The EALTA Special Interests Groups
Classroom-based Language Assessment (CBLA SIG)
&
Multilingual Language Assessment (MLA SIG)

Wednesday, 28 May 2025, 14.00–17.15 CET

(prior to 21st EALTA Conference in University of Education Salzburg, Pädagogische Hochschule Salzburg Stephan Zweig, Akademiestraße 23/25, Salzburg/ Austria https://ealta-annual-conference.org/)

14.00 WELCOME

14.00-14.20 Asma Maaoui & Amel Jarraya: *'Beyond self-reports: Investigating EMI students' real use of AI-powered self-assessment tools'*
14.20-14.40 Vera Vesala & Dmitri Leontjev: *'"I think the most interesting thing is that this would help them": Developing classroom assessment practices in dialectical teacher-researcher collaboration'*
14.40-15.00 Karim Rezagah: *'Transforming Iranian EFL teachers' assessment literacy: The impact of a professional development program'*
15.00-15.20 Melanie Alexandra Zedler: *'Approaching multilingual testing – The more the merrier or lost in translation? Preliminaries and challenges in multilingual-by-design tests for the foreign language classroom'*

15.20-15.30 BREAK

15.30-15.50 Asma Maaoui & Meryem Salhi: *'AI and formative assessment literacies as predictors of AI integration in assessment practices'*
15.50-16.10 Elisa Bertoldi & Sara Gesuato: *'Exploiting AI to provide corrective-constructive feedback on student writing: the case of EFL email requests to faculty'*
16.10-16.30 Eleni Kartalia*: 'EFL teachers' self-reflective awareness of assessment instruments and practices as metacognitive flashbacks'*
16.30-16.50 Valentina Cotesta*: 'Recognizing plurilingual adult migrants' Agency in multilingual language assessment'*
16.50-17.10 Katharina Karges: *'Investigating a multilingual reading assessment: Are tasks comparable? '*

17.10-17.15 CLOSING REMARKS

Beyond self-reports: Investigating EMI students' real use of AI-powered self-assessment tools

Asma Maaoui, & Amel Jarraya
University of Tunis

AI-powered tools have been adopted by learners to enhance their academic performance, especially in various Higher education contexts. Research has demonstrated that AI can facilitate formative assessment, adaptive learning, and personalized feedback, enabling students to enhance self-regulation. Despite the growing body of research on AI-assisted learning, the actual use and real-life experiences of higher education students with AI-driven self-assessment tools remain an under-explored area. Besides, most studies to date have addressed cognitive outcomes or the evaluation of AI models in relation to learning and assessment. To address these gaps, this study investigates English as a Medium of Instruction (EMI) students while relying on samples of learners' real performance data during AI-assisted self-assessment. Thematic analysis was conducted to categorize and interpret patterns in AI-driven self-assessment practices. The findings reveal varying patterns of AI use for self-assessment alongside challenges the learners faced in using the tools effectively and ethically. The findings may contribute to understanding the role of AI in learning while also emphasizing its potential for self-assessment. This may also have significant implications to future pedagogical interventions aiming to enhance learning via AI-powered assessment.

"I think the most interesting thing is that this would help them": Developing classroom assessment practices in dialectical teacher-researcher collaboration

Vera Vesala & Dmitri Leontjev
University of Jyväskylä

We report on a part of the DD-LANG project that combines dynamic (DA) and diagnostic assessment, whose central aim is to develop English teachers' conceptions about assessing reading alongside and intertwined with the development of their classroom assessment practices. Our work with the teachers is strongly informed by Vygotskian praxis, a dialectical teacher-researcher collaboration where theory and practice are interconnected (Poehner & Inbar-Lourie, 2020). In this presentation, we focus, namely, on the teachers' engagement with (computerised) DA and its theoretical underpinnings and, more importantly, on how they use the information from the learners' computerised DA performance in the classroom. DA focuses on diagnosing learner abilities in the process of development and combines teaching and assessment into one development-oriented activity (Poehner & Leontjev, 2020). We base the presentation on interviews (n = 15), which provided the baseline for meetings with teachers where we actively used theoretical concepts. During these meetings, we invited teachers to reflect on their practices through the lens of theoretical concepts

while they simultaneously built their conceptual understanding of classroom-based assessment. We will lead the audience through teacher developmental process, as we together discussed and designed ways that the teachers could use the information from computerised DA in their classroom.

