A Linguistic Perspective on LMS-Mediated Language Testing in ELT Programs

Syamsiarna Nappu – Maemuna Muhayyang – Ratna Dewi – Nurdevi Bte Abduh – Hasnawati

DOI: 10.18355/XL.2026.19.02.01

Abstract
The use of Learning Management Systems (LMSs) in English Language Teaching (ELT), particularly in tertiary instruction, has reconfigured assessment practices, including graduate language testing. This study examines LMS-mediated language testing through a linguistic perspective by exploring how SPADA, an institutional LMS, shapes assessment practices in a graduate Language Testing course at Universitas Muhammadiyah Makassar. Grounded in language testing theory, the study investigates how LMS mediation facilitates or constrains four core assessment principles, validity, reliability, fairness, and washback, and how LMS-based tasks contribute to the development of assessment literacy among ELT graduate students. A qualitative case study design was adopted, drawing on LMS artifacts, students' reflective journals, and instructor feedback. The findings indicate that LMS-mediated assessment supports principled practice through rubric-guided task sequencing, asynchronous peer feedback, and reflective interaction, although these opportunities are conditioned by technical limitations and uneven digital competencies. The study further reframes LMS-based assessment as a language-mediated and discourse-oriented practice in which assessment meanings are co-constructed through interaction, evaluative language, and genre-specific expectations within the platform. It concludes that strengthening the linguistic design of LMS-mediated testing, by attending to discourse features, communicative purposes, and sociocultural context, can more effectively cultivate assessment literacy and meaningful learning in online ELT programs.
Key words: linguistic perspective, LMS-mediated language testing, ELT, SPADA, assessment

Introduction

The incorporation of technology into language learning has led to radical changes in teaching, learning, and assessment, with English Language Teaching (ELT) no exception. The development of Learning Management Systems (LMS) as digital infrastructures has redesigned pedagogical models and established these tools as the main instruments for providing instruction, sharing content, and supervising assessment in blended and fully online learning settings (Rapanta et al., 2021; DeLuca et al., 2022). Of these changes, assessment has become a key feature that needs to be reconceptualized and investigated. The origins of most large-scale, standardized language tests lie in language testing theory grounded in psychometric processes concerned with validity, reliability, washback, authenticity, and fairness (Fulcher, 2020; O'Sullivan, 2021). However, embedded assessment via LMS poses an additional challenge, as the principles need to be translated into asynchronous, multimodal spaces that are often less regulated than face-to-face classrooms (Knoch & Chapelle, 2021; Harding & Sadeghi, 2022).

In ELT, LMS-supported assessment has been largely investigated in relation to its affordances or users' perceptions, but with little consideration of how such

assessments may be situated within the theoretical and moral backdrops of language testing (Yan & Fan, 2021; Khalil & Elkhider, 2022). The learning potential of LMS is its capacity to foster formative, continuous, and collaborative approaches to assessment (Winstone et al., 2021). These include rubric-based tasks, peer assessment, computerized language tests, discussion boards, and reflective writing, which can be aligned with learning outcomes and language testing standards when used wisely (Ockey & Wagner, 2021). However, despite these affordances, research shows that LMS are not always used for assessment informed by language assessment theory, particularly at advanced academic levels such as graduate education (Dahlstrom et al., 2020; DeLuca et al., 2022).

In graduate ELT programs, courses in language testing play a significant role in training new teachers, curriculum developers, and testers. These courses aim to provide students with an orientation to functional testing, interpretation, and use, grounded in theoretical perspectives critical to applied linguistics and language education (Kane, 2021). Nevertheless, in courses mediated by an LMS platform, the practice of testing principles is determined not only by digital literacy but also by the platform's affordances and the pedagogic approach (Yan et al., 2021; Vogt & Tsagari, 2020). This tension also presents some key issues to be addressed: how is it possible for assessments delivered through LMS to fulfill the principles of language testing? Is it possible for such online spaces to foster assessment literacy among graduate students? Finally, what are the problems and potential benefits of integrating LMS into the language testing syllabus?

Current knowledge on the integration of LMSs with summative assessment remains limited and is predominantly drawn from studies in undergraduate education (Rasooli et al., 2021), course-specific evaluation of skills such as writing or speaking (Khalil & Elkhider, 2022), and general perceptions of the effectiveness of online learning (Baird & Fisher, 2005). However, studies based on graduate-level Language Testing courses are still rare. Furthermore, most studies measure LMSs' usage empirically rather than basing it on a linguistic and pedagogical framework (Yan & Fan, 2021; Kocdar et al., 2018). Indeed, there is a significant void in the literature with respect to how LMS-based assessments can be leveraged not only to measure language proficiency or course content but also to interweave basic tenets of language testing and foster enhanced student assessment literacy directly into formal coursework (DeLuca et al., 2022; Xu & Brown, 2016).

Assessment literacy of students is another important, albeit relatively unexplored, dimension in the literature. It is defined as an awareness of the purposes, principles, methods, and outcomes of assessment in practice, together with the realization that this knowledge can be applied effectively (Xu & Brown, 2016; Vogt & Tsagari, 2020). Although teacher assessment literacy has garnered significant attention, especially in professional development, relatively little empirical research has examined how it is developed among graduate students preparing for ELT careers. This omission is especially significant given the increased importance of digital assessment instruments in post-pandemic educational settings (Yan et al., 2021; DeLuca et al., 2022).

In the context of higher education in Indonesia, the government has recommended that all LMS used by universities be based on a government enterprise LMS equivalent system called SPADA (Sistem Pembelajaran Daring). In the English Education Study Program at Universitas Muhammadiyah Makassar, SPADA is only a limited digital environment for conducting Language Testing courses. This is the medium where lectures, share materials, submit tasks, and receive peer feedback, all of which are imperative for assessment-rich teaching and learning. However, the extent to which SPADA-based tests reflect applied language testing and promote assessment literacy remains empirically underexplored (Winstone et al., 2021; Harding & Sadeghi, 2022).

Research by Kocdar et al. (2018) and Alruwais et al. (2018) notes that although LMSs provide functions that enhance the digital flexibility of language education, these are rarely embedded in ways that align with assessment theory. Furthermore, Rahimi et al. (2022) found that LMS-mediated assessment experiences among students are different depending on students' familiarity, digital fluency, and explanatory design. As such, the effective implementation of LMS-based assessment in language testing relies on more than technological proficiency; it requires a theoretically supported and pedagogically attuned discourse that is absent from the literature at this point.

Furthermore, the ongoing trend of educational digitization has created pressing new needs for assessment accountability designs. As noted by Bennett et al. (2020), the integration of digital assessment with learning objectivity is frequently not well aligned in higher education. For example, Broadbent and Poon (2020) argue that the proliferation of Self-Regulated Learning (SRL) and learner autonomy in online settings demands assessment designs that facilitate formative interaction. Meanwhile, Panadero et al. (2018) emphasize the role of feedback literacy in online spaces, including a virtual environment in which an LMS-supported feedback ecology is established. Research has also shown that teacher digital competence (Tondeur et al., 2021), student engagement with criteria (Carless & Chan, 2017), and accessibility (Martin et al., 2020) are critical to achieving equitable outcomes in LMS- based assessments. However, despite this realization, applying language testing principles in practice on LMS platforms remains relatively underdeveloped, with greater attention to linguistic and applied pedagogic perspectives. In this vein, the present study contributes a task-based application grounded in empirical evidence to the literature on language assessment and educational technology. It describes how LMS-based assessment within the context of a graduate Language Testing course aligns with theoretical understandings of language testing, supports student learning, and enhances assessment literacy.

