SAJILE

St Andrews Journal of International and Language Education, 2(1), 3144
https://www.doi.org/10.15664/r1nrqf67

Humanising Generative AI Use in TESOL Assessment

Author Name and Affiliation

Mattia Zingaretti, Cambridge University Press & Assessment, Cambridge, England.
https://orcid.org/0000-0002-5565-2538

Corresponding Author's Email

mattia.zingaretti@cambridge.org

Abstract

Institutional responses to Generative AI (GenAI) in higher education, from outright prohibition to open permission, leave students without the necessary conditions to engage with GenAI critically. In language teacher education, where assessment should support students in consolidating a professional identity, the lack of a principled approach to GenAI is problematic. This article proposes the redesign of two linked assessments in a final-year undergraduate module on Teaching English to Speakers of Other Languages (TESOL), arguing that humanising GenAI use in TESOL assessment offers a way forward that places learner agency, well-being and professional identity at the centre of assessment design. Redesigned briefs are made openly available as supplementary materials, inviting implementation, evaluation and adaptation across contexts.

Keywords

1. Humanising technology
2. Generative AI
3. TESOL assessment
4. Language teacher education
5. Higher education

AI Declaration

No artificial intelligence elements were used in the preparation of this article.

Conflict of Interest

The author declares no conflict of interest.

Acknowledgements

I am very grateful to Hayo Reinders and Chris J. Hall for their insightful comments at various stages of this work.

Humanising Generative AI Use in TESOL Assessment

Introduction

With OpenAI's release of ChatGPT in November 2022, a new class of Artificial Intelligence (AI) tools capable of generating fluent text in response to user prompts has entered the everyday academic lives of university students worldwide. In the UK, the proportion of undergraduates who reported using such Generative AI (GenAI) tools for assessments rose from 53% to 88% in a single year (Freeman, 2025). Institutional responses have spanned a wide spectrum, from outright prohibition to open permission, with many universities occupying a middle ground of hedged guidance, at least as of 2024 (Atkinson-Toal & Guo, 2024). This article argues that neither extreme is pedagogically adequate: prohibition does not prevent covert use, offering no opportunity to develop the critical GenAI literacy students will need as professionals; permission without scaffolding creates conditions in which unreflective over-reliance is the path of least resistance. Existing work has begun to map this territory. For instance, the AI Assessment Scale (Perkins et al., 2024) gives educators a shared vocabulary for specifying permitted levels of GenAI use in assessment, and empirical studies have started to examine what engagement with GenAI develops in students, reconceptualising the technology as a co-learner rather than a mere tool (Fang & Zhou, 2025). What such work does not yet provide, however, is an account grounded in learners' psychological needs and professional identity formation: one that asks not only how much GenAI use to permit at each stage, but what conditions make that use conducive to agency, well-being and the consolidation of a professional self, particularly in language teacher education, where professional identity is part of the curriculum's very object.

Thus, a third position is advanced, aligning with “humanising technology” (Reinders et al., 2024): a values-first approach to technology in education that places learner agency, well-being and professional identity at the centre of design decisions. Drawing on a concrete case from a final-year undergraduate module on Teaching English to Speakers of Other Languages (TESOL) at a UK university, this article describes the redesign of two linked assessments which draws on the Motivation, Engagement and Thriving in User Experience model (Peters et al., 2018) grounded in Self-Determination Theory (Ryan & Deci, 2000), and hybrid autonomy (Reinders, 2026), to propose a framework in which GenAI engagement is made visible, across all stages of the writing process, and educationally productive. The redesign is offered as a principled proposal grounded in observed practice and recent literature, intended for critical adaptation rather than direct adoption.

