

“Respond to Change with the Unchanging”: A Second Language Teacher’s Reflective Journey in Transforming Language Assessment Literacy in the Digital Era

Xiaoli Su

College of Language Intelligence (College of General Education), Sichuan International Studies University, Chongqing, China

Correspondence

Email: suxiaoli@sisu.edu.cn

Abstract

The rapid advancement of digital technologies has transformed language assessment in the past several decades, bringing both benefits and challenges for second language (L2) teachers, who need to advance their language assessment literacy (LAL) in response to the drastic changes. In this reflective essay, drawing on ancient Chinese philosophy, particularly the Shu (the Technique) and Dao (The Way) binary, I first narrate my LAL development from 2010 to 2022, from a pre-service teacher to an in-service teacher-assessor, by emphasizing how my assessment practices (Shu) have changed in response to digital technologies. Then, I reflect on the unchanging Dao of my LAL crystallized during my PhD studies from 2022 to 2025 as a teacher-researcher investigating teachers’ LAL, namely, using language assessment to achieve students’ holistic development. Finally, I advocate a vision in which L2 teachers can respond to the constantly evolving technological environment by grounding their future technology-assisted assessment practices in the unchanging principles of the Dao through human-AI collaboration.

ARTICLE HISTORY

Received: 21 October 2025

Revised: 17 March 2026

Accepted: 13 May 2026

KEYWORDS

Language Assessment
Literacy, Language
Assessment, Second
Language Teacher, Reflective
Practice, Digital Assessment

How to cite this article (APA 7th Edition):

Su, X. (2026). “Respond to change with the unchanging”: A second language teacher’s reflective journey in transforming language assessment literacy in the digital era. *Language Testing in Focus: An International Journal*, 13, 89–104. <https://doi.org/10.32038/ltf.2026.13.06>

Introduction¹

The maxim “change is the only constant,” captured centuries ago by the Greek philosopher Heraclitus and also the core idea of the Chinese Yi Jing, or the Book of Change, has never been more true for the field of language assessment, which has been influenced by digital transformation in recent decades. Language assessment has undergone a transition from paper-and-pencil formats to computer-assisted and, most recently, AI-enabled assessments (Chen et al., 2025; Eedelouei, 2026; Garrett, 1991; Goh & Aryadoust, 2025; Jamieson, 2005; Ma et al., 2026; Ockey, 2009). This technological evolution has posed numerous challenges and ethical dilemmas for second language (L2) teachers, who must capitalize on new digital tools in their classroom assessment (Su & Lee, 2024a). Teachers wonder what types of assessment can facilitate students’ learning, as most assignments can be completed with the click of a button. It is now possible to create not only written assignments but also more complex multimodal assessments due to the improved functions of Generative AI (GenAI), such as text-to-speech and image generation (Goh & Aryadoust, 2025; Goulart et al., 2024; Jiang et al., 2024). Paradoxically, language teachers are also frequent adopters of GenAI for assessment, leveraging it for prompt, detailed feedback and workload reduction (Barnes & Tour, 2025). As a result, L2 teachers must update their language assessment literacy (LAL) to maximize the benefits and minimize the downsides of these new technologies.

According to Scarino (2013), LAL is not a universal constant but is shaped by each teacher’s unique context, beliefs, and theoretical leanings. She emphasizes that self-awareness, or “interpretive frameworks,” is crucial for language teachers’ professional development (p. 322). This call for self-awareness resonates with ancient wisdom from both East and West. The Greek proverb “Know Thyself” and Laozi’s teaching, “Knowing others is intelligence; knowing yourself is true wisdom,” both regard self-reflection and introspection as the highest form of understanding (Zu, 2022).

In response to Scarino’s call, I reflect on my evolving LAL over the past 15 years in this essay, a period that saw my transition from a pre-service teacher to an in-service teacher-assessor and finally to a teacher-researcher. This reflective journey is framed by the dynamic interplay of the Technique (术, Shu) and the Way (道, Dao), two terms from ancient Chinese philosophy (Hu, 2014). Shu represents the flexible application of knowledge, here meaning the changing “how” of assessment to adapt to new digital situations, as in the Chinese proverb “to change according to circumstances” (随机应变, Sui Ji Ying Bian). In contrast, Dao signifies the underlying principles, the unchanging “why” we assess our language learners, as captured by the proverb “to respond to change with the unchanging” (以不变应万变, Yi Bu Bian Ying Wan Bian).

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 should be noted that in this reflective essay, I draw on autobiographical memory, which is inherently reconstructive (Neisser, 1994), so my goal is not absolute factual verifiability but “verisimilitude,” namely, a narrative that is authentic to my experience and believable to the reader (Amsterdam & Bruner, 2000, p. 30). This reflection is more than mere recollection; it is an act of defining my identity as a language teacher-assessor and LAL researcher.

The essay is structured as follows: first, I will narrate my LAL development from 2010 to 2022, emphasizing how my assessment practices have changed in response to digital technologies. Next, I will reflect on the unchanging Dao of my LAL, a set of core principles crystallized during my PhD studies from 2022 to 2025. Finally, I will sketch my vision for future L2 teachers’ LAL.

The Changing Shu (Technique) of LAL

The core idea of the Yi Jing is that the conflict between Yin and Yang causes change in the world. This conflict, however, is not like Western notions of dualism, which connote an absolute separation of two contradictions (Chin et al., 2018). Yin and Yang are relational opposites rather than mutually exclusive; Yin in one regard may be Yang in another, and neither exists in isolation from the other (Lloyd, 1996). As for the change in my LAL, my developmental trajectory has been shaped by the relational conflict between my expertise as a human teacher (Yin) and the power of evolving technologies (Yang). I use Yin to refer to my human strengths because the soft, intuitive “carbon life” is characterized by social-emotional connection, and use Yang to represent technology, as the hard, analytical, logic-driven “silicon life” has become a defining force in modern society.