In an effort to fill the gap, this study examines one case at Universitas Muhammadiyah Makassar, where SPADA is used as a medium of instruction in a Language Testing course within the Master of English Education Study Program. Thus, in this study, the authors aim to learn from authentic teaching practice and provide genuine insights into the pedagogical uses of LMS in language testing. The present study pursues three major research questions: (1) how is LMS-based assessment enacted in an online Language Testing course in a graduate ELT context, specifically within the SPADA LMS space?; (2) to what extent are there visible possibilities and constraints for the development of basic aspects of language testing principles as validity, reliability, fairness, and washback through LMS-based tasks?, and (3) how does the design and implementation of LMS-based assessment influence the development of assessment literacy in ELT graduate students with respect to their conceptual knowledge about language testing and measurement, as well as practical involvement in work-oriented theories of language assessment? The questions are intended to narrow the gap between theory and practice and to analyze LMS integration from a linguistic perspective.

Literature Review

Linguistics Perspective on LMS Integration in Language Testing Courses

The introduction of Learning Management Systems (LMS) into Language Testing classrooms has created a new window for the realization of linguistic theories in e-learning environments. From a linguistic point of view, language competence comprises four aspects: grammatical competence, sociolinguistic competence, discourse competence and strategic competence (Canale & Swain, 1980; Bachman,

1990). In fact, in connection with this, "the test of language is not a technical event concerned only to measure proficiency, but an interpretive act that is definitely grounded in both of the attitudes towards language as communication and as functional (Douglas 2000). All of these raise the issue of how to evaluate tasks in programs delivered through LMS systems such as Moodle or SPADA in ways that are sensitive to the extent to which language is used across modes, registers and functions (Chapelle, 2001).

Moreover, in multilingual settings, online resources facilitate translanguaging practices in which learners can employ their full linguistic repertoire to negotiate meaning from texts (García & Wei, 2014). Middle translanguaging can emerge in LMS-facilitated spaces through features such as multilingual discussion boards and scaffolded writing tasks. In a recent study by Nappu et al. (2025), peer feedback tasks in a digital classroom led learners to critically reflect on language forms, cross-language awareness, and genre-driven conventions of writing. However, to design an LMS assessment with respect for linguistic perspectives is not without its difficulties. The majority of LMSs lack emphasis on linguistic pedagogy, leading to practices that are not embedded and to almost entirely automated exercises that fail to assess interactional competence or discourse fluency. In addressing this, Bachman and Palmer (2010) offer a model of assessment design based on communicative language ability which focuses on authenticity, cognitive engagement and contextual relevance.

Therefore, according to the interpretation of language in LMSs' perspective in Language Testing programs, there should be a constructed convergence among technical artifacts and theories of language learning. It also asks teachers to be assessment-literate and to have second-language pedagogical knowledge in creating authentic, valid, and context-sensitive tasks on which not only a surface level can be evaluated.

Language Assessment on LMS Integration

Language assessment in digital environments requires a principled framework that remains faithful to the foundations of language testing while accounting for the linguistic realities of online interaction. As Widdowson in Currie (2004) argues, language testing has traditionally been dominated by psychometric traditions that emphasize measurement, standardization and what scores mean. Meanwhile, modern applied linguistics has stressed that language performance is context-dependent, discourse-driven, and socially situated. This repositioning suggests that core testing notions: validity, reliability, fairness, and washback have to be reframed in the context of LMS-mediated assessment, where interactional requirements, textual genres and digital participation norms influence how language is generated and assessed. Therefore, the fundamental principles of assessment are essential to master. Validity, defined as the degree to which an assessment measures the construct it claims to measure (Messick, 1989), becomes especially complex in LMS settings because the construct is often expressed through digitally mediated discourse. From a linguistic perspective, a construct representation needs to describe not only linguistic knowledge but also discourse competence, pragmatic fitness and genre competence. In online testing, construct validity might be at risk when non-verbal factors (e.g., typing proficiency, navigation skills or an unreliable internet connection) are confounded with linguistic performance (Kane, 2021). On this side of such practical constraints, however, LMS-based tasks are likely to invite careful consideration of the discursive conditions of assessment: who addresses whom, where, by what genre conventions, in

pursuit of whose interactional goals. This is the crucial point, as linguistic performance will inevitably be influenced by discourse context and audience orientation, especially in asynchronous writing and peer-response situations. The other is reliability, which represents the degree of stability and consistency of ratings across raters and contexts. LMS platforms can support reliability through standardized rubrics, archived submissions, and structured scoring interfaces, which allow assessors to apply consistent criteria over time (Carless & Chan, 2017). However, technical interference and varying access conditions may also contribute to measurement instability (Harding & Sadeghi, 2022). Hence, reliability in digital assessment should not be ensured solely by rubrics, but also through rater training and discourse-aware calibration tasks. Third, there is fairness, which, in language testing, is essentially connected to equality of opportunity and, in digital spaces , overlaps with sociolinguistic variation and the digital divide. Linguistically, fairness also involves avoiding bias toward particular registers, discourse styles, or cultural norms of interaction that may privilege certain learners. In LMS-based testing, fairness involves incorporating inclusive technological design and sociolinguistic awareness of diverse communicative norms. Finally, washback is the effect of testing on teaching and learning. In LMS-supported context, washback is evident in the patterns of revising, peer dialogue and reflective engagement. Positive washback is created when the feedback learners receive is timely, actionable and linguistically relevant; this plays a role in learners' development not only in grammatical correctness but also in clear content organization and pragmatic efficiency. LMS functions, such as peer feedback, teacher annotation and resubmission, may foster iteration in learning, where assessment itself becomes a learning process that strengthens subsequent cycles of learning (Winstone & Carless, 2019).

LMS-Based Assessment in ELT

In ELT higher education, particularly following the worldwide transition to online learning due to the COVID-19 pandemic, LMS-based evaluation has become the new standard (Rapanta et al., 2021). Although LMSs facilitate administrative convenience, their role in pedagogy is leveraged when they are used for formative, interactive assessment that aligns with language-learning objectives. From a linguistics perspective, it should be recognized that LMS-based assessment is a discourse ecology in which language proficiency is enacted through genres, stance-taking, interactional negotiation, and identity performance, rather than a neutral space for the delivery of tests. This aligns with the stance in applied linguistics that language use develops through social interaction and situated communication, such as peer feedback and that items used for measurement should be able to elicit language in use rather than crystallize isolated language features.

Moreover, in interactional linguistics terms, peer feedback is a communicative event involving participants in the negotiation and management of meaning, politeness and face issues and evaluative language or expressions, all of which are closely related to pragmatic competence and academic discourse norms. Thus, when appropriately organized, peer response activities may foster both assessment literacy and discourse competency through dialogic encounters.

