DOI: https://doi.org/10.28925/2412-0774.2026.3.12

UDC 811.111'243+37.016-042.2:004.8

Olesia Stoyka
https://orcid.org/0000-0002-7695-6100
Doctor of Sciences in Pedagogy, Professor,
Professor of the Applied Linguistics Department,
Uzhhorod National University,
14 Universytetska Str., 88000, Uzhhorod, Ukraine;
Chief Research Fellow,
Department of Comparative Pedagogy,
Institute of Pedagogy of the National Academy of Educational Sciences of Ukraine,
52-D Artema Str., 04053, Kyiv, Ukraine
olesya.stoyka@uzhnu.edu.ua

Yelyzaveta Dupyn
https://orcid.org/0009-0005-0061-2490
MA student, Department of Applied Linguistics,
Faculty of International Economic Relations,
Uzhhorod National University,
14 Universytetska Str., 88000, Uzhhorod, Ukraine
dupyn.yelyzaveta@student.uzhnu.edu.ua

THE IMPACT OF AI ON LANGUAGE LEARNING: BETWEEN REAL PROGRESS AND THE ILLUSION OF FLUENCY

The rapid integration of generative artificial intelligence (AI) into English as a Foreign Language (EFL) education has prompted growing debate over whether AI enhances genuine language competence or merely creates an illusion of fluency. Recent studies and the OECD have emphasized both the pedagogical potential of AI for personalized learning and adaptive feedback and the methodological risks associated with learner overreliance on AI-generated output. The aim of this study is to determine how AI influences English language learning in comparison with traditional instruction and to identify the conditions under which AI supports authentic language development. The objectives include reviewing recent research on AI-supported EFL learning, comparing AI-assisted and conventional instruction, operationalizing the concept of the illusion of fluency. The study employed a comparative pedagogical design combining theoretical analysis, a quasi-experimental approach, pre- and post-testing, learner questionnaires, descriptive statistical analysis, and an independence-of-performance criterion. Forty A2–B1 undergraduate learners (aged 17–22) participated in a four-task instructional cycle. The experimental group used ChatGPT and similar chatbot-based tools for preparation, revision, and feedback, whereas the control group followed teacher-led instruction without AI support. The findings indicate that AI enhances selected aspects of language learning but may simultaneously inflate learners' perceptions of their own competence. Consequently, AI should be integrated within a staged blended-learning model in which it supplements, rather than replaces, independent language production and teacher guidance. Further research should investigate the illusion of fluency through larger longitudinal and cross-cultural studies, examine AI literacy as a moderating factor, refine AI-informed assessment practices, and explore AI-supported development of pronunciation and listening skills.

Keywords: *AI literacy, artificial intelligence, blended learning, ChatGPT, grammar-translation, English as a foreign language, illusion of fluency, language acquisition, learner autonomy, self-assessment, traditional methods.*

© Stoyka Olesia, Dupyn Yelyzaveta, 2026

INTRODUCTION

In recent years, the integration of Artificial Intelligence (AI) into educational practices has sparked considerable debate about its impact on language competency development. The central question emerges: does AI genuinely enhance linguistic proficiency, or does it merely create an illusion of fluency and success? This inquiry is vital, particularly when juxtaposing AI-driven learning with traditional educational methods. Initially, the discourse surrounding AI's role in education might suggest that it primarily addresses the «illusion of fluency». This notion implies that while AI tools can simulate language competence, they may not necessarily lead to a deep understanding or mastery of the language. As learners engage with AI, they might appear more proficient due to the immediate feedback and adaptive learning paths these technologies offer. However, this perceived fluency could be superficial if not grounded in a comprehensive understanding of the language.

