

AI Developments in Language Education: What now?

By Cassandra Robertson – School of Languages, Sabanci University

Recently I attended the annual Equals Online conference, held Oct 11-12, which brought together 30 speakers from at least 7 different countries (half the lineup from Turkey) and ELT professionals from across the network. The program featured an illuminating two-part plenary from Dr. Hakan Tarhan of the TOBB University of Economics and Technology on the responsible and effective integration of AI into language education and 24 other sessions on a diverse array of topics, facilitating a dynamic exchange of insights, experiences, and best practices.



Having also participated in the Equals conference last April, it was interesting to observe the evolution of the discourse regarding the ‘AI issue’ now that these tools have been around for two years. AI-focused sessions this time offered a more in-depth look at all aspects of its integration into language education, exploring its limitations in terms of the mechanics of how LLMs process input (through ‘tokenization’) and emphasizing both the growing need for AI literacy, as well as, crucially, the primacy of the uniquely human capacities no machine can ever replicate. Most significantly, the event revealed an emerging clarity for navigating the ambiguities of AI integration in instruction and assessment, marking a definite shift toward a more practical set of principles and possibilities for an effective AI-empowered pedagogy.

The plenary opened with a recounting of the reaction in education to AI since November 2022, acknowledging valid uncertainties and concerns but suggesting a gradual evolution toward ‘cautious optimism’ and growing recognition of AI proficiency as an essential skill for the modern workforce. Addressing challenges to academic integrity, Dr. Tarhan pointed out the inefficacy of bans and overly punitive responses to AI use, and delivered an insightful examination of the issues with detection tools, referencing Perkins et al. (2024b) which compared the accuracy of various detectors and found not only an abysmally low rate for Turnitin (8%) but considerable inconsistencies according to the particular AI tool and version (e.g., Claude 2, GPT-4) used, with paid models more impervious to detection than free ones (p. 15-17). This, he explained, together with the risk for false positives, genuine prospect that development of such detectors will lag behind that of AI technology, and mounting challenge to the notion that AI use has no place in student-written work, ultimately makes detection tools an unreliable determiner of actual academic integrity in the age of AI.

By way of a solution, Dr. Tarhan advocated for robust institutional policies that accommodate a more nuanced understanding of the issue, that create space for the development of AI competencies while deterring technologically-assisted forms of plagiarism by providing explicit guidance as to what constitutes ‘appropriate use’. He remarked that, in his experience (which resonated with mine), the vast majority of students do not want to cheat; but they sometimes lack clear instruction on how to avoid it, especially in this new AI-saturated reality, and often the know-how of using these tools in a constructive way. By defining and modeling appropriate use in a given context, Dr. Tarhan argued, by involving students in the conversation and integrating meaningful interaction with these tools into the learning process, it is possible to create a culture of trust and level of competency that, to a significant extent, eases the likelihood of misuse.



Dr. Tarhan outlined his five-stage approach to the development of such frameworks, with the first three involving a deliberate awareness-building of: *the situation*, by investigating what different stakeholders know about these technologies as well as which tools are being used and how, and by staying informed regarding relevant AI use policy at all levels (national, regional, institutional); *what is needed*, in consideration of the potential benefits, what guidelines are called for, and how AI use can be monitored and assessed; and how the answers to these inquiries can in turn contribute to the broader strategic vision of the institution. Next comes the design and development of frameworks that are human-centered, conducive to building key AI competencies like prompting and output analysis, and adaptable to the specificities of the context as well as able to evolve as the technology does. The final stage concerns implementation which, Dr. Tarhan explained, can be achieved by creating an infrastructure of

support, collaborative culture, and systems for continuous evaluation and improvement. An example that comes to mind is the effort of many organizations to build prompt banks such as [one by AI for Education](#) shared by a conference participant, which may be of interest to language instructors.

To illustrate his points, Dr. Tarhan cited two practical models for AI in assessment design applicable at varying levels of implementation. The first, from an influential paper co-authored and popularized by Leon Furze, was the Artificial Intelligence Assessment Scale (AIAS), which defines appropriate AI use along a spectrum for different

Scale Levels and Descriptions		
1	NO AI	The assessment is completed entirely without AI assistance. This level ensures that students rely solely on their knowledge, understanding, and skills. AI must not be used at any point during the assessment.
2	AI-ASSISTED IDEA GENERATION AND STRUCTURING	AI can be used in the assessment for brainstorming, creating structures, and generating ideas for improving work. No AI content is allowed in the final submission.
3	AI-ASSISTED EDITING	AI can be used to make improvements to the clarity or quality of student created work to improve the final output, but no new content can be created using AI. AI can be used, but your original work with no AI content must be provided in an appendix.
4	AI TASK COMPLETION, HUMAN EVALUATION	AI is used to complete certain elements of the task, with students providing discussion or commentary on the AI-generated content. This level requires critical engagement with AI generated content and evaluating its output. You will use AI to complete specified tasks in your assessment. Any AI created content must be cited.
5	FULL AI	AI should be used as a "co-pilot" in order to meet the requirements of the assessment, allowing for a collaborative approach with AI and enhancing creativity. You may use AI throughout your assessment to support your own work and do not have to specify which content is AI generated.