Pre-Service Teacher: Initial Exploration with Technology-Assisted Language Assessment

As a Chinese student accustomed to countless tests, I primarily understood language assessment as summative, large-scale language testing before my pre-service teacher training in Beijing (2010–2013). I equated language assessment with constructing tests and was unfamiliar with the term “assessment literacy”, though the term was coined back in the 1990s by Stiggins (1991). In retrospect, I viewed it as a component of pedagogical competence, aligning with Black and Wiliam’s (2018) argument that assessment is part of pedagogy.

After taking L2 pedagogy classes, I learned the distinction between formative and summative assessment (Black & Wiliam, 1998; Black & Wiliam, 2009). For formative assessment, I remember practicing how to provide written corrective feedback, using consistent symbols to mark vocabulary and grammatical errors in students’ essays. As a teaching assistant for an English writing class for undergraduates, I helped a professor give feedback on multiple drafts, following the process writing approach (Hyland, 2003). All marking was manual, with no digital technology involved.

As a pre-service teacher still in the “apprenticeship of observation” (Lortie, 1975), I also observed how my applied linguistics professors used formative assessment. They adopted some digital technologies facilitated by Web 2.0 technologies that emerged after 2004 (O’Reilly, 2007), defined by user participation and interaction, such as the online learning management system like the discussion forum on the Blackboard learning system (Seethamraju, 2014). I remember that a renowned professor posted a question for us to discuss language acquisition theories and L2 pedagogy each week. I participated in the discussion with great passion, and I was more thrilled and energized when I received the professor’s feedback. Looking back, this motivating and make-me-visible experience became the starting point of my PhD topic, emphasizing the emotional dimension (Yin) of teachers’ LAL in technology-assisted learning environments (Yang) like blended learning (Su & Lee, 2024a; Su & Lee, 2024b).

Unfortunately, I was not trained on how to use these Web 2.0 technologies in designing, implementing, and evaluating language assessment practices, partly due to the lack of textbooks on technology-assisted language education at that time. Despite the lack of training, I took the initiative to explore using digital tools to assess my students. In 2011, I was fortunate to have an opportunity to teach English speaking and listening skills to a class of 30 undergraduate students. I tried different digital technologies to engage my students. Once, I assigned my students to create a music podcast in which they recorded their interpretation of a favorite music piece and emailed it to me; I then provided feedback through email. I chose some excellent works and played the recordings in class. Due to the limited capabilities of technology back then, I was unable to ask students to do peer feedback on each other’s podcasts, which would have been a better learning opportunity. Despite this, I could feel that students liked the multimodal assignment more than a traditional pen-and-pencil assignment, as they were passionate about promoting their favorite pop stars to me. The assessment experience of teaching these students with digital technologies has also contributed to my later PhD studies, as digital and technological knowledge was conceptualized as an important dimension of teachers’ LAL for blended learning environments (Su et al., 2025).

In-Service Teachers: Riding the Digital Waves

After completing my Master’s in applied linguistics, I began teaching English at a university in western China in 2013. This was the same year China experienced the Massive Online Open Course (MOOC) boom. Against this backdrop, my university and department promoted MOOC creation, and I became a key member of a design team, a role I was naturally drawn to and also encouraged to take. One of the most significant challenges of the MOOC movement was the low completion rate (Hew & Cheung, 2014). To address this, our team integrated built-in assessment tasks, including unit tests, discussion questions, and writing assignments. At the time, all writing assignments had to be manually graded and provided feedback by faculty. During the COVID-19 pandemic, our MOOC became a hit, as it was the only English course for

non-English majors² based on a widely adopted textbook, so it was recommended by teachers from other universities to students. However, this success created an overwhelming workload. Without the aid of GenAI, grading thousands of essays was a significant challenge for L2 teachers in my department. Preoccupied with the sheer volume of grading, the feedback we could offer was general rather than concrete and specific.

For students from our university, we adopted a Small Private Online Course (SPOC) model, integrating our MOOC content with face-to-face teaching. In this blended learning environment, I employed diverse digital assessment tools both inside and outside the classroom. In class, I utilized Rain Classroom, an app developed by Tsinghua University to facilitate interaction (Quadir et al., 2022). This tool enabled me to embed multiple-choice or true/false questions into my slides for students to answer on their smartphones. I could also randomly select students to answer questions, adding an element of surprise that some students enjoyed (and others may not have). My colleagues and I found these features especially useful for encouraging participation among our often-reticent Chinese students. To assess learning from the online MOOC videos, I used WJX, one of China's most popular survey tools, to conduct quizzes on vocabulary and grammar. I also used QQ group chats for activities inside and outside the classroom, where students could instantly share group projects such as drawings or poems for others to see. For collaborative group projects, my colleagues and I used Tencent Docs, a platform similar to Google Docs, allowing students to collaboratively create documents seamlessly.

Outside the classroom, I used tools such as UNIPUS, an AI-powered language learning platform designed by Foreign Language Teaching and Research Press, to post learning tasks, initiate discussion topics, and create written or oral assignments (Wang et al., 2024). Despite the availability of automated grading features, I harbored mixed feelings towards them. For example, I asked students to submit their reading journals and essays on Pigai, an automated writing evaluation system (Yao, 2021). At my busiest, I taught up to six classes with over 200 students in a single semester, so I appreciated the time saved, especially when detailed feedback on reading journals was not necessary. Pigai also offered a peer review function, randomly pairing students to provide feedback on each other's work. I could identify excellent submissions and share them as model texts, and I could track how many times students revised their writing to improve their scores, something many Chinese millennial students care deeply about (A few students can revise a single essay over 40 times!). However, students frequently complained about the platform's feedback quality. Pigai, without being powered by GenAI, sometimes misidentified grammatical or collocation errors, and students expected me to comment on the content and logic of their writing.

