feedback generation, they also risk reinforcing surface-level or procedural approaches if used uncritically. As Crusan et al. (2023) emphasize, assessment literacy is inseparable from pedagogical judgment. Over-reliance on automated outputs may undermine teachers' capacity for principled decision-making. We chose DeepSeek for this study not only for its free accessibility both in China and Cambodia, but because its DeepThink function supports a mode of AI-mediated reflection that aligns with Vygotskian and broader sociocultural conceptions of professional learning.

Operationalizing EFL Teachers' LAL in DeepThink

Drawing on Vygotsky's (1978) Zone of Proximal Development (ZPD), AI can function as a "thinking partner" that supports users just beyond their independent capacity by providing prompts, feedback, and alternative perspectives. In an LAL context, it can

advance assessment thinking through interaction rather than fixed scaffolding. From Inbar-Lourie's (2017) sociocultural perspective, LAL develops through experience, dialogue, and reflection. In this light, DeepThink's real value lies not in the answers it provides, but in how teachers engage with it through questioning, adapting, and reflecting on their fit with specific teaching contexts. Reflection becomes the bridge between AI support and genuine professional learning. As Crusan et al. (2023) note, assessment is not a standalone skill; it is embedded in how teachers design lessons, respond to students, and interpret learning progress. In AI-supported environments, this connection is especially critical, as there is a real risk of favoring efficient outputs over careful pedagogical consideration. Utilizing DeepThink with LAL in this respect, then, is not about making assessment easier, but making it more thoughtful. When teachers engage critically and reflectively, DeepThink can shift from a simple tool to a collaborative partner, helping to sharpen judgment, make informed decisions, and keep assessment grounded in their educational values.

Linking DeepThink to Human–AI Co-Mediation Model

DeepThink situates learning in an AI-mediated ZPD where dialogic reasoning and iterative feedback cycles scaffold teachers' assessment inquiry. This interaction can potentially support teachers in refining diagnostic, evaluative, and decision-making skills. The reflective loop between teacher agency and AI transparency embodies a co-regulative process, aligning with Vygotsky's social constructivism and recent calls for contextualized, practice-based LAL development (Giraldo & Yan, 2025). Consequently, DeepThink transforms assessment literacy into a co-constructed competence, where professional growth emerges from continuous negotiation between human cognition, technological mediation, and educational values. It is important to note, however, that this co-constructed competence is not uniformly distributed across teacher populations. As Crusan et al. (2023) observe in the context of writing instruction, assessment literacy is deeply inflected by teachers' domain-specific pedagogical knowledge, meaning that teachers with stronger subject-area expertise are better positioned to critically interrogate and selectively appropriate AI-generated assessment rationales rather than absorbing them uncritically. This finding adds nuance to the co-mediation model advanced here: it suggests that DeepThink's reflective scaffolding is most productive when teachers bring sufficient pedagogical knowledge to the interaction, a condition that underscores the continued importance of structured professional development as called for by Coombe et al. (2020).

Here, we present our reflections on our experience with DeepThink as a mediation tool in developing teachers' LAL, in both academic supervision and teacher training contexts. First Wenwen, working in China with graduate students, and then Steve in Cambodia, with novice in-service teachers. In both contexts, we used DeepThink as a tool for professional development based on Coombe et al.'s (2020) learn-unlearn-relearn framework. It is worth noting that our reflections are based on our experimentation with DeepThink with a limited sample.

Wenwen: DeepThink in MA Thesis Supervision in China

Since the free-access DeepSeek platform was released in early 2025, I have observed my students using it openly and extensively for her thesis proposal development. However, most seemed unaware of, or simply overlooked the DeepThink feature. Over five months (May–September 2025), I asked one of my post-graduate supervisees to screenshot all her interactions with DeepSeek and submit reflections on how DeepThink's reasoning processes shaped her thinking and influenced her decision-making.