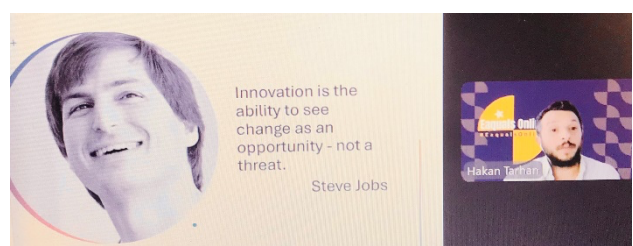
Table 1. The AI Assessment Scale

assessments, establishing clear guidelines on the extent of AI use permissible and, where relevant, expectations regarding the citation of such tools (Perkins et al., 2024a, p. 7). This framework, a variation of which (restricted to the first two bands) I have also integrated into my own research writing lessons with encouraging results, helps communicate clear parameters on ethical AI use and is adaptable across various contexts. The second model presented was the University of Sydney’s two lane approach, designed to ensure the fulfillment of course learning objectives “in a world where generative AI is ubiquitous”, through purposeful placement of in-person summative ‘lane 1’ assessments (with AI restrictions and monitoring) at the *program* level, while permitting a scaffolded, formative use of AI at the *unit* level (lane 2) for tasks like idea generation, structure suggestions, or clarity improvement (Bridgeman, Liu, & Weeks, 2024).

A major concern when it comes to AI integration in education is the potential for such tools to interfere with the development of critical thinking skills. This was a primary focus in the second plenary, centered on maintaining a pedagogy-first approach, and insightfully elaborated on in the session by Christopher Sanders, Associate Director for Assessments at MEF University’s School of Foreign Languages, who in speaking to the matter of responsible use similarly addressed the risk for AI tools to result in ‘skill decay’, referencing a study by The Wharton School which compared the effects of different GPT tools on learning and found not only a drop in performance when AI assistance was removed (below that of the control group), but an inflated degree of self-confidence nonetheless (Bastani et al., 2024, p. 9-10), what he referred to as the ‘Dunning-Krueger Effect’. In consideration of this legitimate risk when it comes to AI integration in education, both Sanders and Dr. Tarhan called for a principled use of such tools that supports but *never supplants* the natural cognitive processes involved in learning.

Toward that end, Sanders shared a concise list of priorities for ‘better usage’, which reiterated the need for clear institutional AI policies as well as a transparent, human-centered approach to integration; urged educators to stay abreast of the latest research and involve students in exploring the educational potential of AI; and underscored the value of quality prompting. On this last point, Sanders also mentioned some key elements of an effective prompt, which from an instructor standpoint should include the AI’s role, context, lesson aims, and pedagogical framework. Conference participants added to this by suggesting inclusion of a link to the CEFR to more clearly define student language proficiency levels (an area in which AI tools notoriously fall short). For his part, Dr. Tarhan challenged the perception that AI use by default inhibits the development of critical thinking by sharing a generous sampling of activities skillfully incorporating use of such tools for different levels and learning objectives, demonstrating how AI can be used to genuinely enhance the “intellectual involvement” in learning and clarifying that ‘the threat to critical thinking development is not AI but the *misuse* of AI’.

The arrival of AI in education has been understandably overwhelming, the process of coming to terms with the magnitude of this paradigm shift a veritable journey from, as Dr. Tarhan described, an excited ‘Oh, boy!’ to an exasperated ‘Come on!’ and everything in-between. Most notably, the presumption that any AI use whatsoever is a violation of academic integrity has given way to a more practical discussion of ‘What now?’, as educators gain familiarity with these tools, engage students in cultivating a culture of integrity with transparent parameters for appropriate use, and develop integration practices that effectively uphold critical thinking while fostering advancement of both AI literacy and core competencies. It’s worth being reminded that AI as a new, albeit exceptionally powerful, ‘mediating artifact’ (Vygotsky,



1978, p. 52-57) still constitutes just one part of a larger, dynamic process of learning in which pedagogical expertise remains critical. A principled approach to responsible AI use ensures student growth is supported in ways that keep pace with technological and market changes while still being grounded in the intellectual capacities that make us uniquely human.

References:

AI for Education. (2023). *Gen AI Chatbot Prompt Library for Educators*.

<https://www.aiforeducation.io/prompt-library>

Bastani, H., Bastani, O., Sungu, A., Ge, H., Kabakçı, Ö., & Mariman, R. (2024, July 15). Generative AI can harm learning. The Wharton School. <http://dx.doi.org/10.2139/ssrn.4895486>

Bridgeman, A., Liu, D., & Weeks, R. (2024, September 12). Program level Assessment design and the two-lane approach. The University of Sydney. <https://educational-innovation.sydney.edu.au/teaching@sydney/program-level-assessment-two-lane/>

Perkins, M., Furze, L., Roe, J., & MacVaugh, J. (2024a). The Artificial Intelligence Assessment Scale (AIAS): A framework for ethical integration of generative AI in educational assessment. *Journal of University Teaching and Learning Practice*, 21(06). <https://doi.org/10.53761/q3azde36>

Perkins, M., Roe, J., Vu, B. H., Postma, D., Hickerson, D., McGaughran, J., & Khuat, H. Q. (2024b). Simple techniques to bypass GenAI text detectors: Implications for inclusive education. *International Journal of Educational Technology in Higher Education* 21(53). <https://doi.org/10.1186/s41239-024-00487-w>

Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Cole, M., John-Steiner, V., Scribner, S., & Souberman, E. (Eds.) Harvard University Press. <https://www.jstor.org/stable/j.ctvjf9vz4>