² Non-English majors in Chinese universities are students studying subjects like education, engineering, business, or medicine who are required to take mandatory English language courses for 2 to 4 semesters as part of their degree requirements, representing the vast majority of English language learners in Chinese higher education.

However, the system was limited in its ability to provide personalized, encouraging feedback, which somewhat reduced its pedagogical effectiveness.

After teaching for five years, I was funded by the Chinese Scholarship Council from December 2018 to December 2019 to conduct research at Ludwig-Maximilians-Universität München in Germany. This unique mobility experience made me seriously think more critically about the digital dimension of assessment, after conducting joint research with my host professor and colleagues comparing English teachers' digital competence between China and Europe (Su et al., 2021; Lütge et al., 2019; Lütge et al., 2021). In the survey we designed to assess pre-service teachers' digital competence, informed by the European Digital Competence Framework for Educators (DigCompEdu) developed by the European Council (Redecker, 2017), we included a dimension "assessment" including three items, "I can use digital technologies to assess English learning", "I can analyze digital evidence on learners' progress critically" and "I can use digital technologies to provide targeted feedback to learners" (Lütge et al., 2021, p.45). These items partially mirror my initial comprehension of digital assessment, contributing to my current research focus on L2 teachers' LAL in digital environments.

I returned to China from Germany on December 15, 2019, and soon after, the COVID-19 pandemic began. As I taught English lessons remotely using new online conferencing technologies during the repeated quarantines, all face-to-face sessions were impossible for a sustained period during the pandemic. Assessment was extremely challenging in remote teaching, as I could not force my typically shy Chinese students to turn on their cameras. Without facial cues, I did not know whether students had understood me or not, even if I did come up with some ways to add interaction, like inviting them to use the chatbox to type in their answers or responses. I felt perplexed and expended considerable emotional labor, much like many language teachers worldwide (Liu et al., 2023; Nyanjom & Naylor, 2020). Many questions emerged at that time: *How do I engage and motivate my students in this online environment? How do I combine online and in-person assessment, given that the quarantine was still on and off? How do I make sure the students finished their home assignments by themselves?*

I felt something was missing during the pandemic, as I was too preoccupied with experimenting with new, cutting-edge digital technologies without critically evaluating their shortcomings. During my eight years as an in-service teacher, my understanding of assessment had already evolved from simple testing to a more social and collaborative practice, but my understanding of the social and collaborative nature of language assessment was not yet visible to me, as I did not have sufficient time to read the relevant literature and think more deeply about it. I eagerly explored how to combine diverse technologies to achieve "assessment for learning" and "assessment as learning," rather than just "assessment of learning" (Earl, 2012; Wiliam, 2011). I believed technology could create a complete profile of a student's learning process, and

I used it to shift my role from the sole judge to a facilitator who invited students into the assessment process. It was time to change my identity from a teacher-assessor to a researcher. I needed time to further understand what teachers can do to use assessment to promote students' learning, which I considered the Dao of LAL.

The Unchanging Dao (the Way) of LAL

In Chinese philosophy, Dao refers to the fundamental principles and holistic wisdom underlying an action. During my three-year PhD studies, my understanding of the unchanging Dao of LAL has crystallized around its most human and unquantifiable elements: understanding the learner as a whole person and guiding their linguistic, intellectual, and emotional growth through language assessment. This narrative recounts my journey in pursuing the Dao.

After eight years of wholeheartedly immersing myself in this teacher-assessor work by implementing digital tools and leading assessment reforms, I began to experience a sense of digital burnout, as a number of language teachers reported in previous studies (e.g. Kurt, 2023). When younger teachers asked for advice I would showcase my “innovative” uses of technology, yet a nagging self-doubt had taken root. I was confronted by fundamental questions: *As technology accelerates, how can I keep up with the latest developments in language assessment? Amidst all the technologically driven changes, what essential principles of LAL remain unchanged? If I were to train new teachers' LAL, what should be the core of that training?*

With these questions lingering, I decided to pause my teaching career to seek deeper reflection in the post-pandemic era. In August 2022, I began my PhD at a university in Hong Kong, focusing on teachers' LAL for blended learning. My manuscript-based PhD thesis allowed me to explore the topic formally through three papers: conceptualizing L2 teachers' blended assessment literacy (L2TBAL, Su et al., 2025), developing and validating a scale to measure it (Su & Lee, 2024a), and exploring teachers' emotion regulation strategies through a case study (Su & Lee, 2024b). A side project is a narrative review paper on LAL through the lens of Habermas's three cognitive interests that allowed me to step back and critically examine LAL (Su et al., 2024). Immersed in the extensive literature on LAL, teacher assessment literacy (TAL), and digital assessment (e.g. Berry et al., 2019; Coombe et al., 2020; Davies, 2008; DeLuca et al., 2015; Eyal, 2012; Fulcher, 2012; Kremmel & Harding, 2019; Levi & Inbar-Lourie, 2019; Pastore & Andrade, 2019; Taylor, 2009; Xu & Brown, 2016), I found my experiences resonating with the voices of teachers globally in the literature and also the teacher participants in my research projects. I connected theories and findings with my decade of practice and was inspired by ingenious assessment practices, both with and without technology (e.g. Barnes et al., 2017; Cui et al., 2025; Gan & Lam, 2023; Nayernia & Mohebbi, 2023). Free from a heavy teaching workload, I could view from a theoretical vantage point and reflect on my role as a language teacher and classroom assessor.

