

University Students' Use of Generative AI for English Language Learning: AI Readiness, Challenges, and Perceived Impact on Language Skills

Rastislav Metruk^{1,2}

¹Faculty of Education, Trnava University, Slovakia

²Faculty of Education, University of Ostrava, Czech Republic

rastislav.metruk@truni.sk

<https://doi.org/10.34190/ejel.24.4.4931>

An open access article under [CC Attribution 4.0](#)

Abstract: Generative AI (artificial intelligence) tools are now widely used by university students for language learning, yet empirical evidence from Central and Eastern Europe remains scarce, and a number of studies rely on closed-ended instruments that capture attitude strength but not the reasoning behind it. This mixed-methods study examines Slovak university students' use of generative AI for English language learning through three research questions addressing AI readiness, perceived challenges, and the impact on language skills development. A purpose-designed 62-item semi-structured questionnaire was administered to 153 students. The instrument combined 59 closed-ended items with three open-ended questions, creating a within-instrument triangulation design in which qualitative responses corroborated, contextualised, and extended the quantitative findings. An interesting tension emerged from the data: students reported high AI readiness and confidence in responsible use, yet expressed genuine concern about academic integrity and the consequences of AI use. Content reliability was the most prominent challenge, with fabricated references receiving the highest agreement score among the challenge items. The open-ended data reinforced this pattern, with content inaccuracy and over-reliance emerging as the most frequently cited problems. A clear split appeared between text-based and oral skills: students perceived strong AI benefits for vocabulary, grammar, reading, and writing, but reported limited or no perceived benefit for speaking, listening, and pronunciation, with listening falling in the Low range. The qualitative data revealed patterns the Likert scale could not detect: a self-reinforcing cycle of over-reliance and diminished critical thinking, and clear divisions on whether institutions should adopt or limit AI. An institutional gap was also documented: students disagreed that their university had provided adequate AI training, and independently identified AI literacy courses as the most frequently mentioned intervention. This study advances e-learning research by providing one of the first empirical analyses in this specific geographical context. It highlights the importance of including open-ended questions in surveys to reveal contradictions that closed-ended questions might miss. Additionally, it offers concrete recommendations for practice: embedding AI literacy in language curricula, teaching source verification alongside prompting skills, and preserving the dedicated speaking and listening practice for which generative AI is not a substitute.

Keywords: Generative artificial intelligence, English language learning, AI readiness, Higher education, Language skills

1. Introduction

Since the public release of ChatGPT in November 2022, generative artificial intelligence (AI) has moved from novelty to near-ubiquity in higher education. Tools such as ChatGPT, Google Gemini, or Claude now offer students capabilities that range from automated writing feedback and grammar correction to interactive conversational practice (Lee et al., 2026), and empirical research has expanded rapidly in response – though unevenly, with some regions and skill areas far better covered than others (Li, B. et al., 2025).

This expansion of research has also exposed genuine problems. Academic integrity concerns have multiplied as students negotiate unclear boundaries around acceptable AI use (Michel-Villarreal et al., 2023). The accuracy of AI-generated content remains unreliable, with fabricated references and plausible-sounding misinformation a known risk (Chan and Hu, 2023). Institutions face pressure to reform curricula in ways that foster digital literacy and responsible AI engagement (Walczak and Cellary, 2023), and in language education specifically, questions about learner dependency and the potential erosion of language skills have grown more urgent (Xiao and Zhi, 2023).

Two gaps in the existing literature motivate the present study. First, most empirical work has been conducted in East and Southeast Asian educational contexts or in well-resourced Anglophone institutions (Li, B. et al., 2025). Central and Eastern Europe is largely absent from this literature. Pikhart, Klímová and Al-Obaydi (2024), one of the few Central European contributions, examined Czech, Taiwanese, and Iraqi students' preferences for digital language learning tools, including ChatGPT, but their scope covered digital resources broadly rather than

generative AI specifically. Second, although mixed-methods designs are increasingly common in this field (Li, B. et al., 2025), several large-scale surveys of learner perceptions have used closed-ended instruments that capture attitude strength but not the reasoning behind it (e.g. Vo and Nguyen, 2024; Maxwell et al., 2025; Huang and Derakhshan, 2025). Likert scales measure the strength of attitudes but cannot capture the reasoning, contradictions, and contextual detail that open-ended questions elicit.

Against this background, the novelty of the present study lies in the combination of three features that, to our knowledge, prior work has not brought together. First, large-scale studies of perceptions of generative AI in higher education have been conducted in Vietnam (Vo and Nguyen, 2024), the United States (Maxwell et al., 2025), Hong Kong (Chan and Hu, 2023), and internationally across 76 countries (Yusuf, Pervin and Román-González, 2024), yet Central and Eastern Europe remains sparsely represented; the present study reports from a Slovak institution. Second, while Vo and Nguyen (2024) assessed perceived usefulness across the four macro-skills, much of the skill-specific research has focused on individual areas, most often writing (Song and Song, 2023; Guo and Wang, 2024) and less often speaking (Fathi, Rahimi and Derakhshan, 2024; Tai and Chen, 2024). The present study extends this work by comparing perceived impact across seven language domains – vocabulary, grammar, pronunciation, speaking, listening, reading, and writing – within a single instrument. Third, integrating open-ended items with quantitative measures within the same survey instrument makes it possible to identify contradictions between what students report about their own practice and what they perceive among other students, patterns that may remain hidden in closed-ended measures alone.

The present study addresses both gaps by examining Slovak university students' use of generative AI tools for English language learning, guided by three research objectives and corresponding research questions, which are set out in Section 3.

2. Literature Review

2.1 AI Readiness and Technology Acceptance

AI readiness in education refers to students' preparedness and competence to use AI tools productively for learning. Early usability research established a baseline: Shaikh et al. (2023) conducted a small-scale usability evaluation with ten university students and found that participants generally perceived ChatGPT as useful and easy to use for conversation, writing, grammar, and vocabulary tasks, though they concluded that AI should complement rather than replace human instruction. This link between AI literacy and effective engagement has since been confirmed from multiple theoretical angles. Vo and Nguyen (2024) applied a technology acceptance model to 369 Vietnamese English-major students and found a telling split – students considered ChatGPT easy to use but held only neutral views on its educational usefulness, suggesting that ease of access does not automatically translate into perceived learning value. Maxwell et al. (2025) surveyed 482 US students and documented that perceived readiness, benefits, and challenges varied across demographic groups, complicating the notion that students form a homogeneous group of AI adopters. From a motivational perspective, Du and Alm (2024) showed, using 24 New Zealand postgraduates, that ChatGPT supported autonomy and competence through a self-determination theory lens, but the picture was less positive for relatedness. Some students felt that AI use reduced human interaction. Huang and Derakhshan (2025) reinforced this multidimensional view through structural equation modelling with 363 Chinese EFL learners, finding that both learning motivation and digital literacy predicted AI adoption, which suggests that readiness cannot be reduced to a single factor. Taken together, these studies converge on a multidimensional view in which technical familiarity, motivational disposition and critical judgement operate jointly, and in which no single dimension predicts productive use on its own.