Furthermore, the conversation often shifts to a comparison between AI and traditional teaching methodologies. Traditional methods, characterized by direct interaction with educators and peers, provide a structured environment for language acquisition. These settings encourage active participation, critical thinking, and personalized feedback, which are crucial for developing genuine language skills. Conversely, AI systems, while offering scalability and personalized learning experiences, may lack the nuanced understanding and emotional intelligence that human educators bring (Chauhan, 2025; Li & Long, 2025) Moreover, the concept of AI literacy has gained traction as an essential skill for navigating modern educational landscapes. Understanding how to effectively use AI tools becomes crucial for both students and educators. This literacy involves not only technical proficiency but also a critical awareness of the limitations and potential biases inherent in AI systems. As such, fostering AI literacy can empower learners to use these tools more effectively, ensuring that their language learning experiences are both meaningful and authentic. Additionally, the rise of blended learning models – where AI technologies complement traditional teaching – presents a promising avenue for optimizing language education. By integrating AI into classroom settings, educators can offer a more dynamic and flexible learning experience. This hybrid approach can harness the strengths of both AI and traditional methods, potentially leading to improved language outcomes. However, the research focus must remain clear and directed toward understanding whether AI truly enhances language competence or simply provides an illusion of success. This clarity is essential to align the study's objectives with its experimental components and to ensure that the investigation contributes valuable insights into the efficacy of AI in language education. In conclusion, while AI presents exciting possibilities for language learning, its role must be critically examined to determine its true impact on linguistic competence (Chauhan, 2025; Li & Long, 2025). By maintaining a focused research agenda on the question of AI's effectiveness compared to traditional teaching, we can better understand how these technologies can be leveraged to genuinely improve language skills rather than merely simulating proficiency.

Few branches of present-day applied linguistics have changed as quickly, or as unevenly, as the teaching of foreign languages. Over a remarkably short stretch of time, generative AI systems have moved from a research curiosity at the edge of the field to an everyday partner of the language learner. Chatbots paraphrase compositions within seconds, mobile applications correct pronunciation in real time, and adaptive platforms tailor grammar drills to the individual answer history of each user. For students of English this shift carries particular weight, since English remains the leading language of academic and professional communication and, as a result, the most heavily targeted by educational technology developers.

The pace of adoption, however, has run well ahead of the empirical scrutiny that should accompany it. While vendors and enthusiasts foreground gains in efficiency, motivation and individualisation, a growing number of teachers and researchers have begun to raise more uncomfortable questions. Does a corrected paragraph indicate what the student can do, or what the model can do on the student's behalf? Is rising confidence a sign of expanding competence, or merely a side-effect of working with a system that conceals difficulty? When a B1-level learner receives a C1-level paragraph after a single AI revision, what, exactly, has been learned?

© Stoyka Olesia, Dupyn Yelyzaveta, 2026

These questions stand at the centre of the present study. The notion of an "illusion of fluency" has recently entered the literature precisely in order to name the distance between AI-mediated output and the learner's own internalised system (Kumar et al., 2026). The illusion is not deception in any deliberate sense: learners and teachers alike are often genuinely persuaded by the linguistic surface of AI-supported work. The trouble is methodological. If language teachers continue to assess the product without examining the process that yielded it, they may unintentionally reward technological mediation while overlooking the very productive struggle that has traditionally been treated as constitutive of acquisition.

International comparative experiments have shown measurable gains in vocabulary acquisition and in writing performance when AI tools are deliberately built into instructional sequences (Chauhan, 2025; Li & Long, 2025). The question, therefore, is not whether AI should be used, but how to use it in a way that visible benefits do not mask hidden costs. The relevance of the topic is reinforced by the speed of institutional change. University writing centres, language departments and educational publishers are revising their guidance on AI use almost every term. Empirical work that brings together test outcomes, learner perceptions and unaided performance within a single design remains comparatively rare. The present article is intended to offer such an integrated picture. The scholarly conversation around AI in language education has, over the past several years, crystallised around three broad themes: personalisation and efficiency; the limits and risks of automated mediation; and the comparative effectiveness of AI-supported versus traditional instruction.

The Cambridge University Press white paper consolidates the research evidence on AI's three principal pedagogical roles in language learning: generation and revision of written text, simulation of conversational exchanges, and provision of explanatory feedback functionally equivalent to a private tutor (Macinska & Vinkler, 2024). A recent review in Language Learning & Technology confirms this typology and adds projections about the institutional integration of AI in formal language curricula (Zhu & Wang, 2025). A. Albedah's (2025). systematic review of AI in multilingual education highlights both pedagogical promise and ethical complications, including data privacy, algorithmic bias and unequal access. M. Dahbi (2023) describes intelligent language tutoring systems as a means of providing personalised instruction and feedback in English grammar learning. H. Shi and V. Aryadoust (2024) review research on AI-based automated written feedback and show that its effects and applications vary across contexts and dimensions of writing development. The cumulative weight of these sources establishes a sober but largely positive picture: AI works, but only under conditions that are explicitly designed for it.