Although this requirement initially made her uneasy, I saw it as a valuable opportunity to understand how AI mediated both my LAL development and her academic growth. Initially, she brought to supervision unedited AI-generated outlines, over-extended templates, and uncritical acceptance of machine-suggested structures. Those issues raised concerns about how I traditionally assessed thesis proposal in terms of research ideas, proposal outline, coherence, relevance, and the integrity of disciplinary knowledge.

In working through these episodes, I came to view my own LAL development as a continuous cycle of learning, unlearning, and relearning (Coombe et al., 2020). This reflection traces that process of renewal, illustrating how the presence of DeepThink in supervision prompted me to rethink assessment design, feedback, and the role of human judgment within GenAI-mediated supervisory practice.

Learning: DeepThink as an Epistemic Scaffold

The initial phase of engagement involved reconceptualizing assessment beyond traditional constructs such as validity, reliability, and fairness toward a more process-oriented and AI-supported practice. DeepThink's reasoning outputs enabled a shift from evaluating textual products to interrogating the underlying logic of student thinking, thereby foregrounding assessment as a diagnostic and interpretive activity. For instance, on one occasion, following the submission of a seemingly flawless draft of the theoretical framework section, most likely largely generated by DeepSeek, I prompted her to evaluate it herself, shifting from a product-oriented stance ("*focusing on structure and language*") to a process-oriented one ("*focusing on learner's critical thinking and decision-making processes*"). In my reflections I wrote:

> *DeepThink's feedback has forced me to notice not only surface structure and language problems, but reasoning gaps. It trained my eyes to see her thinking processes rather than structure and language as most of them are AI-generated.*

Such reflective episodes embody the social constructivist foundation underpinning my work with the student, in which knowledge was co-constructed through dialogic engagement between her, myself as supervisor, and the AI scaffold. Through these iterative dialogues, I learned that assessment is not a terminal judgment on products

but an ongoing conversation. This expanded my personal LAL from a technical domain toward ethical and interpretive awareness (Scarino, 2013). It mirrors what Memarian and Doleck (2023) describe as assessment for learning augmented by AI-mediated feedback, where technology amplifies formative dialogue rather than replaces it. Through interaction with DeepThink, assessment became less about identifying surface-level deficiencies and more about tracing gaps in argumentation, conceptual alignment, and disciplinary reasoning. Importantly, this phase reflects what Coombe et al. (2020) described as learning, but in an expanded sense, that is, not merely acquiring new techniques but reframing the object of assessment itself.

Unlearning: Disrupting Algorithmic Authority and Institutional Habits

A critical turning point in my development emerged when confronted with the temptation to endorse AI-generated structures uncritically. This revealed a deeper issue, that is, the emergence of authority bias, where AI outputs are perceived as inherently valid. Supervising this student exposed my tacit reliance on both institutional assessment routines and AI authority. Earlier in the proposal, she treated DeepThink's suggested outline as a definitive model for her thesis proposal. During one face-to-face supervision, she asked me:

> *Shall I keep all sections suggested by DeepThink as they are all useful?*

My unthinking response was YES as the draft she submitted appeared well-organized and rich in classical and cutting-edge references, but I quickly changed to NO and asked her to explain why she wanted to do so. She justified her decision by arguing that DeepThink provided a clear structure, helped her develop research questions from rough ideas, and offered a model for organizing her writing. These justifications sounded reasonable. However, when I asked her to explain the AI-generated themes in literature review, she could not come up with her own understanding of why and how the literature needed to be structured in that way.

I realized that endorsing AI-generated templates and content without critique contradicted the principles of authentic assessment and learner agency. Instead, I reframed my feedback as follows:

> *This is a starting point, you can use the AI-generated outline and content, but you need to internalize them to be your own research knowledge and skills.*

By withholding an instinctive "yes," I created space for the student to evaluate the relevance, coherence, and necessity of each section, an essential part of developing disciplinary reasoning and writer autonomy. The hesitation also reflects an awareness that DeepThink's suggestions, while generative, are not authoritative. They require human interpretation, contextual tailoring, and alignment with assignment criteria. Thus, stopping myself signaled a shift from AI-as-answer-provider to AI-as-

mediational tool, positioning the teacher not as an approver of AI output, but as a facilitator of reflective decision-making.