Research indicates that graduate students are more actively involved in LMS assessment when the constructs are collaborative, reflective and feedback-centered (Broadbent & Poon, 2020). Structured peer review can further facilitate the development of understanding of validity and reliability by placing weight on learners'

use of scoring criteria and shared rubrics to interpret language performances (Carless et al., 2021). Yet, there are obstacles: LMS use can also be confined to transmitting content and receiving assignments, thereby limiting its ability to support discourse-rich assessment practices. Rahimi et al. (2022) showed that teachers frequently avail themselves only sparingly of the full range and potential of LMS tools for interactive assessment and feedback, thus allowing engagement to remain superficial and limiting the development of learners' analytical and communicative abilities.

Assessment Literacy in ELT Graduate Settings

Assessment Literacy, as the knowledge, skills, and principles that are essential for understanding assessment results (Stiggins, 1995; Xu & Brown, 2016), is an integral part of teacher preparation programs. In graduate ELT settings, however, assessment literacy is not only technical competence in designing and scoring tests; it is also a linguistically mediated professional practice. From a linguistics perspective, assessment literacy includes the ability to interpret learner performance as discourse, recognize how meaning is shaped by grammar, pragmatics, register, and genre, and communicate evaluation in ways that support language development. In other words, it involves competence in evaluative discourse: how assessors use language to judge quality, justify decisions, and negotiate standards in academic settings.

This view is consistent with applied linguistics perspectives that frame professional learning as discourse socialization, where novice practitioners acquire discipline-specific ways of reasoning, writing, and evaluating. Within graduate Language Testing courses, students are socialized into assessment communities by learning to produce rubrics, articulate criteria, write feedback, and justify judgments using specialist concepts such as validity, reliability, fairness, and washback. These practices reflect broader academic literacies traditions in which competence involves participation in disciplinary genres and evaluative norms, not merely the acquisition of content knowledge.

Modern frameworks, such as the Assessment Literacy Framework, define assessment literacy as the integration of pedagogical, sociocultural, and ethical awareness (Brookhart, 2011). In LMS contexts, this expands into digital assessment literacy: the ability to use technology tools responsibly and pedagogically to design tasks, apply rubrics, analyze LMS evidence, and provide feedback (Tondeur et al., 2021). Linguistically, digital assessment literacy also requires awareness that online feedback is often asynchronous and text-based, demanding clarity, explicitness, and pragmatic appropriateness because immediate conversational repair is limited. Therefore, feedback quality depends not only on the tool but also on how evaluative language is phrased, structured, and aligned with genre expectations.

ELT graduate students may have gaps in their assessment literacy because they are rarely exposed to practical procedures for test design. DeLuca et al. (2022) demonstrated that engagement with systematic application tools, particularly in formative and reflective tasks, improved students' ability to translate theoretical concepts into practice. Activities such as test design, peer review, and reflective writing nurture an in-depth understanding of language testing theory (Panadero et al., 2018). Yet, assessment literacy needs to be developed and modeled by teachers, not least in online courses.

Method

This research is a qualitative case study within an interpretive research paradigm. The adoption of a qualitative approach stems from the aim of exploring in greater detail how LMS-based assessment practices are integrated into a graduate Language Testing

course and how these practices reflect or embody central principles of language testing and test and assessment literacy. Through the case study approach, it is possible to analyze real-life experiences in context, providing descriptions and interpretations that align with participants' lived realities (Merriam & Tisdell, 2016; Yin, 2018). Because this study is not trying to generalize but rather to make sense of nuanced teaching practices within an institutional context, the interpretive paradigm is grounded in creating meaning through participants' perspectives and researchers' reflexivity (Creswell & Poth, 2018).

Research Context and Participants

The research was conducted in the Graduate Program of English Language Education at Universitas Muhammadiyah Makassar, Indonesia. The selected site was a Language Testing course conducted online through the university's LMS, SPADA (Sistem Pembelajaran Daring). The course is a program requirement that prepares future EFL teachers by introducing them to basic principles and guidelines for language assessment design, theory, and application.

Participants included a course instructor and 24 master's students taking the Language Testing course in the 2025 academic year. All participants were skilled in digital technologies, had frequent access to SPADA, and some had prior experience with online learning modules. Participants were purposively selected based on meaningful experience with LMS-facilitated assessment tasks to provide rich, relevant, and informative data for analysis (Patton, 2015).

Data Collection

Collection from various sources enabled data triangulation and provided an in-depth overview of LMS assessment practices. These sources of data were: (1) course documents including syllabi, course materials and rubrics used to guide teaching and assessment, as well the descriptions of the SPADA tasks; (2) task-related artefacts including students submissions themselves, peer feedback, reflective journals and annotated rubrics; (3) semi-structured interviews with the teacher active participant in the course concerned and 10 students sampled throughout the research period in relation to their perceptions, construal and engagement with assessment principles; (4) end-of-cycle reflections solicited from all participants providing insights about such understandings as validity, reliability fairness washback; and (5) interaction logs kept by LMS tracking interface patterns on students using for example how engaged were they over time with making postings to peers' commentaries submitting to examine interaction patterns and task completion behavior. Data for all instruments were collected over one semester (16 weeks), encompassing all teaching and testing stages of the Language Testing course.

Data Analysis

Thematic analysis was conducted with a flexible thematic approach (Braun & Clarke, 2006), moving inductively and deductively to construct themes that emerged from existing theoretical typographies. The initial codes were driven by the five principles of language assessment: validity, reliability, fairness, authenticity, and washback; as well as dimensions of assessment literacy such as the knowledge about test design, ethical implications, and pedagogical application (Fulcher & Davidson, 2007; Xu & Brown, 2016).

The analysis was conducted according to Braun and Clarke's (2006) six-phase process: data familiarization, generating initial codes, searching for themes, reviewing

themes, defining and naming themes, and report writing. Interview and reflection data were transcribed and triangulated with document analysis and LMS logs to increase the reliability of the results. Qualitative data were analyzed and coded systematically with NVivo software.

The study complied with the ethical standards recommended by the Ethics Committee of Universitas Muhammadiyah Makassar. All participants provided informed consent and were given an explanation of the aim and the voluntary nature of the research. Trade partners were guaranteed anonymity, and pseudonyms identified all active sex trade participants in the published data to protect their identities. All data were kept in a secure location and used only for academic purposes. Participation did not affect course grading or evaluations due to research protocols.

Results

This section presents the study's findings in response to the three guiding research questions. Each subsection addresses a single research question and integrates data from LMS interaction logs, student task artifacts, interviews, and reflective journals.

LMS-Based Assessment in Graduate Language Testing Course

The initial research question asked how LMS-mediated assessment was enacted in a postgraduate Language Testing course implemented via SPADA. Comparison of teaching and learning materials, as well as LMS log data, revealed repetitions in weekly assessment tasks such as formative quizzes, peer assessment, reflective journals and group presentations. These exercises were intended to provide scaffolding so that students could construct knowledge of language testing theory and its application from the ground up.

From a linguistic perspective, the task also constituted discourse events in which students were asked to appropriate academic language practices, understand task questions, embrace disciplinary registers, construct meaning in peer interactions, and stage evaluative talk when employing rubrics. In the interview, some participants described SPADA as not just a place to input work, but an interactive environment in which assessment was based on language use, through the discourse of critique, justification and stance-taking. "*We had to be cautious with the words that we used: `You could improve this by doing X or Y ',and not say 'This is ugly and you should change it. I was taught how to feedback in a way that is polite but very clear, and to use academic language, not just 'good' or 'wrong*'." (Student 5, Interview).