K. R. Van Horn (2024) reports positive learner perceptions of ChatGPT in university EFL classes and links its use with metacognitive awareness and autonomous learning. These studies are pertinent to the present discussion because they remind us that students arrive in the classroom already carrying experience with AI-mediated language, which complicates the cleaner comparisons that controlled experiments tend to assume.

The most interesting recent developments, for the purposes of this article, concern the comparative effectiveness of AI-supported and traditional instruction. Li and Long report a controlled experiment at the college level in which AI-assisted English teaching produced measurable gains over traditional methods on several criterion measures (Li & Long, 2025). A. Hashim's (2025) comparative analysis of traditional and modern English language teaching argues that traditional approaches retain advantages in some dimensions of language formation. S. Chauhan (2025) similarly concludes that neither AI-based nor traditional teaching is sufficient on its own and advocates a hybrid or blended model that combines their respective strengths. OECD research also emphasizes both the opportunities and the equity-related challenges associated with AI in education (Varsik & Vosberg, 2024). The consistent message of this comparative work is that any uniform claim about superiority is methodologically suspect: outcomes depend on what is measured, how the intervention is structured, and which dimensions of competence are foregrounded.

The notion of the illusion of fluency has begun to receive direct empirical attention only recently. Kumar and colleagues examine the phenomenon of fluency illusion in AI-mediated learning environments and argue that fluent AI-generated responses can increase learners' confidence and

© Stoyka Olesia, Dupyn Yelyzaveta, 2026

perceived understanding without consistently producing deeper or transferable learning (Kumar et al., 2026). This account supports treating fluency illusion as a construct that warrants further operationalisation and empirical testing.

Despite this rich theoretical landscape, several gaps remain. First, much of the existing comparative work either reports test scores in isolation or examines learner perceptions in isolation, without bringing the two into direct relation. Second, the illusion of fluency, although now named in the literature, has rarely been operationalised in a way that would permit its prevalence to be estimated within a specific learner population. Third, the practical question of how AI and traditional methods should be sequenced within a single instructional cycle has received considerably less attention than the more abstract question of which is, in principle, superior. The empirical study reported below was designed to address these gaps within the constraints of a master's-level comparative project.

Goal and tasks of research. The goal of the present study is to determine in what way artificial intelligence influences English language learning in comparison with traditional teaching, and to identify the conditions under which AI mediation produces genuine progress rather than a polished surface of fluency without underlying competence.

The investigation proceeds through several connected tasks. **The first task** is to review the current scholarly conversation on AI in EFL and to distinguish between three thematic clusters: personalisation and efficiency, the limits and risks of automated mediation, and the comparative effectiveness of AI-supported and traditional instruction. **The second task** is to design a small comparative experiment in which AI-assisted and traditional learning formats are compared not only by test scores but also through the alignment between self-assessment and unaided performance. **The third task** is to interpret the results in a way that captures the asymmetric distribution of AI benefits across language components. **The fourth task** is to formulate methodological recommendations for a staged blended model in which AI is positioned as a revision and preparation resource. **The fifth task** is to outline directions for further research, especially in the operationalisation of the illusion of fluency.

RESEARCH METHODS

The empirical part of the work was organised as a comparative pedagogical inquiry that brought together features of a quasi-experimental design, a survey component and a process-oriented evaluation of unaided performance. The general aim was to compare AI-supported and conventional EFL instruction along three lines: measurable progress on a structured test, the learners' subjective experience of the two formats, and the correspondence between self-assessment and independent task performance. The scale of the project was deliberately modest, reflecting the constraints of a master's-level investigation, but the instruments were built so that conclusions could go beyond a simple comparison of means.

Forty undergraduate students of English took part in the study. Their ages ranged from seventeen to twenty-two, and their proficiency, established through a brief placement procedure, was situated between A2 and B1 on the Common European Framework of Reference. Participants were placed into two groups of equal size, designated as experimental and control. The experimental group worked through a four-task instructional cycle with permitted AI support at the stages of preparation, checking and revision. The control group completed the same four tasks under traditional conditions, namely teacher explanation, in-class drills, peer interaction and instructor feedback, without any AI involvement. The four tasks targeted, in order, contextual vocabulary, grammatical accuracy, opinion paragraph writing and a short oral mini-presentation.