Context and Impetus

Final-year undergraduate assessment in language teacher education carries significant weight since, at this stage, students are not only demonstrating disciplinary knowledge but consolidating a professional identity: learning to think, reflect and articulate their practice as teachers. While the expectation of independent and critical engagement is common to final-year assessment across disciplines, it carries additional significance in language teacher education, where professional identity is not incidental to the curriculum but part of its object. Language teacher identity research conceptualises identity as an ongoing

St Andrews Journal of International and Language Education, 2(1), 3144
https://www.doi.org/10.15664/r1nrqf67

negotiation of who one is, and who one aspires to become, as a practitioner, shaped substantially by the activities and artefacts of teacher education programmes, with module-level learning shown to contribute to identity formation beyond knowledge acquisition (Banegas, 2023; Yazan, 2018). Assessment tasks at this level can therefore function as developmental environments, creating conditions in which students engage with the substantive intellectual demands of the field independently and critically. This developmental function is not automatic: summative assessment exists first to certify achievement, and whether it also develops anything depends on what its design asks students to attend to, rehearse and come to value. The position taken here is that in a discipline whose object includes professional identity, assessment design inevitably participates in identity formation, whether by intention or by default. The formative-to-summative bridge described below reflects this position. The growing presence of GenAI tools complicates this function in ways that neither blanket prohibition nor lightly governed permission adequately addresses.

The module described in this article, *Lesson and Materials Design for TESOL*, was designed by the author and refined across three annual iterations from 2023–24 to 2025–26, a period coinciding with the rapid expansion of GenAI use among UK undergraduates (Freeman, 2025). In its third year of delivery (2025–2026), it was assessed through two linked summative components. The written component (2,000 words, submitted via Turnitin) required students to produce either a lesson plan with rationale or an evaluation of language teaching materials. The oral component (a 10-minute live interview conducted via Microsoft Teams) required students to respond to questions on lesson planning, materials evaluation and diversity in English language teaching.

The approaches to GenAI use in the two briefs reflected the author's own evolving thinking at the time each brief was written, and the tensions between them were the starting point for the redesign described here. The brief for the written component warned:

> *GenAI tools like ChatGPT/Claude/Gemini etc. will not be able to create a critical lesson plan with a well thought-out rationale, or critically evaluate existing materials. GenAI use is, therefore, not allowed, and will result in poor work and academic misconduct.*

The brief for the oral component stated:

> *GenAI tools like ChatGPT/Claude/Gemini etc. can be useful in helping you structure your answers when preparing, for example, by following the Situation, Task, Action, Result (STAR) or Context, Action, Result, Learning (CARL) frameworks. However, the concepts need to be studied and thought of critically, which GenAI cannot*

accomplish. Therefore, limit GenAI use to preparation (if absolutely needed), and ensure that your ideas are genuinely your own.

In retrospect, neither position was grounded in a coherent account of what GenAI engagement should develop in students. In the written component, outright prohibition may not have prevented covert use: a UK study estimated that 22% of students used GenAI to cheat in assessed work in 2023–24, with many uncertain about what constituted legitimate use (Newton, 2025). While a submission coversheet invited students to disclose GenAI engagement, the brief's outright prohibition meant that any such disclosure was tantamount to admitting misconduct, which was more likely to deter honesty than encourage critical reflection on GenAI use. Students who avoided GenAI on ethical grounds, meanwhile, were not only unrecognised for that choice, but also denied the opportunity to develop the critical GenAI literacy that scaffolded engagement with such tools could have built. In the oral component, framing GenAI as a last resort combined with the absence of critical scaffolding may have produced one of two patterns: avoidance by students uncertain about the policy, or unreflective reliance by students who generated polished responses via GenAI, rehearsing them without interrogating the output. In both cases, Cukurova (2025) argues, the risk is not technology itself but the conditions under which it is used: when GenAI removes effortful processing rather than scaffolding it, skill development may stall. The redesign below addresses this directly as it makes transparent engagement assessable and, as a result, concealment strategically counterproductive.

Together the two assessments missed the opportunity to provide students with a framework for thinking critically about when and how to use GenAI, with no structured opportunity to make that engagement visible and no scaffolding for the reflective and self-regulated practice that professional formation requires. Nor did either assessment ask students to apply their developing disciplinary knowledge to the evaluation of GenAI output itself, missing an opportunity to align GenAI engagement with the specific professional competencies the module was designed to develop. The redesign proposed here responds to these gaps.