Prompting competence, which can be described as the ability to craft queries that produce useful AI responses, has emerged as a specific dimension of readiness. Tour and Zadorozhnyy (2025) proposed a framework grounded in the Four Resources Model, arguing that prompt literacy involves code-breaking, meaning-making, text-using, and text-analysing practices. In practice, learners need to understand how AI interprets input, evaluate its outputs, and refine their prompts iteratively – skills that overlap with broader language and digital literacies. Xiao and Zhi (2023) provided empirical support at a Chinese university, where EFL students demonstrated critical judgement and iterative prompt modification during AI-assisted tasks. Pikhart, Klímová and Al-Obaydi (2024), in one of the few Central European studies, explored Czech, Taiwanese, and Iraqi students' preferences for digital language learning tools, including ChatGPT, but their scope covered digital resources broadly rather than AI readiness as a distinct construct. Where universities do not offer structured guidance, students develop AI competence informally, risking uneven skill levels and inconsistent ethical practices.

2.2 Challenges of Using Generative AI in Language Learning

The challenges of generative AI in education cluster around several recurring themes, though the research increasingly shows that these are context-dependent rather than universal. Academic integrity, assessment design, and the need for new pedagogical approaches have been identified as the principal institutional-level concerns (Michel-Villarreal et al., 2023). Walczak and Cellary (2023) have argued that addressing them requires genuine curricular reform: not merely adding AI policies to existing syllabi, but rethinking how digital literacy is taught alongside disciplinary knowledge. The scale of the integrity concern appears to vary with context. Yusuf, Pervin and Román-González (2024), surveying 1,217 participants across 76 countries, found that perceptions of AI as an integrity threat depended heavily on cultural norms, institutional expectations, and whether ethical guidelines were available at all. Chan and Hu (2023) documented a related tension at the student level: learners simultaneously appreciated AI's utility and worried about the ethics of using it – an ambivalence that surfaces repeatedly across studies from different regions and educational systems.

Content reliability poses a distinct kind of challenge, particularly for language learners. Generative AI produces fluent, confident text that may be factually wrong, including fabricated references, a phenomenon commonly termed hallucination (Goh and Aryadoust, 2025). For students working in a foreign language, the problem is compounded: they may lack the domain expertise or linguistic confidence to recognise when an AI output is plausible yet incorrect.

Beyond accuracy, affective and social dimensions have begun to receive attention. Huang and Zou (2024) found that AI chatbots could enhance enjoyment and willingness to communicate, but outcomes depended on the specific tool and task. Rezai, Soyooof and Reynolds (2024) showed that emotion regulation mediated the relationship between ChatGPT use and well-being among EFL learners, and Wang, Y. (2025) noted that communication benefits depended on pedagogical integration and students' awareness of AI's limitations.

Two concerns that have received less empirical attention but may prove the most consequential are dependency and skill atrophy. Lee et al. (2026), in their systematic review, flagged both as under-researched, warning that students are increasingly delegating to AI the kinds of tasks, such as drafting arguments, evaluating sources, and structuring ideas, that traditionally built independent cognitive skills. Derakhshan and Taghizadeh (2026) explored this through a phenomenological study of L2 learners' lived experiences and concluded that AI acts as a double-edged sword for higher-order thinking: it can scaffold analytical reasoning by providing diverse perspectives and models; however, it can also undermine that reasoning if learners passively accept outputs and cease engaging with the intellectual work themselves. The challenge, in other words, is not just whether AI content is accurate, but whether the process of using AI leaves students more or less capable than they were before.

2.3 Impact of Generative AI on Language Skills Development

Research regarding AI's influence on particular language skills has progressed inconsistently, with writing receiving the most attention. Song and Song (2023) found significant gains in organisation, coherence, grammar, and vocabulary among Chinese EFL students receiving AI-assisted instruction, though students voiced concerns about dependency. Baskara (2023) highlighted ChatGPT's potential for personalised writing feedback while flagging unresolved ethical challenges. Guo and Wang (2024) compared ChatGPT's feedback with instructor feedback on fifty argumentative essays and found that ChatGPT produced significantly more feedback overall and distributed it relatively equally across content, organisation, and language, whereas teachers concentrated on content and language issues. The two sources also differed in feedback type, suggesting that AI and human feedback may complement rather than replace each other.

The literature on speaking presents a more positive perspective, but with a caveat. Experimental studies have reported improved speaking and willingness to communicate (Fathi, Rahimi and Derakhshan, 2024), reduced anxiety (Zhang, Meng and Ma, 2024), and differential effects by chatbot type (Wang, C. et al., 2024). Tai and Chen (2024) tested individual versus paired interaction modes, and Wang, Li and Zou (2025) examined AI speaking practice through positive psychology. Lyu, Lai and Guo's (2025) meta-analysis of 31 studies confirmed a medium effect ($g = 0.608$). The caveat is that virtually all this evidence comes from structured, researcher-designed interventions. Whether the benefits transfer to informal, self-directed AI use – how most students actually interact with generative AI – has not been tested.

Receptive skills represent the largest gap in the evidence base. Vocabulary and grammar have been documented as areas where AI can help (Song and Song, 2023; Shaikh et al., 2023), but evidence on reading comprehension, listening, and pronunciation is scarce. Çelik, Yangın Ersanlı and Arslanbay (2024) offer one of the few data points:

Turkish EFL undergraduates who read blog texts simplified by ChatGPT scored higher on comprehension tasks, though reading anxiety was unaffected. As Li, B. et al. (2025) observe in their review of 144 empirical studies, writing dominates the research agenda (42.4% of studies), while speaking, listening and reading have received comparatively less attention – an imbalance that may reflect real limitations in what text-based AI tools can do for non-text skills.

Read together, the evidence points to an asymmetry across skill areas: writing is the most extensively studied domain, while speaking, listening, and reading have received comparatively less attention; moreover, the positive evidence for oral skills in the studies reviewed here comes largely from structured chatbot-based interventions rather than informal, self-directed use. The present study contributes to this area by assessing perceived AI impact across seven language domains, while the limited evidence from Central Europe leaves open the question of how far findings generated largely in other educational contexts transfer to this setting.

3. Research Objectives and Research Questions

The present study examines university students' use of generative AI tools for English language learning and academic work in the Slovak higher education context. Three research objectives (ROs) and corresponding research questions (RQs) guide the investigation:

RO1: *To examine the AI readiness of Slovak university students for using generative AI tools in English language learning and academic work.*

RQ1: *How do students assess their own readiness to use generative AI tools for academic purposes, and how does that self-assessment relate to the preparation they report receiving from their institution?*

RO2: *To identify the challenges that students experience when using generative AI tools in their academic work.*

RQ2: *Which challenges do students identify as most significant, and how do the challenges they report in their own practice compare with those they perceive among other students?*

RO3: *To investigate the perceived impact of generative AI tools on the development of English language skills.*

RQ3: *How does perceived impact differ across text-based and oral/aural language skills, and what might explain that difference?*

4. Methodology

4.1 Research Design

This study adopted a mixed-methods design, implemented as a cross-sectional survey in which quantitative and qualitative data were collected concurrently within a single instrument. Descriptive statistics were computed for all Likert-scale items, with results reported by thematic category. Three open-ended questions were included to allow students to elaborate on the challenges they experience when using generative AI tools, and these responses were analysed using thematic content analysis. The strands were given complementary rather than equal weight: the quantitative strand establishes the distribution of attitudes across the sample, while the qualitative strand supplies the reasoning behind them. Following the convergent logic of within-instrument triangulation, the two were analysed separately and then compared, so that qualitative findings could corroborate, contextualise, or challenge the quantitative results.