Just as my perspective was solidifying, the emergence of ChatGPT in late 2022 introduced a new existential challenge. It appears that AI has the potential to disrupt assessment more than any other aspect of education. As I attended countless AI seminars during my PhD studies, the message was clear: AI is here to stay, and teachers who use it will replace those who don't. This process created a profound identity crisis for me at the beginning, because GenAI was hailed as a tireless, emotionally neutral, and objective assessor. For example, it can provide concrete feedback on writing, logic, and organization at a level previously impossible for automated systems (Lee & Moore, 2024). This amplified my long existed professional "inferiority complex" as a non-native English-speaking teacher compared with my native English-speaking colleagues (Medgyes, 1999, p.41). Worse, public discourse was filled with the declining value of language learning and the potential AI displacement of language teachers (Liu & Liu, 2025).

The crisis prompted my turning to educational philosophy for insights, or Dao, drawing on thinkers like Gert Biesta's (2020) argument that three domains of educational purpose are qualification (knowledge, skills, and abilities that education provides to students), socialization (helping students become members of a particular society by introducing existing social and cultural norms), and subjectification or individuation (helping students think and act for themselves, rather than simply conforming to the expectations of society). He criticized contemporary education as it is "characterized by a rather single-minded focus on qualification and socialization: fast and furious rather than slow and with a degree of patience." (Biesta, 2020, p. 98).

This has provided great insights for my understanding of LAL: if we cede the right of language assessment entirely to AI and focus only on achieving "fast and furious" language learning results greater efficiency, we risk not allowing our students to be themselves through a slow-burn process. AI would boost qualification in language assessment because it is excellent at measuring quantifiable skills like grammar and vocabulary with efficiency; AI would also be effective for socialization, as it can measure whether students' language learning conforms to a specific linguistic norm. However, AI may cause potential loss of individual voice, creativity, and personal expression, the very essence of subjectification, and the ultimate learning goals of language learning.

To me, the unchanging Dao of LAL is not about mastering the latest tool but about leveraging my strengths as a human teacher. It is the wisdom to use any tool, be it chalk, a pen, a podcast, or an AI, to promote student language learning. It is the ability to understand theories of language learning and assessment, to think critically, and to appreciate cultural differences. Socio-emotional management plays a crucial role in fostering teacher-student relationships and strengthening human connection, particularly when providing spontaneous feedback to students. In my PhD thesis, I emphasized the socio-emotional dimension of the L2TBAL conceptual framework by utilizing Chan and Luk's (2021) TAL framework; additionally, the only qualitative

study included in my thesis focused on L2 teachers' intrapersonal and interpersonal emotion regulation in blended assessment (Su & Lee, 2024b). I was grateful for the three-year PhD study, where I could slow down, devour excellent books and seminal papers, and talk with my supervisors and peers, which helped me form a vision for my future LAL development: respond to digital change with the unchanging Dao through human-AI collaboration in language assessment.

My Vision: Respond to Change with the Unchanging through Human-AI Collaboration

I have recalled and reflected on the changing Shu by riding the digital waves and the unchangeability of Dao by thinking deeply about the core of teachers' LAL. My future vision is also guided by wisdom from Chinese philosophy: "commanding the Technique with the Dao" (以道御术, yi dao yu shu), meaning one can navigate a changing world by grasping unchanging principles (Hu & Wang, 2020).

There is no denying that AI tools have become super language assessors that can mark assignments tirelessly, diagnose weaknesses with unprecedented efficiency, and maintain objectivity and fairness and even achieve personalized learning (Baskara et al., 2024; Biju et al., 2024; Seo, 2024). Yet, if we only have "the Technique without the Dao" (有术无道, you shu wu dao), students may become quantified data archives, their progress tracked, ranked, and predicted precisely and accurately by AI. The goal of assessment shifts to finding students' imperfections and personalizing guidance to mold them into a standardized ideal language learner profile. As Lee (2023) nicely put it, "while AI teaching machines may have an abundance of information about each student's traits and about effective teaching methods so that they can provide customized learning, the teaching is oriented towards collective, average educational goals because the basis of their knowledge is dependent on the algorithmic analysis of big data" (p.559). This algorithmic analysis of big data constantly averages any differences and diversity and thus may eliminate particularity (Lee, 2023).

This rosy picture of absolute objective assessment, painted by some educational technologists, fails to account for one crucial fact: teachers and students are real, living, unpredictable human beings, not just numbers in an algorithm. If we embrace an AI-only assessment model, we risk objectifying students without the "acknowledgement of each student's singularity of experience" (Lee, 2023, p.562), teaching them to game algorithms rather than engage in deep, independent thought and meaningful encounter with their teachers and peers. We would lose our connection to the Dao of learning itself and, as Max Weber presciently warned by quoting Goethe, risk becoming "specialists without spirit, sensualists without heart" (Weber, 2005, p. 124).

Thus, to add "spirit" and "heart" to LAL, my vision of the development of LAL is human-AI collaboration. I prioritize humans over AI because humans possess subjectification and must use the Dao to direct the technique, rather than the other way around. As Biesta (2020) argues, subjectification is about enabling a student to

exist as the subject of their life, not as the object of educational interventions. Future language assessment, I realized, lies not in competing with AI on its terms, but in focusing on the uniquely human dimensions of our work. L2 teachers can become essentialists in future classroom assessment (McKeown, 2014) by delegating the mechanical and time-consuming aspects of language assessment to AI, allowing them to focus on what truly matters—areas where AI cannot contribute, such as providing social and emotional support or guiding students in contemplating cultural and ethical issues. I wish L2 teachers could be better at identifying the fear and anxiety L2 learners always experience when giving presentations in their second language with verbal and non-verbal cues (Gregersen, 2007). They can also learn how to use humor strategies to better resolve the embarrassing moments in classroom assessment (Neff & Dewaele, 2022). Although AI is said to mitigate human biased judgments in language assessment, particularly regarding L2-accented speech, it may still perpetuate racial or accent biases and stereotypes because standard language ideology is reinforced in the output of large language models, which often position Standard American English as the linguistic default (Curran et al., 2025; Kang & Hirschi, 2025; Smith et al., 2024). In classroom assessment, I hope L2 teachers can spend more time negotiating with their students in collaboration with AI to address the potential cultural and linguistic biases, like native speakerism and racialized language ideologies, owned by both humans and AI.