Literature Review

Before describing the redesign, this section situates it within the relevant literature and establishes the conceptual framework that informs it.

GenAI in Language Education

Research on GenAI in language education reveals a complex picture of affordances and risks. A recently edited volume by Zapata (2025) brings together empirical contributions from across international contexts, demonstrating how GenAI tools can scaffold creativity, support critical engagement with AI-produced text and prompt reflection on culturally responsive pedagogy. Particularly for students writing in an additional language, studies have shown that GenAI tools can support interactive practice, provide personalised feedback and offer accessible linguistic support (Kasneci et al., 2023; Zadorozhnyy & Lai, 2024). In the context of L2 academic writing, Kim et al. (2025) found that students using a ChatGPT-embedded

St Andrews Journal of International and Language Education, 2(1), 3144
https://www.doi.org/10.15664/r1nrqf67

writing system reported GenAI as beneficial across different stages of the writing process (ideation, planning, drafting and revision).

There are also some increasingly well-evidenced risks, however. Zadorozhnyy and Lai (2024) raise concerns that GenAI writing assistance, when adopted uncritically, may erode, rather than develop, independent writing skills. Johnson et al. (2022) and Vaccino-Salvadore (2023) draw attention to cultural bias and the reproduction of dominant linguistic norms in Large Language Model (LLM) outputs, which carry significant impact on language teacher education where students are developing a professional voice. Several contributors to Zapata (2025), too, voice similar reservations, including concerns about academic misconduct and the potential for culturally biased outputs to shape learners' developing professional identities in ways they may not notice. In an experimental study of essay writing, Kosmyna et al. (2025) found that participants who used ChatGPT showed significantly weaker neural connectivity, lower reported ownership of their work and greater difficulty recalling what they had written compared to those who worked without GenAI: evidence of what they term “cognitive debt” accumulating over repeated LLM use. Dergaa et al. (2024) offer a complementary perspective, arguing that unreflective reliance on AI chatbots poses measurable risks to cognitive health, including the atrophy of memory and critical reasoning capacities. Addressing these risks in a principled way requires a framework that places learner agency and well-being at the centre of technology design decisions.

Conceptual Framework

The framework employed in the present redesign draws on three interconnected ideas. The first is “humanising technology” (Reinders et al., 2024) in the context of language learning and teaching. Specifically, Reinders et al. (2024, p. 20) argue that technology should be used to promote human flourishing rather than merely to manage efficiency, defining digital well-being practices as “the use of technology to improve human flourishing and avoid harm from technology use”. The humanising technology approach shifts the evaluative question from whether a specific technology works to whether it serves the learner as a whole person.

The second is the Motivation, Engagement and Thriving in User Experience (METUX) model, developed by Peters et al. (2018). METUX analyses how technology affects users not only at the level of individual tasks but across multiple spheres, from task engagement through to behaviour, life and society. Crucially, drawing from Self-Determination Theory (SDT) (Ryan & Deci, 2000), it identifies three basic psychological needs whose satisfaction is essential to well-being and sustained motivation: autonomy, competence and relatedness. Applied to GenAI use in assessment, the model helps explain why unscaffolded GenAI engagement may satisfy surface-level needs, such as the immediate relief of producing a polished output, while frustrating deeper needs, including the development of genuine competence and a sense of ownership over one's work. Equally, it suggests that GenAI use structured to support effortful processing (e.g., through scaffolded use at different stages) may enhance, rather than undermine, autonomy and competence.

The third is the conceptualisation of “hybrid autonomy” (Reinders, 2026), which draws in part on research on hybrid intelligence (e.g., Cukurova, 2025) to describe the conditions under which human-AI co-regulation is possible. Reinders (2026) argues that genuine hybrid autonomy requires more than access to AI tools: namely, structured

conditions in which learners can interrogate GenAI output, calibrate their own thinking against it, and develop the metacognitive awareness to decide when and how GenAI is useful. Without such conditions, what looks like autonomous AI-assisted work may in practice represent a transfer of cognitive agency to the tool.