Transforming Iranian EFL teachers' assessment literacy: The impact of a professional development program

Karim Rezagah
Jyväskylä University Finland

The Iranian Ministry of Education initiated a shift in 2010, adopting communicative approaches to teaching English in schools. Regarding assessment, teachers are required to harmonize assessments of learning (AoL, evaluating students' achievements at the end of an instructional period) and assessments for learning (AfL, enhancing learning during an instructional period) in classroom (Foroozandeh & Forouzani, 2015). However, due to a lack of exposure to practical AfL training this shift has not yet resulted in integrating AfL and AoL in practice (Gholaminejad & Raeisi-Vanani, 2021). In this presentation, I cover a study that addressed this challenge by proposing a practical AfL-oriented Professional Development Program (PDP) designed as a dialectical teacher-researcher partnership. The study further argues that such engagements, rather than unidirectional training, allow for resolving tensions that Iranian teachers face in merging AoL and AfL. This presentation focuses on the developmental trajectory of one teacher among four participating teachers. Data, from open-ended pre-post-intervention questionnaires, teachers' reflective assignments, and their engagement with each other's reflections on a forum were analyzed using Sociocultural Discourse Analysis (Johnson & Mercer, 2019). The key findings reveal how collaboration between teachers and researchers, as language assessment stakeholders, influences the teaching-learning and assessment processes.

Approaching multilingual testing – the more the merrier or lost in translation? Preliminaries and challenges in multilingual-by-design tests for the foreign language classroom

Melanie Alexandra Zedler,
University of Bremen, Germany

De Angelis' (2021) suggests a trifold test-design distinction (i.e. monolingual, multilingual-by-translation, and multilingual-by-design) and outlines respective perks (e.g. potential benefits for students) and drawbacks (e.g. workload for the test designer). As a first step, we will point out crucial elements of the multilingual test-design process, i.e. (a) knowing the target group, (b) significance of the teacher's choice(s) in test design, and (c) general language-specific suggestions which may improve performance in any type of test. In continuation, we will contrast the impact

of the presence/absence of these elements in sample exercises from the examinations of a Croatian-Polish taster course held at the Language Center at the University of Bremen in Summer 2023. Among others, we will listen to/look at items of a multilingual listening exercise and sample exercises of language comparison/discrimination. Consulting several preliminary versions of the examinations, we will discuss the use/drawbacks of multilingual and translated instructions vs. instructions intentionally written in just one language. Finally, the value of the element of choice (for language and/or the order of approaching the tasks), and mediation during the multilingual oral exam are explored.

AI and formative assessment literacies as predictors of AI Integration in assessment practices

Asma Maaoui
University of Tunis, Tunisia

Meryem Salhi
Medtech-SMU, Tunisia

Research on the integration of AI in formative assessment has shown its potential in enhancing learning through personalized feedback, self-regulation, and critical thinking. Despite the significant role of teacher preparedness in the effective implementation of AI-powered formative assessment, research on educators' AI and formative literacies and their impact on formative assessment practices remains scarce. To address this gap, this study investigated Tunisian secondary and tertiary educators' AI and formative assessment literacies in relation to their use of AI-supported formative assessment tools and methods. Focusing on both language and non-language teachers, the study explored the extent to which these types of literacies influence their adoption of AI for personalized feedback, peer assessments, automated scoring, and gamified learning analytics. Data were collected through an online survey targeting 150 teachers across disciplines. The study primarily relied on a quantitative methodology employing Pearson correlations and simple linear regression to analyze whether AI and formative assessment literacies serve as predictors for AI-assisted formative assessment practices. As a work in progress, the study findings aim to provide insights into these teachers' readiness for AI integration versus their practices in formative assessment, contributing to more effective technology integration for formative assessment purposes in diverse pedagogical contexts.

Exploiting AI to provide corrective-constructive feedback on student writing: the case of EFL email requests to faculty

Elisa Bertoldi & Sara Gesuato
University of Padua, Italy

Digital correspondence remains a dominant form of communication in professional contexts, including in academia, where students frequently email professors. However, students often struggle to adapt their communicative practices to formal email conventions when addressing superiors, especially in second/foreign language contexts (Stephens et el., 2009; Savić et al., 2021). AI can support writing skills development (e.g., Amyatun & Kholis, 2023; Chen & Gong, 2025), particularly through automated feedback (e.g., Silitonga et al., 2023), though its application in pragmatics remains limited (e.g., Enright & Quinlan, 2010; Gerhalter, 2024; Andersson & McIntyre, 2025).

