Technology and Artificial Intelligence for learning, teaching, and assessing across languages and modalities: Recent developments and insights from Finland

Ari Huhta & Dmitri Leontjev
University of Jyväskylä, Finland

This special issue of *Studies in Language Assessment* was born from a colloquium at the AIRiAL (Artificial Intelligence Research in Applied Linguistics) conference in New York in 2024, which the editors of the issue convened. Most of the papers have, therefore, been expanded from the presentations at the said colloquium, which focused on recent developments in Finland that leverage AI and other technological developments to support L2 learning, teaching, and assessment. What makes Finland a particularly interesting context to focus on is the emphasis on formative assessment in its educational policy documents and curricula (e.g., FNAE, 2016) and the increasing use of technologies, particularly in secondary education. Therefore, before introducing the articles in this issue, we give a brief account of the key characteristics of the Finnish educational system, one of which is the teachers' responsibility for almost all assessments, and of the role and impact of technology and AI in language assessment.

The educational system in Finland is characterized by decentralisation and public schools (private schooling or home education is quite rare). While the National Core Curricula for primary as well as for lower and upper secondary education define the objectives and core contents of subjects, local education providers (i.e., municipalities, towns), schools and ultimately teachers have considerable freedom in deciding how to teach and assess and what materials to use. Textbooks play an important role in language teaching as teachers rely on their contents, exercises and tests in their day-to-day work (see, e.g., the large-scale survey of pedagogical practices in lower

Email address for correspondence: ari.huhta@jyu.fi

secondary education by Luukka et al., 2008). Nowadays, textbooks, and thus the tests that are included in them, are increasingly available in digital format. Particularly in general upper secondary education (for 16 to 18-year-olds), textbooks are often available only digitally and the students study using laptops.

In line with the educational system as a whole, assessment is almost entirely decentralised and is the teachers' responsibility, especially in primary and lower secondary schools, including the high-stakes assessments at the end of the latter, in grade nine, after which the students seek entry to their preferred general or vocational upper secondary schools. The only national, school-external examination in the country is the Matriculation Examination (ME) at the end of general upper secondary education. Besides being the summative school-leaving examination for a range of subjects, including several languages, the ME serves a gatekeeping function (see, e.g., Nguyen, 2021) as it is used to gain entry into universities and other post-secondary education. The ME has been digital since the late 2010s.

Artificial intelligence is having an impact on all levels of education in Finland, but since it is a relatively recent phenomenon, both policy makers and practitioners such as teachers struggle to decide how to best approach it and how to utilize it meaningfully. Students, particularly teenagers and young adults in secondary and tertiary education, use AI in their free time, including to complete their homework. Whether to allow various AI tools to write the assignments given by the teacher is an issue that worries educators and draws attention in public discussions of the impact of AI on education. Reactions vary greatly from banning the use of all AI tools in schools (e.g., City of Helsinki in 2025) to allowing differing ways of AI use for schoolwork. Thus, it appears that the role of AI is most controversial whenever students have to write something outside the classroom as part of their studies. One consequence of the ubiquitous availability of AI tools is that teachers are giving up home essays, or they are rethinking how to incorporate AI in the writing process while somehow ensuring that they know what the student's own contribution to the task is and what part AI plays in it. National educational authorities have begun to issue recommendations to schools and teachers regarding AI; however, given the decentralised nature of the educational system, each municipality and school can decide for themselves to what extent they want to regulate AI.

Applying AI to assessing learners' language skills has obviously begun at the same time as the more general introduction of AI to other aspects of language education. We are not aware of any representative surveys of how AI is used in assessment and for what purposes. Anecdotal evidence suggests that both teachers and students are using AI tools for feedback and to get an informal evaluation of the quality of the students' texts. These are clearly formative types of assessment, and we are not aware of any (widespread) uses of AI for summative purposes. One interesting development in high stakes testing in Finland is worth mentioning, however. The Matriculation Examination has started to apply automated assessment of some short-answer questions in the reading and listening parts of its language examinations. These questions are presented to the students in either Finnish or Swedish, depending on the language of education of their school, and the students reply in that language (Finland has two parallel strands of education for the speakers of the two national languages). The ME uses a system based on large language models to automatically score the most straightforward and most obviously correct or incorrect replies. Human raters check the machine scored responses, however, and assess those that the automatic system left unscored. The system was introduced only in the spring of 2025, so by the time of writing this introduction to the special issue, it has been used in just two examination rounds. Currently, the system appears to be capable of scoring 20–50% of the students' responses, depending on the question. Some recent research also paves the way for using AI-based assessment to support human raters in the speaking component to be introduced into the ME language tests in the future (see Hildén et al. in this issue).