4.2 Research Sample

The sample consisted of 153 university students specialising in English language studies at a Slovak higher education institution, selected through convenience sampling. The mean age of participants was 21.3 years, and the sample was predominantly female (82.4%), with 15.0% male and 2.6% identifying as non-binary or preferring not to disclose. The majority self-assessed their English proficiency at B2 level (54.2%), followed by C1 (30.7%), B1 (10.5%), and C2 (4.6%). Participants were drawn from all years of study, with a concentration in the earlier undergraduate years: 69 students (45.1%) were in the first year of Bachelor's study, 58 (37.9%) in the second, and 8 (5.2%) in the third, while 13 (8.5%) were in the first and 5 (3.3%) in the second year of Master's study. Bachelor's students therefore accounted for 88.2% of the sample and Master's students for 11.8%. Although the study was conducted at a Slovak institution, not all participants were Slovak nationals: 82.4% reported Slovak nationality, 11.8% Ukrainian, and 5.9% another nationality.

Data were collected online using Microsoft Forms. The questionnaire link was distributed through institutional communication channels. Participation was voluntary, and the form was configured to record responses

anonymously, so no names or email addresses were collected. Respondents gave informed consent on the opening page and could withdraw at any point by closing the form without submitting it.

Two features of the sample should be acknowledged as limitations at the outset. Convenience sampling from a single institution, combined with a predominantly female sample of English-major students, means that the findings describe this population rather than Slovak university students in general; the gender imbalance reflects the composition of English language programmes at the Slovak institution rather than a sampling decision.

4.3 Research Instrument

Data were collected using a purpose-designed, self-administered questionnaire comprising 62 items distributed across five sections. Section 1 (18 items) gathered demographic and background information. Sections 2–4 employed a five-point Likert scale (1 = Strongly disagree, 5 = Strongly agree): Section 2 (7 items) measured AI readiness and prompting competence, Section 3 (25 items) covered experiences with AI use across five thematic areas (academic integrity, dependency and skill atrophy, content reliability, social and psychological impact, and access and privacy), and Section 4 (9 items) assessed the perceived impact of generative AI on English language skills development. Section 5 comprised three open-ended questions inviting students to describe challenges and suggest interventions.

The questionnaire was developed specifically for this study, as no existing validated instrument comprehensively addressed all three research objectives. Item development was informed by the literature reviewed and guided by the three research objectives. The instrument was reviewed for face validity and content clarity prior to administration.

4.4 Reliability

Internal consistency was assessed using Cronbach's alpha. All three scales exceeded the threshold of $\alpha \geq .70$, considered acceptable for newly developed instruments (Nunnally, 1978; George and Mallery, 2003; Taber, 2018). Section 2 (AI Readiness) yielded $\alpha = .758$ (six items, after removing Q22 for low item-total correlation). Section 3 (Experiences with AI Use) yielded $\alpha = .741$ (22 items, after removing Q30, Q37, Q42). Section 4 (Language Skills) yielded $\alpha = .746$ (all nine items). The four removed items are reported separately as standalone items, as they remain substantively informative.

4.5 Data Analysis

Descriptive statistics – means (M) and standard deviations (SD) – were computed for each Likert-scale item. Statistical computations, including Cronbach's alpha coefficients and descriptive statistics, were performed using Claude (Anthropic, 2025), a large language model with code execution capabilities, and independently verified by the author using JASP, an open-source statistics program (JASP Team, 2026). Mean scores were interpreted using the levels-of-agreement framework proposed by Pimentel (2010): 1.00–1.79 = very low, 1.80–2.59 = low, 2.60–3.39 = neutral, 3.40–4.19 = high, 4.20–5.00 = very high. For the three open-ended questions, responses were first filtered to exclude non-substantive entries (e.g., 'No', 'Nothing', 'I don't know'). Recurring patterns were identified across responses and grouped into thematic categories, such as content reliability, over-reliance, and AI literacy needs. Because individual responses frequently addressed multiple themes, each distinct theme within a response was coded separately, and category frequencies reflect the number of mentions rather than the number of respondents.

5. Results

This section presents the descriptive statistics for the Likert-scale items across the three main sections of the questionnaire. Reverse-coded items are marked with an asterisk (*). Four standalone items (Q22, Q30, Q37, Q42) that were removed from their respective scales due to low item-total correlations are reported separately in Table 4.

Almost all respondents reported using generative AI tools in their studies: 19.6% did so daily, 40.5% a few times per week, 27.5% a few times per month, and 10.5% a few times per semester, while three respondents (2.0%) reported using no generative AI tools at all. Regarding the specific tools, versions, and purposes of use, ChatGPT was by far the most commonly used generative AI tool (92.2%), followed by Gemini (53.6%), NotebookLM (18.3%), Perplexity (13.7%), Copilot (13.1%), DeepSeek (8.5%), Meta AI (5.9%), Grok (5.2%), and Claude (4.6%). The vast majority of students (75.2%) used free versions only, with a further 15.0% using mostly free versions with occasional paid access. The most common academic purposes for AI use were language polishing (69.3%), understanding difficult concepts (62.7%), exam preparation (62.1%), and text summarisation (54.9%).

5.1 AI Readiness and Prompting Competence

Table 1 presents the six items comprising the AI Readiness and Prompting Competence scale ($\alpha = .758$). The overall scale mean was 3.85 (SD = 0.54), placing it in the High range on Pimentel's (2010) framework. The highest-rated item was Q25 ("I understand the limitations and potential risks of using generative AI"; $M = 4.33$, Very high), while the lowest was the reverse-coded Q21 ($M = 3.50$), indicating that a proportion of students still experience confusion when using AI tools despite rating their readiness positively overall.

Table 1: AI Readiness and Prompting Competence (N = 153, $\alpha = .758$)

Item	Statement	M	SD	Level
Q25	I understand the limitations and potential risks of using generative AI.	4.33	0.68	Very high
Q24	I feel confident in my ability to use AI tools responsibly.	4.13	0.72	High
Q23	I feel confident in my ability to use AI tools critically.	3.86	0.81	High
Q19	I feel adequately prepared to use generative AI tools for academic purposes.	3.66	0.83	High
Q20	I know how to write effective prompts to get useful and relevant responses from AI.	3.65	0.81	High
Q21*	I often feel lost or confused when trying to use AI tools effectively. (reverse-coded)	3.50	0.92	High

5.2 Experiences with AI Use

Table 2 presents the 22 items comprising the Experiences with AI Use scale ($\alpha = .741$). The overall scale mean was 3.06 (SD = 0.42), placing the average response at the Neutral level. Items are organised into five thematic groups for clarity, though these groups are not treated as independent sub-scales. The highest individual item mean was Q40 ("AI tools sometimes fabricate references"; $M = 4.18$, High), and students also agreed that AI output does not always meet academic standards (Q39; $M = 3.75$) and that paid AI tools create inequity of access (Q50; $M = 3.76$).