AI use in the experimental group was constrained rather than free. Students were allowed to consult chatbot-style assistants (mostly ChatGPT, with smaller numbers using comparable tools) in order to obtain examples, explanations and comments on their own draft work, but they were not allowed to submit AI-generated text as their own product. For instance, on the vocabulary task connected with the topic "Education and technology", students could request a list of ten B1-level items, choose a subset of six, and produce their own example sentences, while marking which of their formulations had been adapted from AI suggestions and which were original. The pedagogical reason

© Stoyka Olesia, Dupyn Yelyzaveta, 2026

for this restriction was the verdict, supported by recent bursary, that the educational value of AI is established in its mediation of learning rather than in its substitution for it.

Three categories of instrument were applied. The first was a structured pre-test and post-test administered to both groups. The assessment had four components (grammar, vocabulary, reading comprehension and writing) and a total of fifty points, distributed as ten, ten, ten and twenty respectively. The pre-test and post-test were not identical, but they were parallel in structure and difficulty, with grammar items covering tense, comparison, passive voice, prepositions, reported speech and modality, and writing items requiring an opinion paragraph of between one hundred and twenty and one hundred and fifty words. The second instrument was an anonymous questionnaire that gathered information about the frequency of AI use, preferred tools, typical purposes, perceived benefits and risks, ability to complete comparable tasks without AI, and preference for one or another learning format. The questionnaire combined closed items, Likert-scale statements and a single open-ended response. The third instrument was a process-oriented evaluation of unaided performance, in which participants, after completing an AI-supported task, rated their readiness to perform an analogous task without digital assistance on a five-point scale, and were then required to perform that analogous task under controlled conditions. The actual performance was scored on the same five-point scale and compared with the self-rating.

Written and oral tasks were graded according to comparable five-criterion rubrics: task achievement, grammar accuracy, vocabulary range, coherence and logic, and independence of performance for written work; fluency, pronunciation, grammar and vocabulary, response to questions, and independence of response for oral work. The independence criterion was particularly important, since it formalised the intuition that a grammatically correct AI-mediated product does not, in itself, constitute evidence of competence on the part of the learner who submitted it.

The analytic procedure was correspondingly mixed. Test scores were compared at the component and aggregate levels through descriptive statistics, with gains computed as the difference between pre-test and post-test means. Questionnaire data were tabulated as relative frequencies. The self-assessment versus independent-performance comparison generated a “gap score” for each participant, defined as the difference between the self-rated readiness and the actual performance score, both expressed on the same five-point scale. A gap of zero was treated as accurate self-assessment, a gap of ones as moderate overestimation, and a gap of two or more as a pronounced illusion of fluency. Group means and the proportion of participants showing a pronounced illusion were then compared across the experimental and control groups.

Several methodological reservations should be acknowledged. The sample is small and drawn from a single institution, which limits the generalisability of the quantitative results. The four-task cycle is fairly short and does not capture longer-term retention. The placement of students into groups was based on practical scheduling rather than full randomisation, although the pre-test results show that the two groups began the intervention at an almost identical level. These reservations are addressed further in the discussion. They do not, however, undermine the central comparative claims of the study, which rest on the relational structure of the design rather than on absolute population estimates.

A brief note on the use of AI in the preparation of this manuscript: the author consulted AI-based tools (ChatGPT) for grammar checking and language polishing of the English text. All conceptual content, study design, data collection, analysis and interpretation are the author’s own. No AI-generated passages were incorporated into the manuscript without substantive revision by the author.

RESEARCH RESULTS

The results are reported in four subsections matching the principal analytic foci of the study: aggregate test outcomes, component-level progress, questionnaire findings, and the comparison between self-assessment and unaided performance.