This exchange necessitated a process of unlearning at two interrelated levels. Cognitively, it involved resisting the impulse to accept AI-generated coherence as inherently pedagogically appropriate, thereby challenging assumptions about the validity and authority of algorithmic outputs. Institutionally, it required questioning entrenched norms of thesis assessment shaped by standardized expectations and long-standing evaluative conventions.

Relearning: Toward Dialogic and Co-constructed Assessment

The relearning phase involved repositioning DeepThink as a co-reflective interlocutor, enabling a dialogic form of assessment where decisions are jointly constructed and justified. Rather than treating AI outputs as solutions, they became provocations for reflection, prompting me to relearn how to assess my student's reasoned choices. Regarding AI use, I explicitly stated to my student early that: "*You can use AI for brainstorming, but you must keep records of every use in an appendix. The thinking and analysis must be yours.*"

When my student submitted her first draft, I returned it with marginal comments like:

> "*Your framing of 'AI as a co-supervisor' is too provocative. Write your reflection on how and why you put it in that way.*"
> "*You claimed you used AI critically, but where is your evidence? You haven't defined 'critical'.*"

My student addressed my comments. The second draft was much improved, but she deleted her original "provocative" ideas because she was afraid of appearing unscholarly. At that point, I realized my rigorous process-based approach has unintentionally silenced my student's authentic voice. To address this issue, I invited her to co-construct an AI-mediated assessment framework. Together, we developed shared criteria, replaced corrective feedback with reflective prompts, and began maintaining parallel logs of our AI use and assessment struggles, thus shifting from correcting errors and identifying problems to exploring uncertainties, with both parties treating messy, unauthorized AI practices as legitimate data rather than ethical violations. With this re-orientation to co-assessment (see Butler, 2022), supervision became a site of a dialogic space for relearning rigorous assessment. In my supervisory practice, I relearned that my role was not to prohibit AI but to contextualize it, to ensure that students engage with AI reflectively rather than dependently.

Steve: DeepThink in Teacher Development

Learning: Recognizing Gaps and Emerging Practices

My context is a private K–12 school in Cambodia. Our language teaching staff includes twelve novice Khmer-national English teachers who work with the young learners. All

twelve are also undergraduates in TESOL programmes at local universities. Although their degree programmes include modules on testing and assessment, I have found persistent gaps in their LAL, noticeable in their construction of quarterly summative tests, which frequently require extensive revision.

In recent months, I spotted several teachers using AI to design tests. I took this as a positive development, but also noticed that they were doing so furtively, apparently nervous that AI was, perhaps, not allowed or considered a kind of 'cheating'. These attitudes are prevalent at some schools and universities in Cambodia. However, the AI-generated tests submitted were unedited, uninspiring, repetitive, and often of questionable validity. It became clear that the presence of AI exposed rather than resolved existing gaps in teachers' assessment literacy. I decided it was necessary for me to make explicit to the teachers that I view AI as a beneficial tool for supporting teachers, and at the same time I could present an AI tool to promote teachers' professional development.

At this point, my focus was exclusively based on my confidence that a tool like DeepThink could be a catalyst for teachers' unlearning and relearning of their LAL. There was, as yet, no consideration of my own development through Coombe et al.'s (2020) LAL development framework. I will return to my own trajectory shortly.

Unlearning: From Product-Oriented Use to Process Awareness

For the purpose of helping the teachers expand their LAL repertoire, DeepSeek's DeepThink function appeared promising. DeepThink's explicit chain-of-thought response, I felt, might act as a visible expert pathway through the decision-making process involved in assessment design. I saw this as a chance to support a process of unlearning: encouraging teachers to move away from disembodied and pedestrian tests the teachers were encouraged to design in their undergraduate course, and the passive acceptance of AI-generated tests. DeepThink's visible reasoning aligns with the cognitive apprenticeship model of "making thinking visible" (Collins et al., 1991), offering worked examples that can be unpacked, discussed, and adapted. My goal was to reframe AI not as a test-writing engine, but as a scaffold for analysing why particular test decisions are made and the underlying reasoning for these decisions.