At the other end of the scale, students rejected the idea that they accept AI-generated answers without verification (Q35; $M = 2.17$, Low) – the lowest item mean across the entire questionnaire. They also disagreed that AI use had harmed their relationships with classmates or teachers (Q48*; $M = 2.32$) or caused feelings of isolation (Q46; $M = 2.44$). One finding worth noting is Q45 ($M = 3.14$): students neither agreed nor disagreed that they consult teachers less frequently because they can ask AI instead, suggesting an emerging but not yet decisive shift in help-seeking behaviour from human to AI sources.

Table 2: Experiences with AI Use (N = 153, $\alpha = .741$)

Item	Statement	M	SD	Level
Academic Integrity and Ethical Concerns				
Q31	I worry about the consequences if my teacher discovers that I have used AI in my work.	3.44	1.03	High
Q27	I worry about presenting work that was created with AI assistance as my own.	3.35	1.13	Neutral
Q28	I am uncertain about how to properly acknowledge or cite AI-generated content.	3.27	1.05	Neutral
Q29	I feel guilty when I use AI to complete parts of my assignments.	3.20	1.11	Neutral
Q26	I find it difficult to determine how much AI assistance is acceptable in my academic work.	2.80	1.04	Neutral
Dependency and Skill Atrophy				
Q34	Using AI makes me less likely to think critically about texts and tasks.	3.07	1.07	Neutral
Q36	I feel that AI does the intellectual work that I should be doing myself.	3.07	1.09	Neutral

Item	Statement	M	SD	Level
Q32	I am concerned that using AI reduces my motivation to develop my own English language skills.	3.03	1.24	Neutral
Q33	I notice that my own writing ability has declined since I started using AI regularly.	2.65	1.18	Neutral
Q35	I tend to accept AI-generated answers without verifying or questioning them.	2.17	1.01	Low
Content Reliability				
Q40	AI tools sometimes fabricate references or sources that do not actually exist.	4.18	0.90	High
Q39	AI-generated content sometimes does not meet the standards or expectations of academic writing in my field.	3.75	0.98	High
Q38	I find it challenging to deal with inaccurate or misleading information generated by AI tools.	3.01	0.98	Neutral
Q41	I struggle to identify errors or biases in AI-generated content.	2.81	0.99	Neutral
Social and Psychological Impact				
Q45	I consult my teachers less frequently because I can ask AI instead.	3.14	1.12	Neutral
Q43	I experience technical problems when using AI tools.	3.03	1.03	Neutral
Q47	I experience stress or anxiety about keeping up with rapid changes in AI technology.	2.65	1.19	Neutral
Q44	Using AI reduces my interaction with classmates on academic tasks.	2.48	1.03	Low
Q46	I feel more isolated in my studies since I started using AI tools regularly.	2.44	0.85	Low
Q48*	Using AI has not negatively affected my relationships with classmates or teachers. (reverse-coded)	2.32	1.11	Low
Access and Privacy				
Q50	The cost of paid versions of AI tools makes it harder for some students to benefit from AI.	3.76	1.02	High
Q49	I am concerned about the privacy and security of my personal data when using AI tools.	3.63	1.08	High

5.3 AI and English Language Skills Development

Table 3 presents the nine items on perceived AI impact on language skills ($\alpha = .746$). The overall scale mean was 3.17 (SD = 0.59), Neutral on Pimentel's framework. The data split into two distinct clusters, visible in Figure 1. Text-based skills scored in the High range: vocabulary (Q52; M = 3.71), reading (Q56; M = 3.59), grammar (Q51; M = 3.57), and writing (Q57; M = 3.48). Oral and aural skills scored lower: speaking (Q55; M = 2.74), pronunciation (Q53; M = 2.80), and listening (Q54; M = 2.57), with listening falling in the Low range. The 95% confidence intervals for these two clusters do not overlap (Figure 1), confirming that the distinction is not attributable to marginal score differences.

Two items assessed students' self-reflective awareness of their dependence on AI in language learning. Students expressed moderate concern that AI may be doing the language work for them rather than helping them learn (Q58; M = 3.01) and that their skills would improve with less AI reliance (Q59; M = 3.05). Both fell at the Neutral level – students neither dismissed nor strongly endorsed the risk.

Table 3: AI and English Language Skills Development (N = 153, $\alpha = .746$)

Item	Statement	M	SD	Level
Q52	Using AI helps me expand my English vocabulary.	3.71	0.97	High
Q56	Using AI helps me improve my English reading skills.	3.59	0.98	High

Item	Statement	M	SD	Level
Q51	Using AI helps me improve my English grammar.	3.57	0.99	High
Q57	Using AI helps me improve my English writing skills.	3.48	1.05	High
Q59	I feel that my English language skills would be better if I relied less on AI.	3.05	1.07	Neutral
Q58	I am concerned that AI is doing the language work for me rather than helping me learn.	3.01	0.99	Neutral
Q53	Using AI helps me improve my English pronunciation.	2.80	1.09	Neutral
Q55	Using AI helps me improve my English speaking skills.	2.74	1.13	Neutral
Q54	Using AI helps me improve my English listening skills.	2.57	1.02	Low

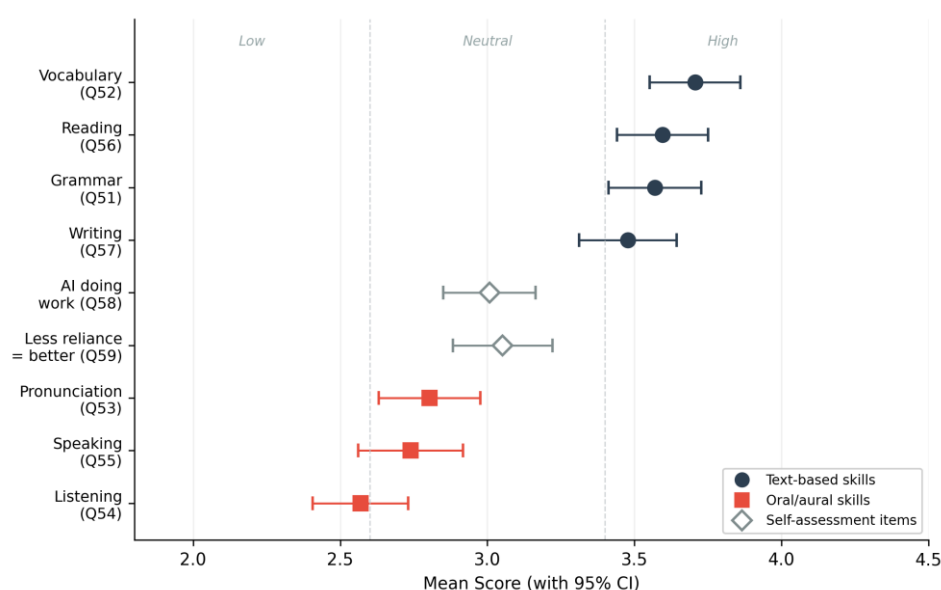


Figure 1: Mean scores with 95% confidence intervals for perceived AI impact on English language skills (N = 153). Dashed lines indicate Pimentel (2010) level boundaries. Text-based skills (circles) cluster in the High range, while oral/aural skills (squares) fall in the Low to Neutral range