Let's imagine an L2 speaking task about a complex ethical dilemma such as gene editing a potentially defective baby. The L2 teacher will use AI-human collaborative assessment. AI can play the roles of various stakeholder like two Chinese parents who have been desperately trying to have healthy babies for ten years, lawyers, doctors, and biologists with different cultural backgrounds. The students are required to have a conversation with these stakeholders played by AI agents in L2, and AI can provide instant feedback on linguistic accuracy and fluency based on their conversations. The L2 teacher, now emancipated from being a mere scorer, can observe how students struggle, choose, and display courage or fear when faced with this ethical dilemma. The teacher can then initiate a deep dialogue, exploring the values and emotions behind the interaction, guiding the student toward maturity, self-awareness, and cross-cultural understanding. Meanwhile, the teacher can reveal vulnerabilities and uncertainties during the interaction by using reflection-in-action (Schön, 2016), which generates spontaneous, intuitive, and improvisational responses while fostering genuine connections with students and promoting self-improvement. The purpose of doing assessment here is no longer simply to answer the question, "*What can you do in this language?*" but rather, "*What kind of person do you want to become through learning this language?*"

The ability to collaborate with AI should therefore be a dimension of L2 teachers' LAL, as classroom assessment is by nature an emotional, social, and cultural practice (Brown et al., 2018; Liddicoat & Scarino, 2020). This combination highlights the unique strengths of both human teachers and AI, with human teachers positioned at

the center. As Sadler (1989) noted, expert human judgment excels at appraising “fuzzy criteria”, namely, abstract qualities like originality or courage, which exist on a continuum rather than a simple correct/incorrect binary. While AI excels at objective criteria, human teachers can engage in the qualitative judgment needed to guide a student’s development as a whole person.

Looking ahead, I expect future research to examine teachers’ professional judgment in human–AI collaboration. Beyond documenting what tools language teachers use and how they use them, researchers can explore how teachers justify decisions about when *to* use AI and when *not to* in support of learners’ language development. In addition, longitudinal research that traces teachers’ emotional struggles and processes of reconciliation with human–AI collaborative assessment is especially meaningful. As teachers are often the vulnerable group when institutions or national policy initiatives drive top-down adoption of AI, it is crucial to understand how they negotiate concerns such as agency loss, ethical responsibility, and professional identity. Giving teachers’ voices greater visibility through self-study research such as auto/duoethnography can reveal how they strive to balance technology use with preserving their Dao-like principles of holistic learner development—an aim that, in turn, can inform both policy and professional learning design.

The essay holds practical implications for teacher development. First, teacher trainers could create a dialogic space where ongoing LAL development is achieved through structured reflection. In this space, teachers are encouraged to articulate their “Dao” in classroom assessment and to share with peers their reactions to AI-enhanced assessment, including moments of amazement as well as disappointment or concern. Through peer exchange, teachers can make their assessment reasoning more visible, strengthen professional judgment, and build collective confidence in human–AI collaboration. Second, institutions and policymakers can provide opportunities to include frontline teachers in policymaking, rather than simply “telling” teachers what they should do with new AI technology. When AI adoption is promoted in a top-down manner, teachers may feel pressured to comply without room to reconcile tools with their teaching philosophies. Supportive dialogue can also help avoid blame narratives that label teachers as technology luddites (Logan et al., 2025), and instead treat teachers’ lived experiences as essential evidence for designing feasible, ethical, and student-centered AI-assisted assessment practices.

Conclusion

In this reflective essay, informed by the dialectical relationship between Shu (the technique) and Dao (the way) from Chinese philosophy, I chronicle my journey with digital tools in language assessment. I frame my passionate adventures with these tools as the changing Shu, while my reflections during my PhD studies represent a period of contemplating the unchanging Dao.

This distinction is made for illustrative purposes. I by no means wish to imply that L2 teachers cannot contemplate the Dao in their daily practice through reflection-in-action (Schön, 2016). Rather, I had already formed a core conviction that the true purpose of classroom assessment is, and always has been, the promotion of student language learning in its richest sense. This essay details my professional development as an individual; my Shu and Dao will naturally be different from those of another L2 teacher.

This is how I envision L2 teachers can respond to change in the digital world: by grounding themselves in the unchanging principles of language assessment. As AI becomes ever more powerful, our LAL must focus more on helping students become individuals who can respond to the changing world in a better, more human way. I hope my students experience intellectual and emotional growth through learning another language, just as I have over the past thirty years as a nonstop, lifelong language learner.

ORCID

 <https://orcid.org/0000-0003-3767-612X>

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 author would like to express her sincere appreciation to the Editor, who was willing to accept a personal reflection on her academic career.

Funding

The paper was funded by Sichuan International Studies University.

CRedit Authorship Contribution Statement

Xiaoli Su: Conceptualization, Writing—Original Draft, Writing—Review & Editing

Generative AI Use Disclosure Statement

During the preparation of this manuscript, Gemini 3.0 is used solely for improving the language and readability of the text. The author reviewed and edited the content as needed and take full responsibility for the accuracy, integrity, and originality of the work.

Ethics Declarations

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

Not applicable.

Competing Interests

None.