Aggregate pre-test scores were nearly identical across the two groups. The experimental group averaged 31.4 points and the control group averaged 31.1 points out of a possible 50. The 0.3-point initial difference is too small to indicate a meaningful baseline disparity. After the four-task instructional

© Stoyka Olesia, Dupyn Yelyzaveta, 2026

cycle, the experimental group reached an aggregate post-test mean of 40.6 points, while the control group reached 38.2 points. The gain in the experimental group was therefore 9.2 points, against 7.1 points in the control group, producing a between-group gain difference of 2.1 points in favour of AI-supported instruction. Table 1 reports the aggregate values.

Table 1

Aggregate pre-test and post-test results

Group	Pre-test (mean, /50)	Post-test (mean, /50)	Gain
Experimental (n = 20)	31.4	40.6	+9.2
Control (n = 20)	31.1	38.2	+7.1
Between-group difference	0.3	2.4	2.1

Source: compiled by the author based on the empirical data of the study

Disaggregating the totals by component reveals a more textured picture. The largest absolute progress in both groups was registered in writing, which is unsurprising given that the writing component carried a 20-point weight and the instructional cycle included two writing-related tasks. The experimental group's writing score rose from 12.2 to 16.2 points (a 4.0-point gain), while the control group's writing score rose from 12.0 to 15.0 points (a 3.0-point gain). The comparative advantage of AI support in writing was therefore 1.0 point. Vocabulary followed a similar pattern, with the experimental group gaining 1.9 points and the control group 1.4 points, yielding a comparative advantage of 0.5 points. Grammar produced a smaller differential: the experimental group gained 1.9 points against the control group's 1.6, while reading showed the smallest difference, with gains of 1.4 and 1.1 points respectively. Table 2 sets out the full component-level breakdown.

Table 2

Component-level progress in the experimental and control groups

Component (max)	Exp. pre	Exp. post	Exp. gain	Ctrl. gain	Difference
Grammar (10)	6.1	8.0	+1.9	+1.6	+0.3
Vocabulary (10)	6.4	8.3	+1.9	+1.4	+0.5
Reading (10)	6.7	8.1	+1.4	+1.1	+0.3
Writing (20)	12.2	16.2	+4.0	+3.0	+1.0
Total (50)	**31.4**	**40.6**	**+9.2**	**+7.1**	**+2.1**

Source: compiled by the author based on the empirical data of the study

The pattern of component-level gains is consistent with the nature of the AI mediation used in the experimental group. Writing benefited the most because AI tools were employed precisely in order to provide examples, paraphrases and revision suggestions for written drafts. Vocabulary benefited noticeably because AI was deployed to generate topic-appropriate lexical items with example contexts. Grammar benefited less because the post-test grammar tasks required active recognition of correct forms under closed-test conditions, where prior AI mediation has less direct purchase. Reading benefited least because comprehension of an unseen text depends on cognitive strategies that are not readily transferred from AI-mediated revision activities.

Questionnaire responses provide an important corroborative perspective on the test-score outcomes. According to participants' self-reports, 68% used AI tools several times a week or more often, indicating that the technology has already become a routine part of their study practice. Chatbot-style assistants, predominantly ChatGPT, were the most frequently mentioned tool, identified by 57% of respondents. The most commonly reported purpose of AI use was the checking of written

© Stoyka Olesia, Dupyn Yelyzaveta, 2026

assignments, mentioned by 62% of participants, followed by vocabulary expansion and grammar clarification. Seventy per cent of respondents agreed that AI helps them grasp grammar more rapidly than traditional explanation alone, a perception that aligns with the modest grammar advantage observed in the experimental group's test scores.

The questionnaire also surfaced a less reassuring pattern. Forty-eight per cent of respondents acknowledged the risk of becoming dependent on ready-made answers, and 35% reported difficulty in completing analogous tasks without AI assistance. These figures matter for the interpretation of the test gains. AI-mediated practice may improve the polished surface of written output without proportionally strengthening the learner's capacity to produce comparable output without mediation. The questionnaire thus offers a bridge between the aggregate gains reported above and the self-assessment analysis that follows.

The comparison between self-assessment and unaided performance constitutes the central analytic contribution of the study. In the experimental group, the mean self-rated readiness, on a five-point scale, was 4.3, while the median actual independent performance, on the same scale, was 3.3. The median gap was therefore 1.0 point, with a pronounced illusion of fluency (gap of 2 or more) identified in 35% of participants. In the control group, the median self-rated readiness was 3.6, while the mean actual independent performance was 3.5. The median gap was only 0.1 point, with a pronounced illusion of fluency identified in 10% of participants. Table 3 reports these comparisons.