5.4 Standalone Items

Table 4 presents the four items removed from their respective scales due to low item-total correlations, reported individually. The lowest score belongs to Q22 (M = 2.46, Low): students disagreed that their university had provided sufficient training on AI tools, a result that stands in tension with the high self-perceived readiness reported in Section 5.1. At the other end, Q42 (M = 3.97) indicates that most students report verifying AI-generated content before using it. The remaining two items fell in the Neutral range: Q30 (M = 3.24) on confidence in using AI without compromising integrity, and Q37 (M = 3.38) on whether AI helps students learn efficiently without eroding their own skills.

Table 4: Standalone Items (N = 153)

Item	Statement	M	SD	Level
Q42	I verify AI-generated content before using it in my work.	3.97	0.92	High
Q37	Using AI helps me learn more efficiently without reducing my own skills.	3.38	0.91	Neutral
Q30	I feel confident that I can use AI without compromising my academic integrity.	3.24	0.86	Neutral

Item	Statement	M	SD	Level
Q22	I have received sufficient training or guidance from my university on how to use AI tools.	2.46	1.00	Low

5.5 Qualitative Insights from Open-Ended Questions

Three open-ended questions were included to allow students to elaborate on their experiences. Of the 153 respondents, 44 (28.8%) responded to the first question, 73 (47.7%) to the second, and 58 (37.9%) to the third. After excluding non-substantive responses, the analysis was based on 21, 71, and 46 substantive responses, respectively. Responses were analysed using thematic content analysis (see Figure 2). Because individual responses frequently addressed multiple themes, frequencies reflect mentions rather than respondents.

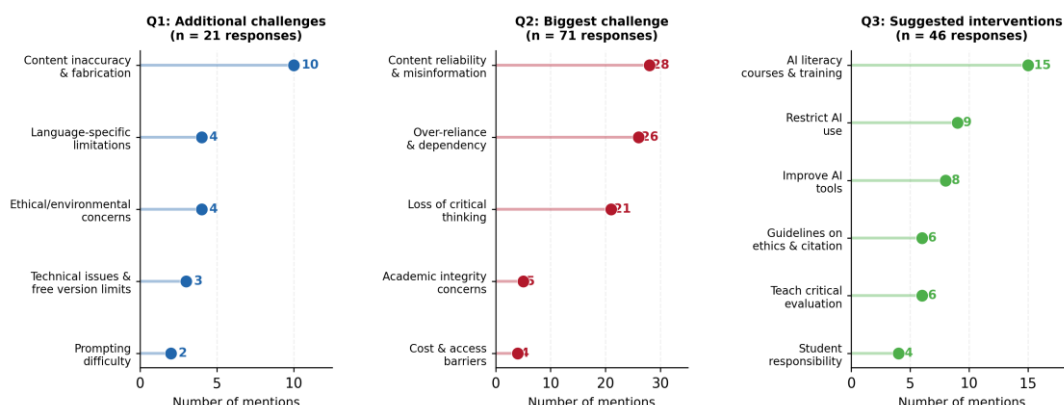


Figure 2: Thematic frequency of mentions across the three open-ended questions. Themes are ranked by frequency within each question. Because individual responses often addressed multiple themes, frequencies reflect the number of mentions, not the number of respondents

Question 1: “Are there any other challenges you have experienced when using AI tools for your studies that were not covered in the statements above?”

Among the 21 substantive responses, the most frequently mentioned additional challenge was content inaccuracy and fabrication ($n = 10$), with students describing fabricated sources and the need for repeated prompting. Language-specific limitations were raised by four respondents, who noted issues with translation quality, clichéd phrasing, and language mixing in AI output. Four respondents expressed ethical or environmental opposition to AI, and three mentioned technical issues and free version limitations. Two respondents described difficulty with prompting, noting that formulating effective queries requires considerable skill.

Question 2: “In your opinion, what is the biggest challenge that students face when using AI for learning and academic work?”

This question yielded the richest data (see Figure 2). Thematic analysis of the 71 substantive responses revealed three dominant, frequently overlapping themes. Content reliability and misinformation ($n = 28$) was the most frequently mentioned theme, followed closely by over-reliance and dependency ($n = 26$), with one respondent writing that students “stop relying on their knowledge and only consult everything with AI.” Loss of critical thinking ($n = 21$) was also prominent. Academic integrity ($n = 5$) and cost barriers ($n = 4$) were mentioned less frequently. The themes of over-reliance, critical thinking, and content reliability were deeply intertwined in many responses, suggesting that students perceive these not as separate issues but as interconnected dimensions of a broader problem.

Question 3: “Is there anything that your university, teachers, or AI tool developers could do to help you deal with these challenges?”

Of the 46 substantive responses, the most frequent suggestion was AI literacy courses and training ($n = 15$), covering prompt writing, AI limitations, and long-term learning strategies. A minority ($n = 9$) advocated restricting AI use rather than teaching it. Guidelines on ethical use and citation ($n = 6$), critical evaluation skills ($n = 6$), and improvements to AI tools ($n = 8$) were also mentioned. Four respondents placed responsibility on

the students themselves. The tension between those calling for integration and those advocating restriction reflects the broader ambivalence in the quantitative findings.

6. Discussion

The picture that emerges from these data is not a straightforward one. Students presented themselves as confident, critical AI users who verify outputs and understand risks, yet the same cohort described a peer culture of uncritical dependency and cognitive offloading. They rated content fabrication as a serious problem but reported only moderate concern about their own skill development. The quantitative data, taken alone, would paint a reassuringly balanced portrait; the qualitative data disrupt that portrait in ways that matter for how the findings should be interpreted and how institutional responses should be designed.

6.1 AI Readiness and the Institutional Gap (RQ1)

The high overall readiness score ($M = 3.85$) warrants closer examination rather than celebration. Students rated their understanding of AI limitations as the strongest dimension of their readiness (Q25; $M = 4.33$), and their prompting competence as adequate (Q20; $M = 3.65$), figures broadly in line with previous findings from Vietnam (Vo and Nguyen, 2024) and China (Xiao and Zhi, 2023). But there is reason to question how far self-report captures genuine competence. Du and Alm (2024) found that students often lacked the ability to effectively prompt AI or engage in post-editing of its output, even when they felt competent in using the tool, and the present data contain a similar tension: if students truly understand AI's limitations so well (Q25; $M = 4.33$), why do so many of their peers, as described in the open-ended responses, accept AI outputs uncritically? One plausible explanation is social desirability bias; another is the Dunning-Kruger adjacent possibility that students overestimate competencies they have never been formally taught to assess.