The Innovation: A Proposed Redesign

The redesign proposed here addresses the fundamental problem that both assessments rewarded outputs that GenAI tools produce with ease: a polished written argument and a fluently structured interview response. If that remained the case, neither prohibition nor cautionary guidance would shift student behaviour in any meaningful way. The redesign therefore changes what the assessments make visible and reward, making the process of GenAI engagement an object of assessment in its own right. It operates across five interlocking dimensions, each addressed by a specific design element described below. The redesigned assessment briefs are available as open supplementary materials on the Open Science Framework (OSF) platform at https://osf.io/tc5ep.

Making GenAI Engagement Transparent and Accountable

Both assessments now require students to respond to the same core reflection prompts: one asking whether and, if so, how GenAI was used and what was kept, modified or discarded, and a tailored follow-up addressing the specific demands of each assessment. In the written component, students are asked whether GenAI suggested an argument, example or phrasing that they decided not to use and why; in the oral component, whether GenAI suggested a framing that did not feel like their own professional voice and what they changed. Using a shared core prompt across both assessments signals a consistent stance: the decision to use GenAI or not rests with the student, and it is the quality of their critical engagement with that decision, whichever it may be, that matters. This is a standard the process portfolio is designed to support equally for both pathways. Students who used GenAI will typically accumulate more discrete decision points to report than those who did not; the marking descriptors, however, reward the quality of critical reflection rather than the volume of material reflected upon, and the records non-users compile (structural options considered, frameworks adopted or set aside, changes between drafts) offer equivalent reflective substance.

The redesign addresses the previous limits of transparency through self-report alone in two ways. First, in the oral interview, the assessor may select a passage from the student's written submission and ask them to talk through how it developed, testing ownership in real time. Second, students are encouraged to retain prompt logs or screen recordings alongside their submissions, providing an evidence base where self-reporting alone might be unreliable.

Scaffolding Cognitive Engagement at Different Stages

The redesign adopts a staged integration model in which GenAI is permitted at every phase of the writing process but with distinct constraints at each. In the ideation stage, students may use GenAI for brainstorming, exploring structural options or clarifying initial ideas, but must document what they generated independently and what emerged from GenAI

St Andrews Journal of International and Language Education, 2(1), 3144
https://www.doi.org/10.15664/r1nrqf67

interaction. In the drafting stage, GenAI is restricted to diagnostic use: identifying gaps, questioning the clarity of arguments or highlighting underdeveloped areas, without producing or rewriting content. In the revision stage, students may consult GenAI more freely, but they must annotate specific changes made in response to GenAI suggestions and justify their decisions.

Unrestricted GenAI use risks eroding the cognitive engagement that both learning and professional formation require (Kosmyna et al., 2025). The staged model responds to this by ensuring that independent thinking precedes GenAI consultation at each phase, while acknowledging that the relationship between the writer and the tool is not a binary before/after but an evolving negotiation across the production of a text (Kim et al., 2025).

In the oral component, the same principle applies. Students must document where their own thinking began and where GenAI entered the process, prepare at least one full STAR or CARL response without GenAI as a baseline, and use GenAI diagnostically for subsequent responses rather than to generate polished alternatives.

Together, the documented traces constitute a process portfolio submitted alongside the written component and drawn on during the oral interview. Students who choose not to use GenAI submit a portfolio documenting their independent decision-making process (how they arrived at their chosen structure, which frameworks they considered and what changed between drafts), ensuring that the documentation requirement applies equally regardless of whether GenAI was used.

Shifting GenAI from Output to Thinking Partner with Disciplinary Engagement

Students are guided to prompt GenAI diagnostically to explain why an argument is unclear, what is missing from their analysis or what a stronger example might look like, without providing that example directly. This shift is grounded in SDT (Ryan & Deci, 2000) as applied to technology design (Peters et al., 2018): feedback that explains its reasoning supports competence development in a way that silent correction does not. The student remains the author; GenAI functions as a critical reader whose suggestions can be interrogated, accepted or rejected, preserving the sense of autonomy that SDT identifies as central to sustained motivation.