Initially, I worried that DeepThink's reasoning would overwhelm teachers with low English proficiency. Some experimentation showed that DeepThink produces its reasoning output in the language of prompt, allowing for a response in Khmer. I first piloted DeepThink with four teachers, two experienced and two novices, using two tasks. The first asked DeepThink to design a test; the reasoning was given in Khmer and the test output in English. The second task asked DeepThink to improve a poorly designed test.

Disappointingly, reflections from the pilot focused largely on the test outputs and not on the DeepThink reasoning process. While the teachers described DeepSeek as

"useful" and "interesting," only one commented on the processes in test design. In her reflections, she writes:

> *I learned that even simple tasks, like True/False questions, require careful planning to make sure they are clear, fair, and directly linked to the text.*

I attributed the poor results of the pilot on my delivery of DeepThink's usefulness, perhaps having made too many assumptions about the teachers' existing confidence in using AI.

I did not, unfortunately, unpack the issue any further, and used these intuitions to contextualize and develop a DeepThink workshop. The workshop, I hoped, would promote guided co-design of assessments using DeepThink and thus begin a process of 'unlearning' in the teachers' LAL. I structured a series of activities that required teachers to examine DeepThink's reasoning, identify the test design principles it used, and collaboratively refine assessment items with AI. The workshop lasted 90 minutes. At the end of the workshop, I believed the teachers to be adequately equipped to engage with DeepThink independently to activate the unlearning and relearning process. I looked forward to seeing teachers taking a process, rather than product, approach.

At this point, my focus was still on the teachers' unlearning and relearning. While I understood that little of the LAL mislearning could be dealt with directly in only 90 minutes, I was confident that the introduction of the AI tool would continue this intervention in directions appropriate for each teacher.

Relearning: Towards Guided Co-design and Sustainable LAL Development

When assessing the impact of in-service training, I follow Guskey's (2000) five-level evaluation model. This includes an evaluation of participants' reactions, knowledge acquisition, change in the organization, teacher uptake, and impact on student learning. During the workshop, the teachers responded positively, seemed to grow in confidence with their use of DeepThink with test design, and expressed relief that their use of AI was not disapproved of by the department. I took this to indicate that the workshop had achieved levels one through three. In follow up conversation with teachers, however, I was (again) disappointed by their limited uptake of a process-oriented usage of AI. Teachers reverted to using it primarily for generating test content rather than for guided decision-making. The modelling I introduced in the workshop did not transfer into independent practice and test submissions returned to being mechanical, repetitive and disconnected from the learners' lived experiences.

The workshop intervention was a failure. It is at this point that I understood the unlearning and relearning could not thrive outside the structured settings of the workshop. The realization became a source for my own development as a teacher educator. My reflection on this insight started with a careful interrogation of the

workshop content, but when it became clear that the problem lay in the transfer, not the knowledge itself, I understood that I needed to explore the larger context in which the workshop took place, both in its lead-up and its follow-through. Rather than the teachers engaging in relearning, I found myself engaging in deep consideration of the need for a reorientation in my own relearning of pedagogical approaches to LAL development. Coombe et al. (2020) identify the persistent gap between assessment training and classroom practice, noting that LAL development requires sustained engagement that connects theory to context. I had set an overly ambitious goal expecting independent LAL development to be catalysed and sustained through interaction with an AI tool when the teachers lacked confidence in both LAL and AI. The approaches that had served me in introducing other new practices to teachers were not transferable here. My exploration of the reasons for the lack of uptake led to authors such as Nemati et al. (2025) and Crusan et al. (2023) who argue in favor of sustained approaches to LAL development. I am yet to discover how exactly to achieve this sustained support for the growth of the teachers' LAL with DeepThink, and so my relearning continues as a form of continuous professional development in this regard. Considering the various constraints of time on both myself and the teachers in our school context, the collaboration between the school management, AI and the teachers to improve assessment practices and promote LAL remains part of my own relearning as a teacher educator.