This last point connects to what is arguably the most consequential finding in the readiness data: the institutional gap. Students disagreed that their university had provided sufficient AI training (Q22; $M = 2.46$), placing this item firmly in the Low range. The implication is that whatever readiness these students possess, they acquired it on their own, through informal experimentation rather than structured guidance. This is not a neutral outcome. Informal learning is by definition uneven; students with greater digital capital, English proficiency, or curiosity will develop stronger AI literacy than those without, reproducing existing inequalities under the guise of individual initiative. Walczak and Cellary (2023) warned that institutions needed to respond with curricular reform; nearly three years later, the present data suggest that response has not materialised, at least not in this context. The qualitative data sharpen this interpretation. AI literacy courses were the single most frequently requested intervention ($n = 15$), covering prompt writing, AI limitations, and long-term learning strategies, and students separately called for guidelines on ethical use and citation ($n = 6$) and training in critical evaluation of AI output ($n = 6$). These are not vague complaints – they amount to a concrete curriculum that the institution has, to date, failed to provide.

6.2 Challenges and the Paradox of Critical Awareness (RQ2)

The overall challenge mean of 3.06 (Neutral) is, on its own, misleading. It averages together a High-level concern about fabricated references (Q40; $M = 4.18$) with Low-level disagreement about social isolation (Q46; $M = 2.44$), flattening meaningful variation into an uninformative midpoint. The item-level data tell a more interesting story. Content reliability was the clear standout: students recognised that AI fabricates sources (Q40; $M = 4.18$) and that its output often falls short of academic standards (Q39; $M = 3.75$). This is consistent with the phenomenon commonly termed hallucination (Goh and Aryadoust, 2025), but the present data add a practical dimension. Students are not just aware of hallucination in the abstract; when asked about challenges they had personally experienced, content inaccuracy and fabrication were the most frequently raised themes ($n = 10$), with students describing fabricated sources and the need for repeated prompting.

A genuine paradox lies at the heart of these data and deserves closer scrutiny. On the Likert scale, students reported diligent verification practices: they rejected uncritical acceptance of AI output (Q35; $M = 2.17$, the lowest item mean in the entire questionnaire) and rated their own verification behaviour highly (Q42; $M = 3.97$). Yet in an open-ended question about the biggest challenge their peers face, two dominant themes were over-reliance ($n = 26$) and loss of critical thinking ($n = 21$). This is not a minor discrepancy.

One reading is straightforward social desirability: students know they should verify, so they report doing so, while freely observing that others do not. But a less comfortable interpretation is also possible: that the gap between stated and actual verification behaviour may be smaller between the respondent and the described peers than the Likert data suggest. This cannot be resolved with the present data, but the pattern should give

pause to any study that relies on self-reported critical engagement as evidence that students are using AI responsibly.

The academic integrity items reveal genuine ethical uncertainty. Students worried about teachers discovering their AI use (Q31; $M = 3.44$) and felt uneasy about presenting AI-assisted work as their own (Q27; $M = 3.35$), but their concern was rooted not in a clear sense that AI use is wrong, but in uncertainty about where the line is. This mirrors Yusuf, Pervin and Román-González's (2024) cross-cultural finding that integrity perceptions vary with the availability of ethical guidelines – where guidelines are absent, anxiety fills the vacuum. The dependency data offer the most consequential example of why mixed-methods designs matter in this field. The quantitative aspects of dependency mostly received neutral ratings, making them easy to overlook. But the open-ended responses painted a different picture: over-reliance and critical thinking loss were two prominent and closely related themes, described not as separate problems but as a single reinforcing cycle. Lee et al. (2026) flagged dependency as under-researched; Derakhshan and Taghizadeh (2026) characterised AI as a double-edged sword for higher-order thinking. The qualitative data from the present study suggest that students already perceive the sword cutting, even if the Likert scale does not register the wound.

6.3 Perceived Impact on Language Skills (RQ3)

The text-based versus oral/aural split is the cleanest pattern in the dataset and, it is argued here, the most practically important. Vocabulary ($M = 3.71$), reading ($M = 3.59$), grammar ($M = 3.57$), and writing ($M = 3.48$) all fell in the High range; speaking ($M = 2.74$), listening ($M = 2.57$), and pronunciation ($M = 2.80$) did not exceed the Neutral threshold, with listening falling into the Low range. The non-overlapping 95% confidence intervals (Figure 1) confirm this is not a product of marginal score differences. The explanation is partly technological: current generative AI tools are text-processing systems at their core. They excel at manipulating written language because that is what they were built to do. Pronunciation, listening comprehension, and spontaneous speech production involve embodied, temporal, and auditory dimensions that a text interface cannot replicate. The pattern mirrors what Song and Song (2023), Shaikh et al. (2023), and Baskara (2023) reported for text-based skills, and the qualitative responses added a corrective: students noted that AI output tends toward clichéd phrasing and occasionally mixes languages – limitations invisible to the Likert items.

The speaking result deserves separate attention because it sits in tension with a body of experimental research reporting positive effects. Fathi, Rahimi and Derakhshan (2024) found improvements in speaking and willingness to communicate, Zhang, Meng and Ma (2024) reported reduced speaking anxiety, and Lyu, Lai and Guo's (2025) meta-analysis confirmed a medium effect across 31 studies ($g = 0.608$). But those studies employed structured, researcher-designed interventions, dedicated speaking practice sessions with AI chatbots under controlled conditions (Tai and Chen, 2024; Wang, C. et al., 2024). The students in the present study, by contrast, reported predominantly text-oriented uses of AI, including language polishing and text summarisation. This pattern may help explain why perceived benefits were stronger for text-based than for oral/aural skills. The discrepancy, then, is not between the present findings and the experimental literature; it is between the affordances of structured pedagogical AI use and the reality of how students actually use these tools when left to their own devices. Çelik, Yangın Ersanlı and Arslanbay (2024) showed that AI-simplified texts can improve reading comprehension, suggesting that even receptive skills may benefit when AI is used thoughtfully for input modification. The broader point is that generative AI's language learning potential is not an intrinsic property of the tool; it is a function of how the tool is used, and informal use skews heavily toward text.

6.4 The Methodological Value of Integrating Quantitative and Qualitative Data

A recurring argument throughout this discussion has been that the open-ended questions did genuine analytical work and were not merely decorative. It is worth making the case explicitly. The qualitative data operated on three levels. First, corroboration: the prominence of content reliability in the open-ended responses ($n = 28$ mentions) independently confirmed the highest-rated challenge item (Q40; $M = 4.18$), and the salience of over-reliance ($n = 26$) tracked the moderate quantitative concern about dependency. When two data sources converge, each strengthens the other's credibility. Second, contextualisation: the qualitative data showed what the numbers describe but cannot explain – what it is like to encounter a fabricated reference mid-assignment, how students develop ad hoc verification routines, and how skill atrophy registers in a learner's own perception of declining fluency. Third, and most important, extension: the open-ended responses revealed patterns that the Likert scale was structurally unable to capture. Students described over-reliance, critical thinking loss, and misinformation not as separate challenges but as a single reinforcing cycle. They split into opposing camps on whether institutions should integrate AI more deeply or restrict it, a division no closed-ended item had

anticipated. And the verification paradox discussed above would have been entirely invisible without the qualitative layer.