This study explores the potential of *Lucrez-IA*, a Large Language Model (LLM) developed at Padua University, to provide feedback on student emails to faculty. Fifty spontaneous requestive emails (7,500 words) written by EFL students were analyzed, focusing on opening salutations, self-introductions, closing pleasantries, and signature lines. A rubric was developed to assess these features, which were then coded into *yes-no* questions with illustrative examples, and transformed into prompts for *Lucrez-IA*. The model's performance was iteratively refined through testing on a subset of the data. Its feedback was then compared with human assessments. Results indicate a high convergence between *Lucrez-IA*'s feedback on and our assessment of the structure and wording of the emails' framing moves. We argue that LLMs such as *Lucrez-IA* could support the teaching of formal email writing in two ways: first, AI queries can serve as a checklist guiding students' writing process, and second, AI feedback can serve as a monitoring tool for focused self- and peer assessment of written products. Overall, this study suggests that refining prompt engineering can enhance AI's ability to provide targeted feedback on formal writing conventions essential for academic and professional success.

EFL teachers' self-reflective awareness of assessment instruments and practices as metacognitive flashbacks

Eleni Kartalia
European University Cyprus, Greece

Assessment-related reflective practice is a key determinant of EFL teachers' Language Assessment Literacy (LAL), shaping their conceptual understanding of assessment culture and pedagogical choices (Tsagari & Vogt, 2017). Teachers' self-reflective awareness is instrumental in classroom-based assessment (CBLA) by impacting the

selection of assessment instruments, adaptation of instructional design and refining teachers' professional identity (Xu & Brown, 2016).

This study introduces the concept of metacognitive flashbacks, a psychological phenomenon rooted in metacognitive reflection that drives teachers' assessment decision making and extends the development of LAL (Merkebu et al., 2023; Purpura, 2016; Bøhn & Tsagari, 2022). Therefore, metacognitive flashbacks are considered as a LAL reflective instrument elucidating teachers' memory-based inferences through the conscious process of retrieval (ecphory), influencing assessment tendencies (Teng et al., 2022; Frankland et al., 2019). Additionally, this research proposes a framework of Self-Reflective Awareness providing actionable strategies for improving LAL, fostering professional growth, and strengthening institutional support.

A qualitative research design was employed, utilizing case studies and semi-structured interviews, analyzed using Reflexive Thematic Analysis (RTA) (Braun & Clarke, 2022). The findings reveal that variability in metacognitive flashbacks is interrelated with professional experience, institutional support, and assessment training. The empirical evidence supports the integration of metacogntive reflection theory to LAL.

Recognizing plurilingual adult migrants' agency in multilingual language assessment

Valentina Cotesta
University of Palermo - University of Foreigners of Siena

Despite the growing recognition of multilingualism, language assessment still lacks shared practices that effectively address plurilingual competence. This stagnation is closely tied to the complexity of plurilingualism as a construct of validity. However, dominant political and cultural ideologies reinforce a "one nation, one language" paradigm, sustaining monolingual assessment models that prioritize native-like proficiency.

In today's globalized world, such assessment practices fail to reflect real-life communication experiences of plurilingual adult migrants. Current language tests not only disregard their plurilingual repertoires and identities – denying them recognition and cultural value – but also fail to acknowledge their agency in multilingual interactions, where linguistic flexibility is a fundamental resource.

This contribution critically examines the limitations of monolingual assessment and its inequities for plurilingual migrants. It explores alternative multilingual assessment practices, drawing on integrated approaches and translanguaging to develop more inclusive and valid methods of evaluating plurilingual competence. By shifting away from monolingual biases, these practices recognize linguistic diversity as a resource and ensure fairer, more representative assessment.

Investigating a multilingual reading assessment: Are tasks comparable?

Katharina Karges
University of Leipzig - University of Fribourg/Switzerland

One of the key challenges in multilingual language assessment is ensuring that tasks and items are equally difficult across languages. To fairly assess a person's multilingual repertoire, we need tasks that allow for comparable conclusions in every language. However, little is known about how item and task difficulty vary across languages.

This study examines 39 reading texts and 94 items in four language versions, analysing both linguistic characteristics—such as lexical diversity, lexical difficulty, and the presence of cognates—and test-taking processes. The data comes from a large-scale assessment of foreign language reading skills in sixth graders (N=19,357) at a low proficiency level (A1/A2.1), assessing French, German and English as foreign languages. The assessment used tasks that were essentially translations of each other, which were created in parallel based on the same construct and blueprint. The findings highlight the complexities of multilingual assessment: Item difficulty varies notably across language versions, as do many linguistic features of texts and items. Quantitative analyses reveal multiple factors influencing item difficulty across languages, while qualitative findings suggest that cognates may contribute to item difficulty, emphasising the role of test takers' multilingual repertoires.