Beyond this process role, GenAI also presents an opportunity to deepen disciplinary engagement. Students might generate GenAI content relevant to their chosen option (e.g., a lesson plan segment, a set of corrective feedback examples, a materials adaptation) and evaluate it against disciplinary knowledge in TESOL. In the oral interview, one question may present students with a GenAI-generated response to a TESOL scenario and ask them to evaluate it on the spot. This ensures the GenAI component is aligned with the professional competencies the module is designed to develop; students with limited prior GenAI experience are supported through the seminar component described below. This disciplinary evaluation task applies equally to students who choose not to use GenAI in their own writing or preparation: appraising AI-generated content against TESOL knowledge is to be considered a professional competence in its own right, distinct from the decision to employ GenAI as a support tool, and it requires no personal GenAI workflow, since sample outputs can be generated in the seminars or, in the oral component, presented by the assessor.

Restoring the Relational Dimension of Learning

Three of the module's twelve seminars are proposed for AI-integrated work. In each session, students first produce an independent response, then write a brief GenAI intention note (e.g., "What do I want to achieve with this tool, and which parts of this response will I keep control of?") before using GenAI. A short, paired comparison follows, in which students discuss what changed, what they kept and which version feels more authentically theirs. This peer comparison reintroduces relatedness into a process that GenAI use tends to privatise, addressing what Ryan and Deci (2000) identify as one of the three basic psychological needs whose frustration undermines sustained motivation, operationalised by Peters et al. (2018) as a design target in the METUX model. The artefacts produced (i.e., the intention note, the before/after comparison and a brief record of the paired discussion) may be included in the process portfolio, creating a formative-to-summative bridge.

Making Process Consequential in Assessment

If the final product continues to carry most weight in assessment, students will opt for output quality and treat process documentation as a compliance exercise. The process portfolio addresses this by making documented critical engagement with the writing process, whether GenAI-assisted or independently produced, a formal and assessable dimension of the written submission. Because all students submit a process portfolio regardless of whether they used GenAI, the documentation requirement does not create an asymmetry in workload that might discourage honest declaration of GenAI use. Students who used GenAI document their decision points at each stage; students who did not use it document how they arrived at their chosen structure, which frameworks they considered, and what changed between drafts. In both cases, the marking descriptors for “professional learning skills” specify what critical engagement looks like at each grade level, with evidence drawn from the portfolio. “Professional learning skills” is one of five equally weighted criteria and therefore carries 20% of the component mark; in addition, the “thinking skills” descriptors at 2:1 level and above reward the capacity to interrogate and selectively use GenAI output, so that documented critical engagement with the process bears on two of the five criteria rather than one. In the oral interview, the assessor's process-related question may draw on a specific passage from the written submission, making the process dimension consequential rather than merely declarative.

Discussion

The redesign described here changes what both pieces of assessment make visible and value: documenting independent thinking at each stage, accounting for GenAI decision points across ideation, drafting and revision, demonstrating critical distance from GenAI output, and applying disciplinary knowledge to GenAI-generated content are now part of what it means to do well.

This shift is consistent with the argument that hybrid autonomy requires structured conditions in which learners interrogate GenAI output, calibrate their own thinking against it, and develop the metacognitive awareness to decide when GenAI is useful (Reinders, 2026). The redesign builds precisely those conditions into the assessment, as well as responding to the Reinders et al. (2024) humanising technology imperative: the question is

St Andrews Journal of International and Language Education, 2(1), 3144
https://www.doi.org/10.15664/r1nrqf67

not whether AI works, but whether its use serves the learner as a whole person, including their developing professional identity. An alternative design would mandate GenAI use to ensure a level playing field; the present design opts against this in favour of student choice, grounded in the autonomy rationale of SDT (Ryan & Deci, 2000).