Data Availability

This paper does not involve new data collection or analysis.

References

- Amsterdam, A., & Bruner, J. (2000). *Minding the law*. Harvard University Press. <https://doi.org/10.4159/9780674020207>
- Barnes, M., & Tour, E. (2025). Teachers' use of generative AI: A 'dirty little secret'? *Language and Education*, 40(1), 24–39. <https://doi.org/10.1080/09500782.2025.2485935>
- Barnes, N., Fives, H., & Dacey, C. M. (2017). US teachers' conceptions of the purposes of assessment. *Teaching and Teacher Education*, 65, 107–116. <http://dx.doi.org/10.1016/j.tate.2017.02.017>
- Baskara, F. R., Puri, A. D., & Mbato, C. L. (2024). Exploring the use of generative AI in student-produced EFL podcasts: A qualitative study. *Language Teaching Research Quarterly*, 43, 81–101. <https://doi.org/10.32038/ltrq.2024.43.05>
- Berry, V., Sheehan, S., & Munro, S. (2019). What does language assessment literacy mean to teachers? *ELT Journal*, 73(2), 113–123. <https://doi.org/10.1093/elt/ccy055>
- Biesta, G. (2020). Risking ourselves in education: Qualification, socialization, and subjectification revisited. *Educational Theory*, 70(1), 89–104. <https://doi.org/10.1111/edth.12411>
- Biju, N., Abdelrasheed, N. S., Bakiyeva, K., Prasad, K. D., & Jember, B. (2024). Which one? AI-assisted language assessment or paper format: An exploration of the impacts on foreign language anxiety, learning attitudes, motivation, and writing performance. *Language Testing in Asia*, 14(1), 1–24. <https://doi.org/10.1186/s40468-024-00322-z>
- Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy & Practice*, 5(1), 7–74. <https://doi.org/10.1080/0969595980050102>
- Black, P., & Wiliam, D. (2009). Developing the theory of formative assessment. *Educational Assessment Evaluation and Accountability*, 21(1), 5–31. <https://doi.org/10.1007/s11092-008-9068-5>
- Black, P., & Wiliam, D. (2018). Classroom assessment and pedagogy. *Assessment in Education: Principles, Policy & Practice*, 25(6), 551–575. <https://doi.org/10.1080/0969594X.2018.1441807>
- Brown, G. T., Gebriel, A., Michaelides, M. P., & Remesal, A. (2018). Assessment as an emotional practice: Emotional challenges faced by L2 teachers within assessment. In *Emotions in second language teaching: Theory, research, and teacher education* (pp. 205–222). Springer International Publishing. https://doi.org/10.1007/978-3-319-75438-3_12
- Chan, C. K. Y., & Luk, L. Y. Y. (2021). A four-dimensional framework for teacher assessment literacy in holistic competencies. *Assessment & Evaluation in Higher Education*, 47(5), 755–769. <https://doi.org/10.1080/02602938.2021.1962806>
- Chen, A., Zhang, Y., Jia, J., Liang, M., Cha, Y., & Lim, C. P. (2025). A systematic review and meta-analysis of AI-enabled assessment in language learning: Design, implementation, and effectiveness. *Journal of Computer Assisted Learning*, 41(1), Article e13064. <https://doi.org/10.1111/jcal.13064>
- Chin, T., Rowley, C., Redding, G., & Wang, S. (2018). Chinese strategic thinking on competitive conflict: Insights from Yin-Yang harmony cognition. *International Journal of Conflict Management*, 29(5), 683–704. <https://doi.org/10.1108/IJCM-09-2017-0101>
- 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 6. <https://doi.org/10.1186/s40468-020-00101-6>
- Cui, Y., Liu, Y., & Yu, H. (2025). Enhancing English teachers' language assessment literacy through a professional learning community: A collaborative autoethnography study. *System*, 134, Article 103799. <https://doi.org/10.1016/j.system.2025.103799>