These high-stakes assessments were facilitated through the SPADA platform , which contained easy-to-follow modules (readings, discussion questions, rubrics and assignment upload facilities). Tasks with rubric support were provided throughout the course. Students had to apply analytical rubrics to evaluate peers' test designs or the quality of the sample assessment (validity and reliability). There were indications that these contributions made a relatively strong contribution with purposeful responses to the criteria and theory.

Linguistically, the extensive use of rubrics meant that students had to read assessment language as genre-specific metalanguage. Through this process, students were not only practicing applying criteria, but they were also learning the language of assessment, words like construct alignment, task representativeness, and scoring consistency, which affected the way that students articulated their judgments about performance. "*In the beginning the rubric wording seemed hard to me, but then I understood that it was teaching me how to think like a tester. I had to be able to explain why it's valid, not just say that it is.*" (Student 9, Interview). This suggests that

rubric feedback helped foster assessment literacy by facilitating participants' socialization into the professional language-testing community.

Based on the interviews, students identified that they learned from the LMS design due to its accessibility, clarity and well-structured feedback instructions. Most SPADA students reported in reflective journal entries that its interface supported planning, time management and regular interaction. Logs of LMS interactions also revealed extensive peer-to-peer participation in the discussion forums, particularly during challenge weeks when students were required to work together or conduct peer reviews.

"*The rubric made me actually understand what I was aiming to accomplish. It was not so much grading, it was learning to grade properly*." (Student 2, Interview)

"*Enjoyed getting the feedback and revising my work. It got me thinking that assessment was not like it's the final thing, but more a process*." (Student 8, Reflective Journal WK5)

Crucially, students also reported that LMS peer talk influenced their awareness of the nature of discourse and pragmatics in assessment communication. "*Assessment, I realized, is also about how we talk to one another," observed one student. Too direct, my friend feels attacked and too soft, it's not helpful. So, I attempted to moderate the language."* (Student 4, Reflective Journal Week 7). Comments such as these suggest that LMS-mediated assessment prompted students to rehearse pragmatically valid evaluative discourse and to view assessment as communication embedded in interpersonal meaning-making.

However, some challenges were reported. They faced a challenge in making sense of the task, especially for participants with less experience with theoretical frameworks in language testing. These concerns were partially addressed through LC posts and asynchronous group consultation sessions. The teacher noticed that SPADA provided strong support for content delivery through some formative evaluations, but it lacked native tools for fine-grained feedback (annotation) and had to rely on external files or messaging services.

Some students also associated these difficulties with linguistic rather than merely technical aspects. For instance, one participant reported that "the information was sometimes confusing because of the testing terminology, which we were not familiar with. "*Once the lecturer explained what some of those key terms were, it was easier to follow*." (Student 11, Interview). This draws attention to LMS task accessibility because it is influenced not only by platform usability, but also by the makeup of the assessment discourse and metalanguage in which task instructions are cast.

Thus, the constructive alignment of LMS-integrated assessment in the Language Testing course was realized through phased tasks, rubric directives and interaction tools that encouraged student engagement with theoretical content and peer rater tasks. Although there were some small technical challenges, the LMS system could be a helpful resource for planning assessment, delivery and follow-up. From a linguistics perspective, the results also reveal that LMS-based assessment is realized through language-in-use: It mediates students' participation in academic assessment discourse and their expression of evaluative meaning, and facilitates the negotiation of feedback in ways that are integral to professional practice in language testing.

Applying Testing Principles to real language testing through LMS-driven assessment

The second research question concerned how LMS-based assessment helped fulfill the key language testing principles: validity, reliability, fairness, and washback. There was a rationale for aligning assessment tasks with course learning outcomes. Review of the documents revealed that the activities focused on whether students grasped particular theoretical concepts. For example, students' task was to judge real language tests based on their validity aspects (construct, content or face). Students agreed that these exercises helped them grasp the concept of validity through practical analysis (interview data). Furthermore, capstone projects in which students generated their own language assessments implemented the course-introduced validity-directed design approach. From a linguistics perspective, however, the validity category does not simply refer to a technical meaning; it appeared relevant to how well the tasks actually elicited intended discourse features in terms of coherence, genre control, pragmatic appropriateness and communicative purpose. One participant reported that considering validity forced them to think about the language demands inherent in a task: "*I started questioning if the task really makes students use the language function we want, such as arguing, explaining or negotiating meaning not just filling some blanks*." (Student 12, Interview). This suggests that students viewed validity through a discourse-and-function perspective, rather than solely through score-based arguments.

"*Having dissected a real test using the concepts of validity, it astounded me about how much thinking goes into even the simplest part of any test - probably more than I imagined."* (Student 1, Interview)

Credibility was further strengthened through the use of standardized rubrics and peer-evaluation activities. Both the instructor and learners had access to rubrics with specific descriptors and scoring criteria to evaluate their peers' work. Task artifacts showed that peer scores were well calibrated with teacher attribution, indicating inter-rater reliability. Through the rubric tool in the LMS and continuous peer evaluation, students became experienced in estimating the reliability of scores. As they wrote in their reflections: "*I've learned the subjectivity bias [in assessment and standards; by doing peer evaluations.* Crucially, students also said that reliability work relied on shared understandings of assessment language and the meanings of terms and scores in rubrics. Multiple participants in interviews stressed that consistency was achieved only after they had negotiated the linguistic meaning of several descriptors, which referred to what constituted coherent, appropriate register or clear argumentation. *"At the beginning, we interpreted 'coherence' in the rubric in divergent ways. We discussed, and decided what the discourse markers and organization should be*." (Student 6, Interview). This demonstrates that reliability in LMS peer scoring was also constructed through discursive negotiation and shared genre knowledge alongside technical standardization.

"*It was difficult at first giving scores to my friend's test design, contrasting with the rubric and other people's comments, but [later on] I got better at doing it."* (Student 6, Reflection Journal Week 9)

Fairness was facilitated by asynchronous task delivery, open chat places and anonymous peer review. The flexibility SPADA provided was popular among students who were able to work on their own time. Equitable treatment was also supplied through checklist rubrics and the opportunity to revise some assignments after feedback. Several students had concerns about unequal digital literacy, and they requested that continuous support for digital literacy be provided. From a language

perspective, fairness was also considered by students in terms of sociolinguistic variation: learners may use different registers, rhetorical styles or culturally-conditioned modes of critiquing and taking a stance. One student remarked that fairness meant accepting variation in voice style, not forcing everyone to adopt a single ideal academic voice: "*Sometimes my friends write differently from me [because of their background language style], but it might be logical. So the fair thing is not to be pushing everyone into one style.*" (Student 14, Interview). Statements like these reflect the fairness seen not only in equal opportunity but also in the representation of diverse language varieties in academic success.