The METUX model (Peters et al., 2018) offers a useful lens for evaluation. Applied to the written component, it asks whether the redesign supports learners not only at the task level (structuring an argument) but at the behaviour level (developing the habits of critical self-regulation that professional teaching requires) and at the life level (building a sense of professional identity that extends beyond the module). The staged integration constraints, the process portfolio and the diagnostic prompting elements are designed to operate at all three levels. The redesign also positions GenAI literacy as a dimension of TESOL professional competence rather than a generic academic skill. By asking students to evaluate GenAI output against TESOL knowledge, it embeds critical engagement with AI within the epistemic practices of the discipline.

Several limitations merit acknowledgement. The staged integration model introduces a complex set of expectations, and students may need explicit modelling in seminars before they can navigate these expectations productively. A brief student-facing summary of the staged expectations, distributed alongside the full brief, might be a straightforward mitigation for adopters to consider, complementing the explicit modelling provided in the seminars. The reflective and portfolio elements add to every student's overall workload and, although the addition is symmetrical across GenAI users and non-users, they may feel like compliance mechanisms if not carefully framed as professional development scaffolds. Furthermore, the process portfolio is itself a text artefact that GenAI could produce; the oral interview partially addresses this by testing ownership in real time, with the two assessments functioning as a system of mutual verification. The redesign also assumes a degree of GenAI literacy that not all students will have, and while the proposed seminars model the process explicitly, students with limited prior experience may find diagnostic GenAI prompting difficult to enact without additional support. Importantly, although the redesign has not been implemented, it is a principled response to observed patterns across three years of module delivery and engagement with recent GenAI literature in language education. It was developed after the author left the institution, and, at the time of writing, no successor had yet been appointed to the module who might take it forward. This makes publication all the more necessary: by making the redesign openly available alongside the revised assessment briefs via OSF, this article invites other TESOL practitioners to implement, critically evaluate and adapt it in their own contexts.

Conclusion

This article has described a redesign of TESOL assessment in response to the challenges posed by uncritical GenAI use, starting from the premise that when assessment rewards what GenAI already does well, neither prohibition nor unstructured permission adequately serves learners: the former may drive GenAI use underground, while the latter risks leaving students without the support needed to develop the professional competencies that assessment is intended to foster.

By making GenAI engagement visible, staged, disciplinarily grounded and part of what it means to do well, the redesign proposed here creates the structured conditions on

which, Reinders (2026) argues, hybrid autonomy depends. By attending to basic psychological need satisfaction across task, behaviour and life spheres, the redesign aims to support not just risk avoidance but the kind of flourishing that Reinders et al. (2024) and Peters et al. (2018) place at the centre of humanised technology design.

Although developed for TESOL, the underlying principle generalises. Recognising GenAI in assessment, rather than banning or ignoring it, is the tenable way forward wherever assessment currently rewards outputs that GenAI produces with ease. The architecture proposed here (staged integration, process portfolios, diagnostic prompting, disciplinary evaluation of GenAI output and reflection prompts that apply equally to users and non-users) can arguably be adapted to English language teaching more broadly and to higher education assessment at large. What is discipline-specific is the knowledge against which GenAI output is evaluated; the design principle of visible, staged and consequential engagement is not.

The revised assessment briefs are offered as a transferable resource for TESOL practitioners seeking to move towards humanised and principled approaches to GenAI in assessment, demonstrating that the integration of GenAI can be designed to serve the specific disciplinary and professional purposes that assessment in language teacher education exists to support.

St Andrews Journal of International and Language Education, 2(1), 3144
https://www.doi.org/10.15664/r1nrqf67

References

Atkinson-Toal, A., & Guo, C. (2024). Generative Artificial Intelligence (AI) education policies of UK universities. *Enhancing Teaching and Learning in Higher Education*, *2*, 70–94. https://doi.org/10.62512/etlhe.20

Banegas, D. L. (2023). Four spheres of student-teachers' professional identity formation through learning about curriculum development. *Journal of Education for Teaching*, *49*(3), 370–383. https://doi.org/10.1080/02607476.2022.2105644

Cukurova, M. (2025). The interplay of learning, analytics and artificial intelligence in education: A vision for hybrid intelligence. *British Journal of Educational Technology*, *56*(2), 469–488. https://doi.org/10.1111/bjet.13514