Table 3

Self-assessment compared with actual unaided performance

Group	Self-rated readiness (1 to 5)	Actual performance (1 to 5)	Median gap	Pronounced illusion (gap >= 2)
Experimental	4.3	3.3	1.0	35%
Control	3.6	3.5	0.1	10%

Source: compiled by the author based on the empirical data of the study

The contrast between the two groups on this dimension is the most consequential finding of the study. While the AI-supported group outperformed the control group on the aggregate test, it simultaneously overestimated its own independent capability to a degree more than ten times greater than the control group. In other words, AI-mediated practice produced both real progress and a measurable illusion of fluency, and these two effects coexisted in the same population. The control group's progress was more modest in absolute terms, but it was matched by a self-assessment that closely tracked actual performance, which suggests that learners who worked under traditional conditions developed a clearer awareness of their own level.

A qualitative reading of the open-ended questionnaire responses reinforces the quantitative findings. Several participants in the experimental group described feeling more capable after an AI-supported task and then encountering unexpected difficulty when asked to perform an analogous task unaided. For instance, one respondent reported confidence in writing about a familiar topic after AI revision, but described difficulty with sentence structure and transitions when writing on the same topic without AI on a subsequent occasion. The phenomenon is not universal: many AI users completed independent tasks with little degradation. It is, however, sufficiently widespread to deserve attention as a systematic risk rather than an isolated anomaly.

DISCUSSION

The findings of the present study converge with, and in certain respects extend, the existing literature on AI in EFL. The aggregate test advantage of the AI-supported group is consistent with the comparative results of C. Li and J. Long (2025) and with the meta-analytic suggestion in C. Dhanapal and colleagues (2024) report on learners' perceptions of AI-based and conventional language instruction and discuss the potential benefits and limitations of both approaches. The pattern of component-level

© Stoyka Olesia, Dupyn Yelyzaveta, 2026

advantage, strongest in writing and vocabulary, moderate in grammar, weakest in reading, is broadly consistent with previous accounts of AI's affordances and limitations (Macinska & Vinkler, 2024; Zhu & Wang, 2025). The novel contribution of this study, however, lies in the simultaneous demonstration that aggregate gains and an illusion of fluency can coexist within the same population of learners. This finding extends the concern articulated by S. Kumar and colleagues (2026) that fluent AI output may strengthen perceived understanding without guaranteeing deeper learning.

A first interpretive point concerns the asymmetric distribution of AI benefit across language components. Writing and vocabulary benefited substantially because AI tools can function as drafting assistants and sources of personalised feedback, particularly in writing and grammar learning (Macinska & Vinkler, 2024; Dahbi, 2023). Grammar showed a more modest advantage because grammatical competence has a productive dimension that does not transfer automatically from AI-mediated revision to closed-test recognition. Reading benefited least because comprehension of a previously unseen text mobilises cognitive strategies that AI exposure does not directly cultivate. The implication for EFL methodology is that AI's contribution is not uniform: instructional design should target AI mediation to the components where it is most productive and rely on traditional approaches where its leverage is weakest.

A second point concerns the relationship between the questionnaire data and the self-assessment results. The fact that 70% of respondents perceive AI as accelerating their grammatical understanding is broadly consistent with the modest grammar advantage observed in the test, yet it sits in tension with the much larger self-assessment gap. One plausible explanation is that AI use creates a kind of cognitive smoothing: when an explanation arrives quickly and is phrased clearly, learners experience comprehension as immediate, and that experience of immediacy is mistaken for stable acquisition. The outcome is a perception of competence that is more confident than the underlying competence warrants. This is consistent with the broader cognitive-psychological literature on fluency effects in metacognition, which holds that processing fluency is systematically misattributed to depth of knowledge.

A third point concerns the control group's accurate self-assessment. The 0.1-point gap observed in the control group is striking not because it is small in absolute terms but because it suggests that traditional methods, often criticised for their slowness, may have a hidden benefit that has been undertheorised in recent debates (Hashim, 2025). The slower pace of teacher explanation, peer interaction and written feedback appears to give learners more accurate information about their own level. Each error met in a controlled exercise, each misstep corrected by a peer, each pause for clarification with the instructor functions as a calibration signal that helps the learner construct a more realistic model of their own competence. AI mediation, in contrast, tends to suppress these signals: corrections appear instantly, errors are minimised in the visible output, and the learner has less occasion to register the gap between intention and execution.