"*Another big thing was time to complete work and the internet also being (unreliable) at home. It felt more equalizing to me.*" (Student 10, Interview)

Washback was described in terms of the impact of testing on students' learning behavior. The reflective journal essays also made clear that some students returned to the course/test-theory readings to improve their practice following peer/lecturer feedback. Repetitive practice and feedback support a learning-center assessment culture consistent with the repeated nature of assessments. A number of students reported feeling more confident in their ability to appraise assessments objectively and in recognizing the clinical relevance of tasks for future professional work. In terms of language, washback was manifested in students' monitoring the discourse quality not only of their own test tasks but also of the feedback language. Other participants mentioned that their pragmatic tone and clarity apparently tightened, as did their emphasis on understanding the genre when providing peer feedback and when designing test prompts. One student said: "*When I got feedback that we couldn't test on the instructions that I had, they were a bit too nebulous. I rewrote them so that the language function and audience are apparent*." (Student 9, Interview). This is a sign of washback, not just at the level of content learning but also at the levels of discourse acquaintance and communication skills enhancement.

"*The criticism was so such that I had to take a step back and say, 'Is this test the only time when I know what's best for my patients?*' she said. (Student 3, Reflective Journal Week 11).

The study , taken together, demonstrated how LMS-based testing could enable the translation of core principles of language testing into an authentic and valid language assessment setting. Task building and establishing feedback systems and rubrics were necessary to make the abstract principles concrete in the learning experiences. Linguistically, the results also demonstrate that these principles were operationalized via language-mediated processes, such as negotiation of rubric meanings, interpretation of discourse -level performance and use of evaluative language in peer feedback, thereby suggesting that principled assessment in LMS environments is not based solely on technical design but also on discourse, pragmatics and sociocultural meaning-making.

Development of Assessment Literacy Among Graduate ELT Students

The third research question was to determine the role of LMS-based assessment in fostering assessment literacy among graduate ELT students.It emerged from interviews and observational data that three key constructs of assessment literacy were involved: conceptual, skills, practice (use), and reflection. Conceptually, students became more aware of theoretical models informing language assessment (e.g., Bachman's communicative language ability framework) and principles introduced in classroom assessment. Such awareness was cultivated through a

structured set of LMS modules, scaffolded readings and targeted reflection prompts. Linguistically, students also spoke of learning to 'name' features of discourse in relation to performance and feedback—for example, coherence, register and pragmatics—through the application of assessment metalanguage. "*It used to be that I'd judge writing by how it felt, but now I can articulate the discourse reasons, things like genre moves and tone — why it worked or didn't,"* wrote one student. (Student 13, Interview)

In applied competence, students appeared to develop greater competence over time in design critique and in the revision of assessment tasks through their task-artifact peer-review work. Findings from the project reports indicate that students integrated validity, reliability and fairness into the design of their tests. The LMS supported this through incremental submission and instructor scaffolding.

"*I never really thought about it until now, how many layers there are in a test. Following this workshop, I will be able to explain why each of them are in the test me all parts.*" (Student 7, Interview)

Reflective engagement was a key feature in the development of assessment literacy. Students were encouraged to maintain weekly journals that addressed the learning process and included prompts that corresponded to assessment criteria. These answers suggest greater awareness of assessment as a pedagogical tool and less as an instrument for grading. Students reported that assessment influenced their motivation, engagement, and learning outcomes. Some found that the way peer feedback influenced how they perceived scoring and fairness as they navigated differences in interpretation was problematic.

"*You learn that peer feedback is not just about correcting but learning to see through someone else's eyes."* (Student 4, Reflective Journal Week 7)

Trends in self-regulation and peer support can be observed in LMS interaction logs. Students often reread feedback, requested details and engaged in follow-up discussions with peers. These practices suggested a shift towards more autonomous and reflective evaluation. By the end of the semester, Raghuram said students' questions became more complicated: "what do I get for that score?" to "how can I make this rubric better?" On the whole, the LMS-embedded evaluation was found to support the multifaceted development of assessment literacy in this program. Alongside the tasks that were inherently connected in terms of content and that introduced and reinforced their diminishing learning performance, this resulted in students attaining a certain level of understanding, whether theoretical knowledge or practical skills, through critical reflection on assessment. These enhancements suggest that LMSs can be effective in supporting deep and sustained learning for assessment when pedagogically matched.

Discussion

This part summarizes findings related to the three research questions, embedded within the theoretical concepts of language testing principles and assessment literacy, as well as digital mediation enabled by LMS platforms. The findings validate LMS-driven assessment in pedagogy, with a need for linguistic grounding, instructional alignment and scaffolding for reflection in graduate ELT.

LMS-Based Assessment in Graduate Language Testing Course

From a linguistic perspective on an LMS-based assessment, it should be noted that incorporating these forms of assessment into a Language Testing course at the graduate level is not merely about technological change, because language use fundamentally shapes the practices of assessment-as-text, discourse interaction, and

communicative meaning-making in digital spaces. In that sense, LMS are not only testing spaces but also areas in which linguistic competence, academic discourse, and the communication of assessment can be continuously negotiated.

The first research question: "How was an LMS-supported assessment integrated into the graduate-level course, Language Testing?" The results show that SPADA enabled sequenced, theory-driven assessments that aligned with central language assessment goals. This aligns with the belief that LMS systems, when designed purposively, are not merely administrative conveniences but also pedagogical ecosystems in their own right, managing ongoing engagement and learning (Rapanta et al., 2021; Carless et al., 2021).

The combination of rubrics, collaborative feedback, and asynchronous modules provided opportunities for learners of varying backgrounds to engage with the theory of language testing. These findings are in line with those recently reported by DeLuca et al. (2022), which focused on the opportunities afforded by digital learning environments to support formative, dialogic assessment processes. Task sheets that rely on LMS functionality facilitated learning-oriented assessment design, following the feedback literacy model of Winstone and Carless (2019), with asynchronous discussions and peer feedback tasks.

Additionally, structured weekly modules promote self-regulated learning (Panadero et al., 2018), a key competency needed in graduate school. Consistent with Broadbent and Poon (2020), this report demonstrates that LMS design can influence learner agency, task engagement, and metacognitive growth. Crucially, the shift in students' focus from grade orientation to reflection on improvement is consistent with Vygotsky's (1978) socio-constructivist theory of learning, in which interaction and feedback mediate conceptual development.

Nevertheless, issues such as confusion at the outset about rubric use and LMS feedback tools point to a need for an instructor facilitator and digital literacy support (Tondeur et al., 2021). These findings support the claims of Martin et al. (2020), who argue that LMS use requires pedagogical intentionality and sufficient training to prevent shallow digital engagement.

Furthermore, this research contributes to the debate by demonstrating how a local institution-based LMS – SPADA was not just a delivery tool, but also a site of disciplining. The increasing ubiquity of the assessment design process and students' exposure to aspects of rubric use reveal a pedagogy that reflects what Lea and Street (2006) call an assumption within academia for the 'academic literacies' model, where students are simultaneously navigating content, discipline conventions, and forming an identity as they move through the disciplinary system. Here, LMS activities served as sites of disciplinary sense-making, especially in applied language studies and testing.

Educational LMSs such as SPADA, while lacking some of the toolsets available in commercial systems, have been found effective when integrated with tightly aligned courses. This is consistent with previous research by Boelens et al. (2017), who assert that effective digital assessment relies less on technical sophistication than on instructional integration. Moreover, incorporating formative assessment practices into the LMS architecture aligns with Nicol and Macfarlane‐Dick's (2006) guidelines for formative feedback, especially in promoting time-on-task and peer learning. Such findings are especially relevant in resource-limited settings, where open-source or government-hosted LMS platforms dominate. It echoes Beetham et al.'s (2019) point that pedagogical quality depends not on technology but

on teachers' design practices. The study also offers empirical support for Boud and Molloy's (2013) assertion that assessment should be conceived of as a learning process and highlights how digital environments can facilitate this process through transparent, dialogic, and recursive feedback.