Dergaa, I., Ben Saad, H., Glenn, J. M., Amamou, B., Ben Aissa, M., Guelmami, N., Fekih-Romdhane, F., & Chamari, K. (2024). From tools to threats: A reflection on the impact of artificial-intelligence chatbots on cognitive health. *Frontiers in Psychology*, *15*, Article 1259845. https://doi.org/10.3389/fpsyg.2024.1259845

Fang, L., & Zhou, X. (2025). From tool to co-learner: Exploring student engagement with GenAI through the lens of social constructivism. *Evaluation Review*. Advance online publication. https://doi.org/10.1177/0193841X251411618

Freeman, J. (2025, February). *Student generative AI survey 2025*. Higher Education Policy Institute. https://www.hepi.ac.uk/wp-content/uploads/2025/02/HEPI-Kortext-Student-Generative-AI-Survey-2025.pdf

Johnson, R. L., Pistilli, G., Menédez-González, N., Dias Duran, L. D., Panai, E., Kalpokiene, J., & Bertulfo, D. J. (2022). *The ghost in the machine has an American accent: Value conflict in GPT-3* [Preprint]. arXiv. https://arxiv.org/abs/2203.07785

Kasneci, E., Sessler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Günnemann, S., Hüllermeier, E., Krusche, S., Kutyniok, G., Michaeli, T., Nerdel, C., Pfeffer, J., Poquet, O., Sailer, M., Schmidt, A., Seidel, T., Stadler, M., Weller, J., Kuhn, J., & Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences*, *103*, Article 102274. https://doi.org/10.1016/j.lindif.2023.102274

Kim, J., Yu, S., Detrick, R., & Li, N. (2025). Exploring students' perspectives on generative AI-assisted academic writing. *Education and Information Technologies*, *30*(1), 1265–1300. https://doi.org/10.1007/s10639-024-12878-7

Kosmyna, N., Hauptmann, E., Yuan, Y. T., Situ, J., Liao, X.-H., Beresnitzky, A. V., Braunstein, I., & Maes, P. (2025). *Your brain on ChatGPT: Accumulation of cognitive debt when using an AI assistant for essay writing task* [Preprint]. arXiv. https://doi.org/10.48550/arXiv.2506.08872

Newton, P. M. (2025). How vulnerable are UK universities to cheating with new GenAI tools? A pragmatic risk assessment. *Assessment & Evaluation in Higher Education*, *50*(8), 1332–1343. https://doi.org/10.1080/02602938.2025.2511794

Perkins, M., Furze, L., Roe, J., & MacVaugh, J. (2024). 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*(6). https://doi.org/10.53761/q3azde36

Peters, D., Calvo, R. A., & Ryan, R. M. (2018). Designing for motivation, engagement and wellbeing in digital experience. *Frontiers in Psychology*, *9*, Article 797. https://doi.org/10.3389/fpsyg.2018.00797

Reinders, H. (2026). From learner autonomy to hybrid autonomy: A research agenda. *Artificial Intelligence in Language Education*, *2*(1), Article 104233. https://doi.org/10.29140/aile.2026.104233

Reinders, H., Kukulska-Hulme, A., & Wilson, A. (2024). *Humanising technology in language learning and teaching*. Oxford University Press.

Ryan, R. M., & Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist*, *55*(1), 68–78. https://doi.org/10.1037/0003-066X.55.1.68

Vaccino-Salvadore, S. (2023). Exploring the ethical dimensions of using ChatGPT in language learning and beyond. *Languages*, *8*(3), Article 191. https://doi.org/10.3390/languages8030191

Yazan, B. (2018). A conceptual framework to understand language teacher identities. *Journal of Second Language Teacher Education*, *1*(1), 21–48.

Zadorozhnyy, A., & Lai, W. Y. W. (2024). ChatGPT and L2 written communication: A game-changer or just another tool? *Languages*, *9*(1), Article 5. https://doi.org/10.3390/languages9010005

Zapata, G. C. (Ed.). (2025). *Generative AI technologies, multiliteracies, and language education*. Routledge.