Discussion

Our experience of AI highlighted a number of problems in its application. For Wenwen in China these included students' uncritical view of its output and an alarming reliance on it as a source of direction, to the point that she doubted her own value as a research supervisor. For Steve, in the context of teacher management in Cambodia, a different set of concerns arose: teachers' uncertainty about its use and fear of reprisals for using it from the school management. Though our contexts differ, we were both interested in the possibility of using DeepThink to move from a product to process orientation and make it a tool for mediating the development of LAL.

We approached our investigation from a Vygotskyan perspective (Vygotsky, 1978) of AI as a tool for mediating learning. To inspect this mediation process in action, we used Coombe et al.'s (2020) learn-unlearn-relearn framework. Following this framework, we inspected the value of the mediation tool in terms of how it highlighted shortcomings in learning and facilitated the process of unlearning and relearning.

The outcomes of the two cases are instructive. Since Wenwen's supervision was through a one-to-one relationship and the influence of AI could be tracked over multiple sessions, it created a sustained and iterative dialogue between AI, the supervisee, and with the supervisor's own LAL development. By contrast, the one-off nature of Steve's intervention did not lead to this depth of learning. The episodic nature of Steve's AI workshop and poor teacher uptake is explained in Coombe et al.'s framing of LAL development. Learn-unlearn-relearn is not a linear experience, but a

dialogue which iterates forward and backward through the practitioner's LAL experience. In Steve's case, he mistakenly made the assumption that such dialogue would automatically become part of the teachers' individual dialogue with AI.

The lack of LAL development in Steve's context provides an insight into AI's effectiveness as a mediational affordance. From a Vygotskian perspective, both unlearning and relearning require careful scaffolding and then a gradual release toward independence. This is particularly pertinent in the context of the complexity of LAL, and with a tool as easily misused as AI. In both cases, the key was not the tool itself but how the student/teachers could be led to using it effectively. The move from a focus on process rather than product using AI demands scaffolding, support in how to read DeepThink output, and guidance in engineering prompts away from a 'do for me' to a 'explain to me' mindset.

As Wenwen's supervisory reflections show, interaction with DeepThink externalized tacit reasoning and deepened conceptual understanding, supporting Giraldo and Yan's (2025) argument that LAL develops through authentic, practice-based inquiry. In Steve's case, it becomes clear that the possible benefits of AI as a tool for professional development, while exciting, are not sufficient to sustain teacher's interest, and need to be incorporated into a systematic professional development program.

Conclusion

This reflective paper demonstrates how DeepSeek's DeepThink function, an AI-powered dialogic reasoning mechanism, mediates the development of LAL among EFL teachers and supervisors in China and Cambodia. The insight we have taken from our investigation is that the key is not AI, but the need to nurture its use and understand its usefulness. DeepThink as a tool is not yet ideal for mediation in LAL. For instance, DeepThink's responses can be verbose, which deterred some of Steve's teachers from engaging with it. In addition, AI functions like DeepThink are limited as a tool for professional development if users simply default to their need to get their job done as efficiently as possible. What is needed, then, is a systematic strategy to introduce the tool, but also wean teachers from their product-oriented habits through sustained coaching. In this respect, we propose that procedures and methods for using AI need explicit pre-teaching and scaffolded follow up coaching. Both Wenwen and Steve presented the AI procedure, but Steve failed in following it up with continued support in reading, interpreting, critiquing and reflecting on AI output.

Our study provides a very limited insight into the exciting but complicated question of the role of AI as a sociocultural mediation tool in facilitating unlearning and relearning in the context of LAL. That we were engaged with this question at a personal level in our professional practice inspired our decision to reflect carefully on and report on our explorations. In spite of the small sample and lack of reportable data, our findings may be useful in guiding a larger and more systematic study of the interaction between the

use of prior knowledge, scaffolded AI use and long-term coaching as a means of facilitating unlearning and relearning in the context of LAL.