Implementing Language Testing Principles via LMS-Assisted Testing

The second research question was to see how LMS-based assessment facilitated the application of four core principles of language testing: validity, reliability, fairness, and washback. The study revealed strong alignment between task design and these theoretical pillars, providing evidence of LMSs' promise to facilitate principled assessment practices. From a linguistics perspective, these principles are not only psychometric standards but also reflect how language is socially constructed, pragmatically negotiated, and discursively performed within LMS-mediated assessment environments.

Validity was established through analysis tasks in which students were asked to critique actual test results using theoretical approaches. This mirrors Messick's (1989) model of construct validity, which emphasizes the significance of relevant task representations of the intended construct. As Knoch and Chapelle (2021) note, validity in digital environments depends on task design, which they illustrate through a scaffolded application.

The validity of the tasks was also enhanced by having students develop their own testing instruments. These assessments involved what Brown and Abeywickrama (2019) define as construct-referenced decision-making: where students justify each component of the challenge in terms of theory. This degree of engagement promotes deeper conceptual understanding and enhances assessment competence, as well as the fidelity and validity of the assessment. Reliability was supported by rubric-mediated peer assessment and the ongoing use of benchmark scoring guides. Consistency in scoring, as stressed by McNamara and Knoch (2012), is crucial to ensure the fairness and accountability of assessment. It showed that with exposure to rubric-based feedback, students' scoring practices became more consistent over time. This result contradicts the study by Fulcher and Davidson (2007), which claims that initial training in rating scales and collaborative norming improves scoring reliability.

Moreover, using SPADA logs, we can trace peer-to-peer grading interactions, thereby extending the literature and providing supporting evidence for real-time reliability calibration within an LMS environment. Such documentation supports O'Sullivan's (2021) assertion that, when carefully structured, digital platforms can facilitate standardization and increased inter-rater reliability.

The ethical and operational principle of fairness was upheld through equal access to tasks, clear success criteria, and open, rolling submission deadlines. However, the analysis also highlighted equity issues in digital literacy and access. These results uphold the claim made by Harding and Sadeghi (2022) that equity in online assessment is not a given, especially in diverse, under-resourced learning contexts. Drawing on digital equity principles (Watts, 2019) and practices (Meyer et al., 2014) provides another means for promoting inclusive assessment.

Washback manifested itself in how students behave and approach learning. All students not only reworked their assignments in response to feedback but also felt more connected and motivated to read the course literature and theory more broadly. This is consistent with Watanabe's (2004) categorization of washback as systemic course design or individual learner behavior, and the present study has shown that both are at work within the LMS architecture.

Significantly, students' reflections demonstrated recognition of these ramifications in assessments under consideration for studies, as well as an emerging ethical awareness of the power relations embedded in how tests are designed. This echoes Green's (2007) concept of critical washback, in which learners not only react to assessment pressures but also begin to question the fairness and potential repercussions of such assessments. This kind of critical engagement indicates the possibilities offered by LMS-taught assessment for moving beyond procedural learning and towards principled practice.

Unlike much prior LMS-assessment research that foregrounds general platform affordances and learner perceptions, the novelty of this study lies in treating LMS-mediated testing as a linguistic discursive practice that reveals how rubrics, feedback language, and task genres within SPADA were deployed to make meaning about assessments undertaken in a graduate Language Testing course.

By delinking language-mediated processes from the four testing principles (validity, reliability, fairness and washback) as well as from the construction of assessment literacy, the study expands previous work on practical relationships to reveal how LMS-based assessment becomes effective not just through instructional alignment but also through linguistic design within assessment communicative contexts.

Theoretically, this research contributes to language-testing scholarship by extending tenets of validity, reliability, fairness and washback into the ecosystem of an LMS as language-mediated constructs, showing how principles of assessment are actualized through task-rubric-feedback discourses and not merely viewed as psychometric properties. Practically, the study offers an evidence-based blueprint for graduate ELT programs using institutionally hosted LMSs like SPADA by aligning sequenced modules with assessment principles, standardizing rubrics and benchmark guides to achieve rater consistency, and designing feedback-and-revision cycles that make assessment literacy observable through reflective writing. In resource-limited contexts, the findings also suggest that pedagogical and linguistic design (not platform sophistication) is the lever that delivers scalable, defensible LMS-mediated language assessment.

Conclusion

The study investigates the integration of LMS-based assessment into a graduate-level Language Testing course in an ELT program, focusing on whether it proceeds in accordance with core LT principles and cultivates test literacy. Using the SPADA LMS at Universitas Muhammadiyah Makassar, the research centered on three dimensions: how assessment tasks were implemented through the LMS; how validity, reliability, fairness and washback were employed; and what these practices did to foster graduate students' assessment literacy as prospective ELT practitioners.

The results suggest that LMS-based assessment can create focused and relevant learning when informed task design and theoretical alignment of the assessment communication occur. SPADA facilitated scaffolded assessment cycles through structured modules, rubric-guided tasks, peer discussion and reflection. Learners were able to asynchronously interact with the material, work collaboratively with fellow students , and engage in feedback activities to facilitate increased learning. These activities were not procedural alone, but were also linguistically rich events in which students participated in academic talk, negotiated the meaning of assessment criteria, and created evaluative language that represents professional or expert-like assessment practice.

The study also indicates that core test design principles can be successfully implemented in LMS environments, provided they are interpreted through the paradigms of language use in technology-mediated spaces. Validity was enhanced through tasks that elicited communicative performance and asked students to explain design choices in terms of unit constructs and discourse demands, respectively. Reliability of the rating was assured through rubric application and iterative peer calibration, but relied on common readings of linguistic category descriptors such as coherence, register and appropriateness. Fairness entailed not only equal opportunities and time response, but also awareness of social linguistic mismatching and style variation, particularly in the case of peer assessment and academic tasks. Washback developed through cycles of feedback revision and reflective use, suggesting that LMS spaces can support reformist assessment cultures.

One central contribution of this study is its linguistic perspective: LMS assessments are not a neutral, technical process but rather a language-mediated activity. Testing in SPADA occurred through discourse such as task instructions, rubric descriptors, feedback comments, peer negotiation, and reflective writing, and the language of the test context was simultaneously a content of testing and an instrument with which testing happened. This reinforces the idea that successful LMS-based language testing should consider discourse, genre, pragmatics and the sociocultural construction of meaning. If the assessment activities and feedback are designed with these linguistic considerations in mind, the LMS can be seen as an ecology for developing both communicative competence and assessment literacy.

Pedagogically, the results emphasize that effective LMS-based assessment depends on deliberate design: clear criteria, accessible task instructions and feedback, timely feedback and opportunities to reflect and review. Institutions should also enhance digital and assessment literacy support for faculty and students alike, acknowledging that technological access and acuity influence participation and achievement. While this study was situated within one course at one institution, its transferable findings can be applied to comparable postgraduate ELT contexts in which the LMS is increasingly a dominant component of assessment practice.