A fourth point concerns assessment design. The 50-point pre/post-test, taken on its own, would have led to a straightforward conclusion in favour of AI-supported instruction. The 2.1-point aggregate gain advantage is real, measured under controlled conditions, and unlikely to be an artefact of measurement error. Yet the same controlled design would have rendered the illusion of fluency invisible. The implication for EFL assessment is significant: traditional product-based assessment, even when carefully constructed, is no longer sufficient under conditions of widespread AI use. Process-oriented elements (explanation of corrections, oral defense of one's own text, controlled in-class tasks without prompts, comparison between drafts) are required to triangulate the product-based score and to detect the illusion of fluency where it occurs.

A fifth point concerns the integration of AI and traditional methods. The findings of the present study do not support a polarised choice between the two modes. The AI-supported group outperformed the control group in aggregate, while the control group displayed more accurate self-assessment. These two profiles are complementary rather than mutually exclusive. The most defensible pedagogical model is therefore one in which AI mediation and traditional pedagogy are sequenced so that each compensates for the limitations of the other (Chauhan, 2025; Varsik & Vosberg, 2024). A sequence in which the teacher first explains the target material, learners attempt an initial independent task, AI is used at the

© Stoyka Olesia, Dupyn Yelyzaveta, 2026

revision stage, and a final controlled task is performed without prompts would, in principle, preserve the productivity gains of AI while restoring the calibration signals that traditional methods provide.

A sixth point concerns the role of AI literacy. The illusion of fluency is not an inherent property of AI tools: it is a property of the way in which learners use them. A learner who can articulate the difference between recognition and production, who can distinguish their own contribution from the tool's, and who can submit themselves to occasional unaided checking is unlikely to develop a pronounced illusion. By contrast, a learner who uses AI as a form of simulation, accepting all corrections without analysis, is at much higher risk. This implies that AI literacy should be cultivated explicitly within EFL curricula, not as an optional add-on but as a fundamental component of communicative competence in the digital age.

Several limitations should be acknowledged. The sample of forty learners is modest and recruited from a single institution, which limits generalizability. The four-task instructional cycle, while internally coherent, is too short to capture longer-term retention or transfer to genuinely novel communicative situations. The placement of participants into groups was non-random, although the near-identical pre-test scores mitigate this concern. The reliance on self-report for AI use frequency and perceived benefit introduces the usual risks of socially desirable responding. Finally, the operationalization of the illusion of fluency through a single self-assessment versus independent-performance comparison is a first approximation: future work should explore alternative operationalization and combine them within a more comprehensive measurement model (Albedah, 2025).

Notwithstanding these limitations, the results provide a coherent empirical basis for the methodological recommendations that follow. They suggest that AI is neither the panacea its enthusiasts claim nor the threat its critics fear, but a complex pedagogical resource whose value depends on how it is embedded in instructional design. The polarised framing of "AI versus traditional methods" obscures more than it illuminates: the relevant question is how the two modes should be combined in such a way that real progress is supported and the illusion of fluency is actively counteracted.

The practical translation of these results into classroom protocols deserves a brief separate discussion. The recommended sequence for vocabulary work begins with the teacher introducing a thematic frame, followed by an initial elicitation in which the learner attempts to list known items. Only after this initial step should AI be consulted, and the consultation itself should be structured: the learner asks for ten B1-level items connected with the topic, selects a subset of six, and produces original example sentences, marking which formulations were adapted from AI suggestions and which were independent. A final unaided task, performed in class without digital prompts, completes the cycle. This protocol preserves the lexical breadth that AI offers while preventing passive copying.

Grammar work calls for a similar but more conservative structure. AI should not be the first port of call when a new structure is introduced. The teacher explains the structure, the learner attempts a small number of controlled exercises, and only then is AI consulted, typically to clarify two or three points that remain unclear. The closed-test recognition tasks that follow should be performed without prompts. This sequencing recognises the asymmetry observed in the experimental data: AI mediation produces only a modest grammar advantage, and the productive dimension of grammatical competence does not transfer automatically from AI-mediated revision to test performance.