- Curran, N. M., Gu, B., Zhen, L., & Jenks, C. (2025). AI and native speakerism: The intersections of technology, language assessment, and linguistic objectivity. *RELC Journal*. Advance online publication. <https://doi.org/10.1177/00336882251367459>
- Davies, A. (2008). Textbook trends in teaching language testing. *Language Testing*, 25(3), 327–347. <https://doi.org/10.1177/0265532208090156>
- DeLuca, C., LaPointe-McEwan, D., & Luhanga, U. (2015). Teacher assessment literacy: A review of international standards and measures. *Educational Assessment Evaluation and Accountability*, 28(3), 251–272. <https://doi.org/10.1007/s11092-015-9233-6>
- Earl, L. M. (2012). *Assessment as learning : Using classroom assessment to maximize student learning*. Corwin Press.
- Eedelouei, S. (2026). Factors influencing language teachers' judgements and decision-making about the use of generative AI tools: A systematic review. *Teaching and Teacher Education*, 171, Article 105333. <https://doi.org/10.1016/j.tate.2025.105333>
- Eyal, L. (2012). Digital assessment literacy—The core role of the teacher in a digital environment. *Journal of Educational Technology & Society*, 15(2), 37–49.
- Fulcher, G. (2012). Assessment literacy for the language classroom. *Language Assessment Quarterly*, 9(2), 113–132. <https://doi.org/10.1080/15434303.2011.642041>
- Gan, L., & Lam, R. (2023). A tale of two teachers: Construction of novice university English teachers' assessment identity. *Chinese Journal of Applied Linguistics*, 46(2), 219–236. <https://doi.org/10.1515/cjal-2023-0205>
- Garrett, N. (1991). Technology in the service of language learning: Trends and issues. *The Modern Language Journal*, 75(1), 74–101. <https://doi.org/10.1111/j.1540-4781.1991.tb01085.x>
- Goh, C. C. M., & Aryadoust, V. (2025). Developing and assessing second language listening and speaking: Does AI make it better? *Annual Review of Applied Linguistics*, 45, 179–199. <https://doi.org/10.1017/S0267190525100111>
- Goulart, L., Matte, M. L., Mendoza, A., Alvarado, L., & Veloso, I. (2024). AI or student writing? Analyzing the situational and linguistic characteristics of undergraduate student writing and AI-generated assignments. *Journal of Second Language Writing*, 66, Article 101160. <https://doi.org/10.1016/j.jslw.2024.101160>
- Gregersen, T. (2007). Breaking the code of silence: A study of teachers' nonverbal decoding accuracy of foreign language anxiety. *Language Teaching Research*, 11(2), 209–221. <https://doi.org/10.1177/1362168807074607>
- Hew, K. F., & Cheung, W. S. (2014). Students' and instructors' use of massive open online courses (MOOCs): Motivations and challenges. *Educational Research Review*, 12, 45–58. <https://doi.org/10.1016/j.edurev.2014.05.001>
- Hu, W. (2014). Dao yu shu: Guoji zhengzhi chuanbo de zhanlue siwei. [Dao and Shu: Strategic thinking in international political communication]. *Shehui Kexue*, 12, 3–10.
- Hu, Y., & Wang, B. (2020). Yijiao de sanchong xiangdu: Jianyi, Buyi, Bianyi [The three dimensions of Yi doctrine: Simplicity, unchangeability, and changeability]. *Zhouyi Yanjiu*, 4, 35–41.
- Hyland, K. (2003). Genre-based pedagogies: A social response to process. *Journal of Second Language Writing*, 12(1), 17–29. [https://doi.org/10.1016/s1060-3743\(02\)00124-8](https://doi.org/10.1016/s1060-3743(02)00124-8)
- Jamieson, J. (2005). Trends in computer-based second language assessment. *Annual Review of Applied Linguistics*, 25, 228–242. <https://doi.org/10.1017/S0267190505000127>
- Jiang, L., Lee, I., & Yu, S. (2024). Conceptualizing multimodal feedback literacy for L2 writing teachers in the digital age. *International Journal of Applied Linguistics*, 34(4), 1479–1496. <https://doi.org/10.1111/ijal.12578>
- Kang, O., & Hirschi, K. (2025). Bias and stereotyping: Human and artificial intelligence (AI). *Annual Review of Applied Linguistics*, 45, 69–84. <https://doi.org/10.1017/s026719052500008x>
- Kremmel, B., & Harding, L. (2019). 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>
- Kurt, A. (2023). *Perspectives on digital burnout in second language acquisition*. IGI Global. <https://doi.org/10.4018/978-1-6684-9246-8>
- Lee, S. (2023). Otherwise than teaching by artificial intelligence. *Journal of Philosophy of Education*, 57(2), 553–570. <https://doi.org/10.1093/jopedu/qhad019>
- Lee, S., & Moore, R. (2024). Harnessing generative AI (GenAI) for automated feedback in higher education: A systematic review. *Online Learning*, 28(3), 82–104. <https://doi.org/10.24059/olj.v28i3.4593>

- Levi, T., & Inbar-Lourie, O. (2019). Assessment literacy or language assessment literacy: Learning from the teachers. *Language Assessment Quarterly*, 17(2), 168–182. <https://doi.org/10.1080/15434303.2019.1692347>
- Liddicoat, A., & Scarino, A. (2020). Assessing intercultural language learning. In J. Jackson (Ed.), *The Routledge handbook of language and intercultural communication* (2nd ed., pp. 395–409). Routledge. <https://doi.org/10.4324/9781003036210-30>
- Liu, H., Li, H., & Fang, F. (2023). “Behind the screen, I still care about my students!”: Exploring the emotional labour of English language teachers in online teaching during the COVID-19 pandemic. *International Journal of Applied Linguistics*, 34(2), 450–465. <https://doi.org/10.1111/ijal.12501>
- Liu, X., & Liu, Y. (2025). Developing and validating a scale of artificial intelligence anxiety among Chinese EFL teachers. *European Journal of Education*, 60(1), Article e12902. <https://doi.org/10.1111/ejed.12902>
- Lloyd, G. (1996). *Finite and infinite in Greece and China*. *Chinese Science*, 13, 11–34. <http://www.jstor.org/stable/43290378>
- Logan, C., Nichols, T. P., & Garcia, A. (2025, December 9). *Teach like a Luddite*. *Phi Delta Kappan*. <https://kappanonline.org/teach-like-a-luddite/>
- Lortie, D. C. (1975). *Schoolteacher: A sociological study*. University of Chicago Press.
- Lütge, C., Merse, T., & Su, X. (2019). Becoming a digitally competent teacher: The European framework for the digital competence of educators. *Praxis Fremdsprachenunterricht Englisch*, 16(4), 4–7.
- Lütge, C., Merse, T., & Su, X. (2021). The digital competence of English language educators: Exploring the DigCompEdu framework with an empirical case study. In C. Lütge & T. Merse (Eds.), *Digital teaching and learning: Perspectives for English language education* (pp. 31–60). Narr Francke Attempto.
- Ma, X., Zhang, C., & Lee, I. (2026). L2 learners' engagement with AI-generated feedback on writing. *Assessing Writing*, 68, Article 101020. <https://doi.org/10.1016/j.asw.2026.101020>
- McKeown, G. (2014). *Essentialism: The disciplined pursuit of less*. Crown.
- Medgyes, P. (1999). *The non-native teacher* (2nd ed.). Macmillan.
- Nayernia, A., & Mohebbi, H. (2023). Teachers' online language assessment literacy. In Z. Tajeddin & C. Griffiths (Eds.), *Language education programs: Perspectives on policies and practices* (Volume 34) (pp. 121–140). Springer. https://doi.org/10.1007/978-3-031-38754-8_8
- Neff, P., & Dewaele, J. M. (2022). Humor strategies in the foreign language class. *Innovation in Language Learning and Teaching*, 17(3), 567–579. <https://doi.org/10.1080/17501229.2022.2088761>
- Neisser, U. (1994). Self-narratives: True and false. In U. Neisser & R. Fivush (Eds.), *The remembering self: Construction and accuracy in the self-narrative* (pp. 1–18). Cambridge University Press.
- Nyanjom, J., & Naylor, D. (2020). Performing emotional labour while teaching online. *Educational Research*, 63(2), 147–163. <https://doi.org/10.1080/00131881.2020.1836989>
- Ockey, G. J. (2009). Developments and challenges in the use of computer-based testing for assessing second language ability. *Modern Language Journal*, 93(s1), 836–847. <https://doi.org/10.1111/j.1540-4781.2009.00976.x>
- O'Reilly, T. (2007). What is Web 2.0: Design patterns and business models for the next generation of software. *Communications & Strategies*, 65, 17–37.
- Pastore, S., & Andrade, H. L. (2019). Teacher assessment literacy: A three-dimensional model. *Teaching and Teacher Education*, 84, 128–138. <https://doi.org/10.1016/j.tate.2019.05.003>
- Quadir, B., Yang, J. C., & Wang, W. (2022). Factors influencing the acquisition of English skills in an English learning environment using Rain Classroom. *Interactive Learning Environments*, 32(1), 1–19. <https://doi.org/10.1080/10494820.2022.2075015>
- Redecker, C. (2017). *European framework for the digital competence of educators: DigCompEdu (JRC107466)*. Joint Research Centre. <http://publications.jrc.ec.europa.eu/repository/handle/JRC107466>
- Sadler, D. R. (1989). Formative assessment and the design of instructional systems. *Instructional Science*, 18(2), 119–144. <https://doi.org/10.1007/bf00117714>
- 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>
- Schön, D. A. (2016). *The reflective practitioner: How professionals think in action*. Routledge.
- Seethamraju, R. (2014). Effectiveness of using online discussion forum for case study analysis. *Education Research International*, 2014, 1–10. <https://doi.org/10.1155/2014/589860>