Thus, LMS-facilitated language assessment can be an ally to means and literacy-oriented assessment in ELT programs as long as its effectiveness depends on linguistic grounding. Organizing LMS-based test and assessment as discourse practice, considering genre, register, interaction and pragmatic meaning, enables assessment to reflect real language use and prepares graduate students to become linguistically consciously and ethically responsible and professionally competent testers.

The study implies that LMS-mediated language testing in graduate ELT programs should be designed as a linguistically informed discourse ecology**,** where task prompts, rubric descriptors, and feedback interactions are intentionally aligned to strengthen validity, reliability, fairness, and positive washback while simultaneously cultivating students' assessment literacy. Practically, institutions should pair rubric-guided, feedback-rich task sequencing with explicit support for digital competencies and platform limitations, because the quality of assessment depends not only on tools but also on how assessment meanings are co-constructed through language and interaction within the LMS.

Future research should test the transferability of these findings across different LMS platforms and institutional contexts, for instance, commercial LMSs versus government-hosted systems, to determine which linguistic and pedagogical design features are robust and which are context-sensitive. Further studies are also needed to

examine how specific discourse elements, such as rubric wording, feedback moves, and the genres of reflection prompts, shape assessment literacy development over time, ideally using longitudinal designs and multimodal discourse analysis of interaction logs, annotations, and revisions. Finally, larger-scale or comparative investigations should address equity and fairness by mapping how digital access, language proficiency, and digital literacy mediate participation and scoring outcomes, and by evaluating targeted interventions such as rater calibration training, feedback-literacy microtasks, and accessibility supports within LMS-mediated language testing.

AI Tools Declaration

The authors used AI tools (Grammarly and Gemini) solely for language editing and proofreading to improve clarity and grammatical accuracy. All intellectual content, data analysis, interpretations, and conclusions presented in this manuscript are the original work of the authors.

Bibliographic references

Alruwais, N., Wills, G., & Wald, M. (2018). Advantages and challenges of using e-assessment. International Journal of Information and Education Technology, 8(1), 34–37. https://doi.org/10.18178/ijiet.2018.8.1.1008

Andrade, H., & Brookhart, S. M. (2020). Student self-assessment: Making it work in the classroom. ASCD.

Bachman, L. F. (1990). Fundamental considerations in language testing. Oxford University Press.

Bachman, L. F., & Palmer, A. S. (1996). Language testing in practice: Designing and developing useful language tests. Oxford University Press.

Bachman, L. F., & Palmer, A. S. (2010). Language assessment in practice: Developing language assessments and justifying their use in the real world. Oxford University Press.

Baird, J., & Fisher, L. (2005). Pedagogical practices and assessment: Methodological issues in a classroom-based project. Oxford Review of Education, 31(2), 211-225. https://doi.org/10.1080/03054980500130072

Beetham, H., Sharpe, R., & Benfield, G. (2019). Rethinking pedagogy for a digital age: Principles and practices of design. Routledge. https://doi.org/10.4324/9780429506345

Belz, J. A., & Kinginger, C. (2003). Discourse options and the development of pragmatic competence by classroom learners of German: The case of address forms. Language Learning, 53(4), 591-647. https://doi.org/10.1111/1467-9922.00236

Boelens, R., De Wever, B., & Voet, M. (2017). Four key challenges to the design of blended learning: A systematic literature review. Educational Research Review, 22, 1-18. https://doi.org/10.1016/j.edurev.2017.06.001

Bennett, R. E., Zhao, Y., & Jenkins, F. (2020). Assessing the quality of technology-enhanced assessments in higher education.

Assessment & Evaluation in Higher Education, 45(4), 547-561. https://doi.org/10.1080/02602938.2019.1682111

Boud, D., & Molloy, E. (2013). Feedback in higher and professional education: Understanding it and doing it well. Routledge. https://doi.org/10.4324/9780203074336

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

Broadbent, J., & Poon, W. L. (2020). Self-regulated learning strategies and academic achievement in online higher education learning environments: A systematic review. The Internet and Higher Education, 51, 100814. https://doi.org/10.1016/j.iheduc.2021.100814
Brookhart, S. M. (2011). Educational assessment knowledge and skills for teachers. Educational Measurement: Issues and Practice, 30(1), 3-12. https://doi.org/10.1111/j.1745-3992.2010.00195.x
Brown, H. D., & Abeywickrama, P. (2019). Language assessment: Principles and classroom practices (3rd ed.). Pearson Education.
Canale, M., & Swain, M. (1980). Theoretical bases of communicative approaches to second language teaching and testing. Applied Linguistics, 1(1), 1-47. https://doi.org/10.1093/applin/I.1.1
Carless, D., & Chan, K. K. H. (2017). Managing dialogic use of exemplars. Assessment & Evaluation in Higher Education, 42(6), 930-941. https://doi.org/10.1080/02602938.2016.1211246
Carless, D., Winstone, N., Chan, K. K. H., & Tuck, J. (2021). Teacher feedback literacy and its interplay with student feedback literacy. Teaching in Higher Education, 28(2), 147-163. https://doi.org/10.1080/13562517.2021.1948783
Chapelle, C. A. (2001). Computer applications in second language acquisition: Foundations for teaching, testing and research. Cambridge University Press.
Chapelle, C. A., & Chung, Y. R. (2021). The evolving construct of communicative language ability in an age of multilingualism.
Language Testing, 38(4), 587-606. https://doi.org/10.1177/02655322211020467
Creswell, J. W., & Poth, C. N. (2018). Qualitative inquiry and research design: Choosing among five approaches (4th ed.). Sage Publications.
Currie, P. (Ed.). (2004). Discursive constructions of identity in academic writing. John Benjamins Publishing Company.
Dahlstrom, E., Brooks, D. C., & Bichsel, J. (2020). The LMS evolution: Learning management systems in higher education. Educause Center for Analysis and Research. https://library.educause.edu/resources/2014/5/the-evolution-of-the-lms-from- management-to-learning
DeLuca, C., Coombs, A., & LaPointe-McEwan, D. (2022). Assessment for learning in online environments: A scoping review of empirical studies. Assessment in Education: Principles, Policy & Practice, 29(3), 321-342. https://doi.org/10.1080/0969594X.2022.2054390
Dijk, Van, T. A. (2006). Discourse and manipulation. Discourse & Society, 17(3), 359-383. https://doi.org/10.1177/0957926506060250
Douglas, D. (2000). Assessing languages for specific purposes. Cambridge University Press.
Douglas, D. (2021). Language testing and technology: Past directions and future developments. Language Testing, 38(2), 221-240. https://doi.org/10.1177/0265532220966724
Ellis, R. (2003). Task-based language learning and teaching. Oxford University Press.
Fulcher, G. (2020). Essentials of language testing (2nd ed.). Routledge. https://doi.org/10.4324/9781315679945
Fulcher, G., & Davidson, F. (2007). Language testing and assessment: An advanced resource book. Routledge.
García, O., & Wei, L. (2014). Translanguaging: Language, bilingualism and education. Palgrave Macmillan.