The methodological implication is direct. Survey-based research relying exclusively on Likert scales risks producing a sanitised picture that might not reflect how students experience these tools. Self-report scales are vulnerable to social desirability, acquiescence bias, and the averaging-out of complex attitudes into uninformative midpoints. The within-instrument triangulation employed here exposed contradictions that a purely quantitative instrument would have flattened, and researchers designing future surveys should consider qualitative components not as optional addenda but as core elements of instrument validity. Finally, this study is one of the first to report from Central Europe. The Slovak students exhibited patterns that share features with other contexts, including strong perceptions of text-based skills and concerns about content reliability, while diverging in others. The institutional training gap documented here may reflect a broader pattern in contexts where AI-related curricular reform is still at an early stage. It is worth distinguishing which findings are likely to travel and which are tied to this setting. The text-oral asymmetry should be the most transferable, since it follows from the architecture of current generative AI tools rather than from any local condition. The verification paradox is likewise a property of self-report methodology, and would be expected wherever socially desirable responding is possible. The size of the institutional training gap, by contrast, is a local finding, as is the prominence of cost as a barrier, which is sensitive to national income levels. Whether that structural factor accounts for the gap, or whether it is equally prevalent elsewhere but less frequently measured, remains a question for comparative research.

7. Conclusion

This study set out to examine how Slovak university students use generative AI for English language learning, and the answer turns out to be more complicated than any single data source could reveal. Students are confident, largely self-taught AI users who recognise the technology's limitations in the abstract but navigate a daily reality of fabricated references, unclear ethical boundaries, and an institutional environment that is still adapting to the tools they are already using. The tension between high self-reported readiness and limited institutional training carries practical implications, because informal learning produces uneven competencies across the student body.

Three findings carry direct practical weight. First, AI literacy deserves a more prominent place in university curricula; the students themselves identified this as the most needed intervention. Second, the sharp text-oral skills divide means that generative AI cannot substitute for the interactive, embodied language practice that develops speaking, listening, and pronunciation; institutions should therefore preserve dedicated practice in these skills and be explicit with students about why it matters. Third, the qualitative-quantitative divergence regarding dependency and verification behaviour should concern anyone designing assessments or policies based on what students say they do with AI; what they report about their peers suggests a more troubled picture than self-report alone can capture.

The study has limitations. The sample from a single Slovak institution, with a predominance of female participants, limits generalisability; self-report data carry inherent biases, and the cross-sectional design cannot capture how attitudes evolve over time. Future research should track students longitudinally, compare self-reported verification with actual practice through observational methods, and test whether structured AI integration into curricula can close the readiness gap. Cross-national comparisons within Central and Eastern Europe would clarify whether these patterns are regional or local.

The question facing higher education is no longer whether students will use generative AI for language learning. They already do so overwhelmingly, largely without institutional guidance. The question is whether universities will shape that use or merely react to it. Every semester of inaction risks widening the gap between what students need and what institutions provide, potentially affecting students with fewer digital, linguistic, or financial resources most strongly. The data presented here, from a context rarely represented in this literature, suggest that students are not passive consumers of AI but active, ambivalent agents navigating a technology that is simultaneously powerful and unreliable. Understanding how students engage with these tools is a necessary first step toward informed policy and curriculum design.

Acknowledgement

This article is from Project no. B4/2026, "Using Generative Artificial Intelligence Tools in the Foreign Language Teaching Process – Challenges and Perspectives", funded by the Faculty of Education, Trnava University, Slovakia

and from the project "Implementation of the Reform of Teacher Education at the University of Ostrava: Competence-based learning founded on practice, research, internationalisation and cooperation with the clinical school CZ.02.02.XX/00/23_019/0008272", which is co-financed by the European Union.

Ethics Statement: Ethics approval was not required for this study. All participants gave their informed consent for inclusion before they participated in the study. Participation was entirely voluntary, and respondents could withdraw at any point before submitting without consequences. No personally identifiable information was collected, and all data were stored securely and used solely for research purposes. The study was conducted in accordance with the Declaration of Helsinki and the ethical standards of the institution.

AI Declaration: Generative AI tools were used in the preparation of this manuscript. Specifically, Claude Opus 4 (Anthropic, 2025) was used to perform statistical computations (Cronbach's alpha coefficients and descriptive statistics) and to assist with manuscript formatting, language editing, and the preparation of tables and figures. GPT-5.6 Sol (OpenAI, 2026) was also used during the revision process to cross-check statistical calculations, verify the accuracy of literature-based claims and references, and assist with language and internal consistency checks. All AI-assisted statistical outputs and figures were independently verified by the author against the underlying data, and the statistical analyses were independently reproduced in JASP (an open-source statistics program). The author takes full responsibility for the content and integrity of the final manuscript.

Conflict of Interest Statement: The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Funding: This article was funded by Project no. B4/2026, "Using Generative Artificial Intelligence Tools in the Foreign Language Teaching Process – Challenges and Perspectives", funded by the Faculty of Education, Trnava University, Slovakia.

References

- Anthropic, 2025. Claude Opus 4 [Large language model]. Available at: <https://www.anthropic.com/claude> [Accessed 27 May 2026].
- Baskara, F.X.R., 2023. Integrating ChatGPT into EFL writing instruction: benefits and challenges. *International Journal of Education and Learning*, 5(1), pp 44–55. <https://doi.org/10.31763/ijelev.v5i1.858>.
- Çelik, F., Yangın Ersanlı, C. and Arslanbay, G., 2024. Does AI simplification of authentic blog texts improve reading comprehension, inferencing, and anxiety? A one-shot intervention in Turkish EFL context. *The International Review of Research in Open and Distributed Learning*, 25(3), pp 287–303. <https://doi.org/10.19173/irrodl.v25i3.7779>.
- Chan, C.K.Y. and Hu, W., 2023. Students' voices on generative AI: perceptions, benefits, and challenges in higher education. *International Journal of Educational Technology in Higher Education*, 20, 43. <https://doi.org/10.1186/s41239-023-00411-8>.
- Derakhshan, A. and Taghizadeh, M.S., 2026. Does artificial intelligence (AI) nurture or hinder language learners' higher-order thinking skills (HOTS)? A phenomenological study on L2 learners' perspectives and lived experiences. *International Journal of Applied Linguistics*. <https://doi.org/10.1111/ijal.12824>.
- Du, J. and Alm, A., 2024. The impact of ChatGPT on English for Academic Purposes (EAP) students' language learning experience: a self-determination theory perspective. *Education Sciences*, 14(7), 726. <https://doi.org/10.3390/educsci14070726>.
- Fathi, J., Rahimi, M. and Derakhshan, A., 2024. Improving EFL learners' speaking skills and willingness to communicate via artificial intelligence-mediated interactions. *System*, 121, 103254. <https://doi.org/10.1016/j.system.2024.103254>.
- George, D. and Mallery, P., 2003. *SPSS for Windows step by step: a simple guide and reference, 11.0 update*. 4th ed. Boston: Allyn & Bacon.
- Goh, C.C.M. and Aryadoust, V., 2025. Developing and assessing second language listening and speaking: does AI make it better? *Annual Review of Applied Linguistics*, 45, pp 179–199. <https://doi.org/10.1017/S0267190525100111>.
- Guo, K. and Wang, D., 2024. To resist it or to embrace it? Examining ChatGPT's potential to support teacher feedback in EFL writing. *Education and Information Technologies*, 29, pp 8435–8463. <https://doi.org/10.1007/s10639-023-12146-0>.
- Huang, F. and Derakhshan, A., 2025. Learning motivation and digital literacy in AI adoption for self-regulated English learning. *European Journal of Education*, 60(4), e70254. <https://doi.org/10.1111/ejed.70254>.
- Huang, F. and Zou, B., 2024. English speaking with artificial intelligence (AI): the roles of enjoyment, willingness to communicate with AI, and innovativeness. *Computers in Human Behavior*, 159, 108355. <https://doi.org/10.1016/j.chb.2024.108355>.
- JASP Team, 2026. JASP (Version 0.98.0) [Computer software]. <https://jasp-stats.org> [Accessed 15 August 2026].
- Lee, S., Choe, H., Zou, D. and Jeon, J., 2026. Generative AI (GenAI) in the language classroom: a systematic review. *Interactive Learning Environments*, 34(1), pp 335–359. <https://doi.org/10.1080/10494820.2025.2498537>.
- Li, B., Tan, Y.L., Wang, C. and Lowell, V.L., 2025. Two years of innovation: a systematic review of empirical generative AI research in language learning and teaching. *Computers and Education: Artificial Intelligence*, 9, 100445. <https://doi.org/10.1016/j.caeai.2025.100445>.