ORCID

https://orcid.org/0000-0002-7601-5757

https://orcid.org/0000-0003-0738-496X

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

The authors would like to express their sincere gratitude to the editor and reviewers of this manuscript for their valuable time and constructive feedback.

Funding

This research received no external funding.

CRediT Authorship Contribution Statement

Wenwen Tian: Conceptualization, Methodology, Investigation, Data Curation, Writing – Original Draft Writing, Review & Editing

Stephen Louw: Conceptualization, Methodology, Investigation, Data Curation, Writing – Original Draft Writing , Review & Editing

Generative AI Use Disclosure Statement

During manuscript preparation, the authors used ChatGPT or DeepSeek to improve clarity and coherence of author-developed content. The AI tool was not used to generate research data, conduct statistical analysis, fabricate references, or interpret findings. All research design, data collection, analysis, interpretation, and final academic decisions were made solely by the authors.

Ethics Declarations

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

This study was conducted in accordance with the ethical principles of the World Medical Association Declaration of Helsinki. Retrospective ethical approval was obtained from Institutional Review Board (IRB), School of Foreign Studies, Northwestern Polytechnical University, China (Approval Number: IRB-2026-001). All participants provided informed written consent prior to participation, and participation was voluntary.

Competing Interests

The authors declare that they have no competing interests.

Data Availability

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

References

Butler, Y. G. (2022). Exploring the idea of "assessment with learners" in discussing language assessment literacy. *Language Teaching Research Quarterly, 29*, 92–101. https://doi.org/10.32038/ltrq.2022.29.06

Collins, A., Brown, J. S., & Holum, A. (1991). Cognitive apprenticeship: Making thinking visible. *American Educator, 15*(3), 6–11, 38–46.

Coombe, C., Vafadar, H., & Mohebbi, H. (2020). Language assessment literacy: What do we need to learn, unlearn, and relearn? *Language Testing in Asia, 10*(1), Article 3. https://doi.org/10.1186/s40468-020-00101-6

Crusan, D., Panahi, A., & Mohebbi, H. (2023). Teachers' pedagogical knowledge and assessment literacy. In H. Mohebbi & Y. Wang (Eds.), *Insights into teaching and learning writing: A practical guide for early-career teachers* (pp. 45–55). Castledown. https://doi.org/10.29140/9781914291159-4

Davoodifard, M., & Eskin, D. (2024). Exploring the role of generative AI in second language education: Insights for instruction, learning, and assessment: Introduction to the forum. *Studies in Applied Linguistics & TESOL, 24*(1). https://doi.org/10.52214/salt.v24i1.12863

Giraldo, F., & Yan, X. (2025). Evaluating an online assessment course: Teachers' voices on their language assessment literacy. *PROFILE: Issues in Teachers' Professional Development, 27*(1), 97–114. https://doi.org/10.15446/profile.v27n1.113304

Guskey, T. R. (2000). *Evaluating professional development* (2nd ed.). Corwin.

Inbar-Lourie, O. (2017). Language assessment literacy. In E. Shohamy, I. G. Or, & S. May (Eds.), *Language testing and assessment* (3rd ed., pp. 257–268). Springer. https://doi.org/10.1007/978-3-319-02261-1_19

Memarian, B., & Doleck, T. (2023). A review of assessment for learning with artificial intelligence. *Computers in Human Behavior: Artificial Humans, 2*(1), Article 100040. https://doi.org/10.1016/j.chbah.2023.100040

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

Nemati, F., Bagheri, B., & Abbasian, G. R. (2025). An investigation into learning-oriented assessment literacy and effective teaching in Iranian EFL context. *Language Teaching Research Quarterly, 49*, 113–132. https://doi.org/10.32038/ltrq.2025.49.06

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

Sidorkin, A. M. (2026). Leapfrogging effect hypothesis: Generative AI as a permanent scaffold in higher education. *Social Sciences & Humanities Open, 13*, Article 102958. https://doi.org/10.1016/j.ssaho.2026.102958

Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Harvard University Press.