Green, A. (2007). IELTS washback in context: Preparation for academic writing in higher education. Cambridge University Press. https://doi.org/10.1017/CBO9780511667322
Harding, L., & Sadeghi, B. (2022). Test fairness in online language assessments: Perspectives from language testers. Language Assessment Quarterly, 19(1), 1-20. https://doi.org/10.1080/15434303.2021.2004131
Kane, M. T. (2021). Validity as a pragmatic, scientific concept. Language Testing, 38(3), 367-377. https://doi.org/10.1177/0265532221990811
Khalil, R., & Elkhider, I. (2022). Students' perceptions of e-assessment during COVID-19: Evidence from an ELT program. Journal of Language Teaching and Research, 13(2), 219-226. https://doi.org/10.17507/jltr.1302.05
Kocdar, S., Karadeniz, A., Bozkurt, A., & Büyük, K. (2018). Measuring self-directed learning skills in e-learning environments: A scale validation study. The International Review of Research in Open and Distributed Learning, 19(1), 25-43. https://doi.org/10.19173/irrodl.v19i1.3255
Knoch, U., & Chapelle, C. A. (2021). Validation of writing assessments in the digital age: Issues and opportunities. Assessing Writing, 50, 100571. https://doi.org/10.1016/j.asw.2021.100571
Kress, G. (2015). Multimodality: A social semiotic approach to contemporary communication. Routledge. https://doi.org/10.4324/9781315748818
Lea, M. R., & Street, B. V. (2006). The "academic literacies" model: Theory and applications. Theory into Practice, 45(4), 368-
377. https://doi.org/10.1207/s15430421tip4504_11
Martin, F., Polly, D., & Ritzhaupt, A. D. (2020). Bichronous online learning: Blending asynchronous and synchronous online learning. TechTrends, 64, 128-135. https://doi.org/10.1007/s11528-019-00436-0
McNamara, T., & Knoch, U. (2012). Theoretical underpinnings of test validation. In G. Fulcher & F. Davidson (Eds.), The Routledge Handbook of Language Testing (pp. 9-21). Routledge.
Messick, S. (1989). Validity. In R. L. Linn (Ed.), Educational measurement (3rd ed., pp. 13-103). Macmillan.
Merriam, S. B., & Tisdell, E. J. (2016). Qualitative research: A guide to design and implementation (4th ed.). Jossey-Bass.
Messick, S. (1989). Validity. In R. L. Linn (Ed.), Educational measurement (3rd ed., pp. 13-103). Macmillan.
Meyer, A., Rose, D. H., & Gordon, D. (2014). Universal design for learning: Theory and practice. CAST.
Nappu, S, Muhayyang, M, Dewi, R, & Hamid, R. (2025). Peer feedback on EFL students' essay writing through Google Classroom: A multilingualism study. XLinguae, 18(4), 98-112. https://doi.org/10.18355/XL.2025.18.04.08
Nappu, S., Dewi, R., Hasnawati, & Hamid, A. (2022). Online learning and EFL students' academic writing performance: Exploring engagement and instructional design. VELES: Voices of English Language Education Society, 6(2), 175-187. https://doi.org/10.29408/veles.v6i2.5220
Nicol, D. J., & Macfarlane-Dick, D. (2006). Formative assessment and self-regulated learning: A model and seven principles of good feedback practice. Studies in Higher Education, 31(2), 199-218. https://doi.org/10.1080/03075070600572090

Nicol, D., & McCallum, S. (2022). Making internal feedback explicit: Exploiting the potential of digital feedback. Assessment & Evaluation in Higher Education, 47(6), 839-852. https://doi.org/10.1080/02602938.2021.1934525
Ockey, G. J., & Wagner, E. (2021). Assessing L2 students in the digital age. Cambridge University Press. https://doi.org/10.1017/9781108865791
O'Sullivan, B. (2021). Issues in assessing language through technology. Language Testing, 38(2), 161-168. https://doi.org/10.1177/0265532221990812
Panadero, E., Jonsson, A., & Botella, J. (2021). Effects of self-assessment on self-regulated learning and self-efficacy: Four meta-analyses. Educational Research Review, 22, 74-98. https://doi.org/10.1016/j.edurev.2017.08.004
Panadero, E., Andrade, H., & Brookhart, S. (2018). Fusing self-regulated learning and formative assessment: A roadmap of where we are, how we got here, and where we are going. Australian Educational Researcher, 45(1), 13-31. https://doi.org/10.1007/s13384-018-0258-y
Patton, M. Q. (2015). Qualitative research & evaluation methods (4th ed.). Sage Publications.
Rasooli, A., Timmermans, A. C., & Pilot, A. (2021). The role of assessment literacy in teachers' classroom assessment practices. Studies in Educational Evaluation, 72, 101106. https://doi.org/10.1016/j.stueduc.2021.101106
Sun, Y., & Zhang, Y. (2022). Discourse-oriented assessment in online writing: Exploring peer feedback through genre analysis. Language Testing in Asia, 12, https://doi.org/10.1186/s40468-022-00178-y
Stiggins, R. J. (1995). Assessment literacy for the 21st century. Phi Delta Kappan, 77(3), 238-245.
Tai, J., Ajjawi, R., Boud, D., Dawson, P., & Panadero, E. (2018). Developing evaluative judgement: Enabling students to make decisions about the quality of work. Higher Education, 76(3), 467-481. https://doi.org/10.1007/s10734-017- 0220-3
Watts, J. (2019). Equitable access to digital assessment in higher education: The role of inclusive learning design. Assessment & Evaluation in Higher Education, 44(5), 720-733. https://doi.org/10.1080/02602938.2018.1531747
Watanabe, Y. (2004). Methodology in washback studies. In L. Cheng, Y. Watanabe, & A. Curtis (Eds.), Washback in language testing: Research contexts and methods (pp. 19-36). Lawrence Erlbaum Associates.
Winstone, N. E., Nash, R. A., Rowntree, J., & Bannister, S. (2021). Supporting learners' agentic engagement with feedback: A systematic review. Educational Psychologist, 56(1), 1-20. https://doi.org/10.1080/00461520.2020.1841839
Xu, Y., & Brown, G. T. L. (2016). Teacher assessment literacy in digital environments: A review of research and practice. Assessment in Education: Principles, Policy & Practice, 30(2), 133-151. https://doi.org/10.1080/0969594X.2022.2148895

Words: 9753
Characters: 72 697 (40 standard pages)

Dr. Syamsiarna Nappu, M.Pd.
Magister Pendidikan Bahasa Inggris
Universitas Muhammadiyah Makassar,
Makassar, 90221
Indonesia

syamsiarna.nappu@unismuh.ac.id
ORCHID ID: 0000-0001-7115-7011

Prof. Dr. Maemuna Muhayyang, M.Pd.
Department of English Education
Universitas Negeri Makassar, 90222
Indonesia
maemarasyid@unm.ac.id
ORCHID ID: 0009-0005-3804-8338

Dr. Ratna Dewi, S.S., M.Hum.
Magister Pendidikan Bahasa Inggris
Universitas Muhammadiyah Makassar,
Makassar, 90221
Indonesia
ratnadewi@unismuh.ac.id
ORCHID ID: 000-0002-8959-919X

Dr. Nurdevi Bte Abduh, M.Pd.
Department of English Education
Universitas Muhammadiyah Makassar,
Makassar, 90221
Indonesia
nurdevi@unismuh.ac.id
ORCHID ID: 0000-0003-1922-1594

Dr. Hasnawati, M.Pd.
Department of English Education
Universitas Muhammadiyah Makassar,
Makassar, 90221
Indonesia
hasnawati@unismuh.ac.id
ORCHID ID: 0000-0002-1582-1800