- Lyu, B., Lai, C. and Guo, J., 2025. Effectiveness of chatbots in improving language learning: a meta-analysis of comparative studies. *International Journal of Applied Linguistics*, 35(2), pp 834–851. <https://doi.org/10.1111/ijal.12668>.
- Maxwell, D., Oyarzun, B., Kim, S. and Bong, J.Y., 2025. Generative AI in higher education: demographic differences in student perceived readiness, benefits, and challenges. *TechTrends*, 69, pp 1248–1259. <https://doi.org/10.1007/s11528-025-01109-6>.
- Michel-Villarreal, R., Vilalta-Perdomo, E., Salinas-Navarro, D.E., Thierry-Aguilera, R. and Gerardou, F.S., 2023. Challenges and opportunities of generative AI for higher education as explained by ChatGPT. *Education Sciences*, 13(9), 856. <https://doi.org/10.3390/educsci13090856>.
- Nunnally, J.C., 1978. *Psychometric theory*. 2nd ed. New York: McGraw-Hill.
- OpenAI, 2026. GPT-5.6 Sol [Large language model]. Available at: <https://developers.openai.com/api/docs/models/gpt-5.6-sol> [Accessed 14 August 2026].
- Pikhart, M., Klímová, B. and Al-Obaydi, L.H., 2024. Exploring university students' preferences and satisfaction in utilizing digital tools for foreign language learning. *Frontiers in Education*, 9, 1412377. <https://doi.org/10.3389/feduc.2024.1412377>.
- Pimentel, J.L., 2010. A note on the usage of Likert scaling for research data analysis. *USM R&D Journal*, 18(2), pp 109–112.
- Rezai, A., Soyoo, A. and Reynolds, B.L., 2024. Disclosing the correlation between using ChatGPT and well-being in EFL learners: considering the mediating role of emotion regulation. *European Journal of Education*, 59(4), e12752. <https://doi.org/10.1111/ejed.12752>.
- Shaikh, S., Yayilgan, S.Y., Klímová, B. and Pikhart, M., 2023. Assessing the usability of ChatGPT for formal English language learning. *European Journal of Investigation in Health, Psychology and Education*, 13(9), pp 1937–1960. <https://doi.org/10.3390/ejihpe13090140>.
- Song, C. and Song, Y., 2023. Enhancing academic writing skills and motivation: assessing the efficacy of ChatGPT in AI-assisted language learning for EFL students. *Frontiers in Psychology*, 14, 1260843. <https://doi.org/10.3389/fpsyg.2023.1260843>.
- Taber, K.S., 2018. The use of Cronbach's alpha when developing and reporting research instruments in science education. *Research in Science Education*, 48, pp 1273–1296. <https://doi.org/10.1007/s11165-016-9602-2>.
- Tai, T.-Y. and Chen, H.H.-J., 2024. Improving elementary EFL speaking skills with generative AI chatbots: exploring individual and paired interactions. *Computers & Education*, 220, 105112. <https://doi.org/10.1016/j.compedu.2024.105112>.
- Tour, E. and Zadorozhnyy, A., 2025. Conceptualizing and operationalizing prompt literacy for English language learners. *Journal of Adolescent & Adult Literacy*, 69(3), e70020. <https://doi.org/10.1002/jaal.70020>.
- Vo, T.K.A. and Nguyen, H., 2024. Generative artificial intelligence and ChatGPT in language learning: EFL students' perceptions of technology acceptance. *Journal of University Teaching and Learning Practice*, 21(6). <https://doi.org/10.53761/fr1rkj58>.
- Walczak, K. and Cellary, W., 2023. Challenges for higher education in the era of widespread access to generative AI. *Economics and Business Review*, 9(2), pp 71–100. <https://doi.org/10.18559/eb.2023.2.743>.
- Wang, C., Li, X. and Zou, B., 2025. Revisiting integrated model of technology acceptance among the generative AI-powered foreign language speaking practice: through the lens of positive psychology and intrinsic motivation. *European Journal of Education*, 60(1), e70054. <https://doi.org/10.1111/ejed.70054>.
- Wang, C., Zou, B., Du, Y. and Wang, Z., 2024. The impact of different conversational generative AI chatbots on EFL learners: an analysis of willingness to communicate, foreign language speaking anxiety, and self-perceived communicative competence. *System*, 127, 103533. <https://doi.org/10.1016/j.system.2024.103533>.
- Wang, Y., 2025. A study on the efficacy of ChatGPT-4 in enhancing students' English communication skills. *SAGE Open*, 15(1), pp 1–17. <https://doi.org/10.1177/21582440241310644>.
- Xiao, Y. and Zhi, Y., 2023. An exploratory study of EFL learners' use of ChatGPT for language learning tasks: experience and perceptions. *Languages*, 8(3), 212. <https://doi.org/10.3390/languages8030212>.
- Yusuf, A., Pervin, N. and Román-González, M., 2024. Generative AI and the future of higher education: a threat to academic integrity or reformation? Evidence from multicultural perspectives. *International Journal of Educational Technology in Higher Education*, 21, 21. <https://doi.org/10.1186/s41239-024-00453-6>.
- Zhang, C., Meng, Y. and Ma, X., 2024. Artificial intelligence in EFL speaking: impact on enjoyment, anxiety, and willingness to communicate. *System*, 121, 103259. <https://doi.org/10.1016/j.system.2024.103259>.