- Seo, J. (2024). Exploring the educational potential of ChatGPT: AI-assisted narrative writing for EFL college students. *Language Teaching Research Quarterly*, 43, 1–21. <https://doi.org/10.32038/ltrq.2024.43.01>
- Smith, G., Fleisig, E., Bossi, M., Rustagi, I., & Yin, X. (2024). Standard language ideology in AI-generated language. ArXiv. <https://arxiv.org/abs/2406.08726>
- Stiggins, R. J. (1991). Assessment literacy. *Phi Delta Kappan*, 72(7), 534–539.
- Su, X., & Lee, I. (2024a). Is the magic in the mix? The development and validation of the L2 teachers' blended assessment literacy scale. *System*, 123, Article 103354. <https://doi.org/10.1016/j.system.2024.103354>
- Su, X., & Lee, I. (2024b). Emotion regulation of EFL teachers in blended classroom assessment. *The Asia-Pacific Education Researcher*, 33(3), 649–658. <https://doi.org/10.1007/s40299-023-00761-x>
- Su, X., Ma, W., Lütge, C., & Merse, T. (2021). “Ouzhou jiaoyu gongzuzhe de shuzi shengrenli kuangjia” jiedu he qishi. [The introduction and implications of DigCompEdu]. *Kaifang Xuexi Yanjiu*, 26(3), 47–54. <https://doi.org/10.19605/j.cnki.kfxyj.2021.03.006>
- Su, X., Lee, I., & Wang, M. D. (2025). Conceptualizing L2 teachers' blended assessment literacy (L2TBAL): A scoping review. *International Journal of Applied Linguistics*. <https://doi.org/10.1111/ijal.70002>
- Su, X., Yin, H., & Lee, I. (2024). Let a hundred flowers bloom: Towards a coexistence of paradigms in language assessment literacy. *Linguistics and Education*, 84, 101346. <https://doi.org/10.1016/j.linged.2024.101346>
- Taylor, L. (2009). Developing assessment literacy. *Annual Review of Applied Linguistics*, 29, 21–36. <https://doi.org/10.1017/S0267190509090035>
- Wang, Y., Wu, J., Chen, F., Wang, Z., Li, J., & Wang, L. (2024). Empirical assessment of AI-powered tools for vocabulary acquisition in EFL instruction. *IEEE Access*, 12, 131892–131905. <https://doi.org/10.1109/ACCESS.2024.3446657>
- Weber, M. (2005). *The Protestant ethic and the spirit of capitalism*. Taylor & Francis.
- William, D. (2011). What is assessment for learning? *Studies in Educational Evaluation*, 37(1), 3–14. <https://doi.org/10.1016/j.stueduc.2011.03.001>
- Xu, Y., & Brown, G. T. (2016). Teacher assessment literacy in practice: A reconceptualization. *Teaching and Teacher Education*, 58, 149–162. <https://doi.org/10.1016/j.tate.2016.05.010>
- Yao, D. (2021). Automated writing evaluation for ESL learners: A case study of Pigai system. *The Journal of Asia TEFL*, 18(3), 949–958. <https://doi.org/10.18823/asiatefl.2021.18.3.14.949>
- Zu, L. (2022). Five principles of Taoism for purpose-driven leadership. In L. Zu (Ed.), *Responsible and sustainable business* (pp. 197–225). Springer. https://doi.org/10.1007/978-3-030-91117-1_14