The five papers forming this special issue collectively explore the applications of artificial intelligence in language assessment in Finland, approaching this topic from various angles, focusing on different languages, subskills, and proposing different roles that AI can play in language assessment and the development of language abilities.

A number of considerations shaped the order of the papers in this issue. As all papers have a strong connection to assessment supporting learning, their order was rather informed by (a) the language(s) which they deal with (b) the completeness of the research reported in them, and (c) the focus of each article (assessment vs a combination of learning, teaching and assessment).

The first three papers all focus on the national languages of Finland, Finnish and Swedish, whereas the remaining two articles either cover a number of languages, including the national languages, or focus on an international language (English) that plays an important role in the educational system in Finland. The first paper on the DigiTala project (Hildén et al.) covers the work done on developing automatic speech recognition and assessment of spoken L2 Finnish and L2 Swedish – or more precisely, for Finland Swedish, a variant of Swedish spoken in Finland. The second paper, on the AASIS project (Ullakonoja et al.), is authored largely by members of the research team that carried out DigiTala, and the project described in the paper is in many ways a continuation of that earlier research. AASIS involves L2 Finnish but moves the focus from the automatic assessment of speech production (i.e., monologue tasks) to interactive speaking and investigates how non-verbal features of interaction could be evaluated automatically. The third paper (Toivanen et al.) also concerns L2 Finnish but changes the focus to writing and investigates the suitability of deep learning models for predicting the human-rated CEFR levels of L2 Finnish texts by using several corpora of L2 writing. The fourth article (Hou et al.), reports on Revita, an online tutoring system that was originally designed to automatically create linguistic exercises and feedback from texts in several small, endangered languages with the goal of supporting their revitalisation. Over time, the scope of Revita has increased, and it now includes Finnish, for example, and has also been used for research purposes in other projects. Finally, Leontjev et al.'s paper explores the potential of generative AI to become a part of the mediation process in computerised dynamic assessment of L2 English reading comprehension delivered on the Revita platform. The authors study how an AI-based chatbot that was enabled after the learners arrived at correct responses, with or without standardised mediation, yielded deeper diagnostic insights into learner abilities, becoming thus complementary to the standardised mediation as a part of computerised dynamic assessment.

The studies reported in the five papers differ in terms of their completeness, which had some impact on their order in the special issue. For example, the first article on DigiTala reports on a completed project and the ongoing follow-up project stemming from it, AASIS, comes second, after its parent project. The order of the fourth and fifth articles in this issue has a somewhat similar reason: Since the DD-LANG project uses the Revita platform that was created several years prior to the start of the project, it is

logical to describe Revita before moving on to one of the research-based further developments of the platform in DD-LANG.

Furthermore, there is some transition from assessment focused articles to ones discussing assessments that are integrated into teaching and learning. Thus, the specific applications of AI vary in the papers, roughly falling into one of the two categories: AI in the service of automated language assessment (ASA) and AI supporting the development of L2 learners' abilities. In Hildén et al., Ullakonoja et al., and Tarvainen et al, AI is used (1) to automate assessments; (2) to assess various features of monologic speech, extending the assessment of speech to spoken interaction; and (3) in automating writing proficiency assessment, respectively. In Hou et al. and Leontjev et al., while automatic assessment of texts is a part of the procedure required to create exercises and provide graduated guidance to learners during the assessment process and analysing texts to respond to learners' queries, respectively, there is a greater focus on supporting learner development by providing personalised instruction and guidance to learners. This is not to say that the first three papers do not focus on supporting learning. Indeed, in all the papers of this special issue, there is an overarching goal of supporting language learning, teaching, and assessment through technological innovation.

While these purposes of assessment have been discussed as binaries, oftentimes focusing on assessment of learning (AoL) and assessment for learning (AfL) (see, e.g., Davison & Leung, 2009, for a discussion), it is more fruitful to discuss these as a dimension, where there is a larger focus on one or the other while both purposes are taken into account (see, e.g., Wiliam & Leahy, 2015, for a discussion). Indeed, while the project in Ullakonoja et al., for example, is heavily informed by research and areas for development of high-stakes assessments in Finland – with the challenges in incorporating automatic-speech-recognition-based assessment in high-stakes testing being among its key considerations – its findings are also discussed as valuable for developing language learning tutors with whom language learners can interact . In Hildén et al., the eventual goal is to potentially use the developed procedures to increase the efficiency and reliability of high-stakes decisions, including in the Finnish Matriculation Examination. Currently however, the formative function is emphasised. Tarvainen et al., in turn, explicitly state that both AoL and AfL informed the development of their project. They further discuss the relevance of their findings and

development work for both formative feedback in support of writing development and high-stakes proficiency purposes, such as language requirements for citizenship. Hou et al. and Leontjev et al., on the other hand, concentrate strongly and explicitly on assessment in support of learner development. Hou et al., throughout their paper, focus on using assessment for adapting language training to the needs and growing abilities of the individual learner, by means of graduated feedback and guiding learners to correct responses. The authors also see the potential for a complementary role to support teaching by continuously analysing and synthesising learning data to enrich teaching decisions. Leontjev et al., focusing on the framework of L2 dynamic assessment, explore how open-ended learner interactions with an AI-based chatbot complement more standardised mediation whose goal is to develop learners' thinking processes about reading in English. They also discuss how this approach to simultaneously assessing and guiding learner development can be used in the language classroom, including providing diagnostic insights into learners' emerging abilities and areas of struggle.

Finally, we should note the range of languages and language subskills targeted collectively by the papers in this special issue. Hildén et al., Ullakonoja et al., and Tarvainen et al., focus on the second language context (Finnish and Swedish). Leontjev et al., in turn, discuss supporting the development of reading in English – a global language. Hou et al., however, support a range of L2s, including both Finnish and Swedish as second languages, English as a global language, and a number of endangered and minority languages such as Komi-Zyrian, North Saami, and Sakha. With regards to the subskills, the papers collectively target L2 speaking (Hildén et al.; Ullakonoja et al.), writing (Tarvainen et al.), reading (Leontjev et al.), and grammar and vocabulary, but also aural comprehension, in the shape of asking learners to type in missing grammatical forms, and pronunciation (Hue et al.).

Despite the focus in this special issue on one educational context – Finland – the papers discuss an extensive range of questions regarding the role of AI, assessment purposes, and targeted languages and subskills. They offer several directions for the use of technology in assessment supporting the learning of languages, and research on

this topic both in Finland[2] and internationally. We, therefore, hope that it can offer useful insights to a variety of readers of the journal.

ORCID iDs

Ari Huhta https://orcid.org/0000-0003-3124-449X

Dmitri Leontjev https://orcid.org/0000-0003-0177-3681

References

FNAE (2016). New national core curriculum for basic education: focus on school culture and integrative approach. Finnish National Agency for Education. https://www.oph.fi/en/statistics-and-publications/publications/new-national-core-curriculum-basic-education-focus-school

Davison, C., & Leung, C. (2009). Current issues in English language teacher-based assessment. *TESOL Quarterly*, *43*(3), 393–415. https://doi.org/10.1002/j.1545-7249.2009.tb00242.x

Luukka, M.-R., Pöyhönen, S., Huhta, A., Taalas, P., Tarnanen, M. & Keränen, A. (2008). *Maailma muuttuu - mitä tekee koulu? Äidinkielen ja vieraiden kielten tekstikäytänteet koulussa ja vapaa-ajalla* [The world changes - how does the school respond? Mother tongue and foreign language literacy practices at school and in free-time.] University of Jyväskylä: Centre for Applied Language Studies. https://jyx.jyu.fi/jyx/Record/jyx_123456789_36607

Nguyen, P. (2021). Uses of language assessments. In C. Chapelle (Ed.), *The encyclopedia of applied linguistics*. Wiley-Blackwell. https://doi.org/10.1002/9781405198431.wbeal1237.pub2

Wiliam, D., & Leahy, S. (2015). *Embedding formative assessment: Practical techniques for K-12 classrooms*. Learning Sciences International.

[2] Indeed, two of the projects reported in this special issue – DD-LANG (Leontjev et al.) and Revita (Hue et al.) – will play a key role in the newly started six-year research project DALAI-FIN, funded by the Strategic Research Council of the Research Council of Finland, that aims to support adult migrants' L2 Finnish learning with the help of AI.