For writing, the most defensible model is "draft, comment, independent revision". The learner produces a 120 to 150-word paragraph without external assistance. AI is then asked to comment on grammar and logic, but not to rewrite the text. The learner reads the comments, decides which to incorporate, and revises the paragraph independently. A final assessment task is performed under controlled conditions, without prompts. For speaking, AI use should be restricted to preparatory work: the generation of five key words or a short outline. The two to three-minute mini-presentation and the answers to follow-up questions should be performed without prompts, and the assessment criteria should explicitly reward independence of response.

Across all four protocols, the central principle is that AI mediation should be sandwiched between two unaided performances: an initial attempt that surfaces the learner's starting point and a final task that confirms transfer to independent work.

© Stoyka Olesia, Dupyn Yelyzaveta, 2026

CONCLUSIONS

The present article has examined the impact of artificial intelligence on English language learning through a combination of theoretical synthesis and a small-scale comparative experiment. Three conclusions follow from the analysis.

First, AI-supported instruction can produce measurable gains in English language competence, particularly in writing and vocabulary, where its mediation aligns closely with the cognitive demands of the relevant tasks. The 9.2-point gain registered in the experimental group, against a 7.1-point gain in the traditional control group, is not negligible and supports the integration of AI into EFL curricula, provided that this integration is purposeful and constrained.

Second, AI-supported instruction simultaneously produces an illusion of fluency in a non-trivial proportion of learners. The 1.0-point gap between self-assessment and unaided performance in the experimental group, set against the 0.1-point gap in the control group, and the 35% prevalence of pronounced illusion against 10% in the control group, show that AI mediation systematically inflates learners' perceptions of their own competence. This pattern is not an accident of the present study: it is consistent with recent theoretical accounts and with broader cognitive-psychological findings on fluency effects in metacognition.

Third, these two findings together support a staged, blended methodology rather than a polarised choice between AI and traditional instruction. The most defensible pedagogical sequence is one in which the teacher explains the target material, the learner attempts an initial independent task, AI is used at the revision stage under explicit guidelines, and a final controlled task is performed without prompts. Within this sequence, written tasks should employ a "draft, comment, independent revision" model, oral tasks should restrict AI to preparatory work, and assessment should be process-oriented as well as product-oriented, with mechanisms in place to detect the illusion of fluency where it occurs.

The broader implication of the study is that the integration of AI into language education is neither inevitable progress nor inevitable decline. It is a methodological project whose outcome depends on the conceptual clarity, pedagogical discipline and ethical responsibility with which it is undertaken. Real progress in English language learning, in the AI era as before, requires more than visible polish: it requires the slow, often inelegant work through which a learner internalises a foreign language as their own. The task of EFL methodology, accordingly, is to ensure that the new tools are deployed in support of that work rather than in substitution for it.

Prospects for further research. Several directions for further research follow naturally from the conclusions of this study. The illusion of fluency deserves more sophisticated operationalisation, ideally through multi-method designs that combine self-assessment, unaided performance, expert judgement and possibly process tracing through keystroke logging or screen recording. The optimal sequencing of AI and traditional methods should be tested through longer interventions with retention assessments at multiple intervals. The role of AI literacy as a moderator of fluency-illusion susceptibility should be examined in controlled studies. The implications of AI mediation for assessment design should be addressed at the institutional level, with attention to questions of academic integrity, equity of access and the validity of widely used assessment genres under conditions of pervasive AI use. Finally, it would be productive to extend the present comparative framework to other linguistic levels, including pronunciation and listening, where the affordances of AI tools differ markedly from those examined here. Cross-cultural comparison of AI use in EFL contexts, including the role of learner first language and cultural attitudes to corrective feedback, also offers a promising direction for further work.

References

Albedah, A. (2025). Artificial Intelligence in Language Education: A Systematic Review of Multilingual Applications, Large Language Models, and Emerging Challenges. *Language Teaching Research Quarterly,* 49, 247–268. https://doi.org/10.32038/ltrq.2025.49.13

© Stoyka Olesia, Dupyn Yelyzaveta, 2026

Chauhan, S. (2025). AI vs traditional teaching: A comparative analysis across key dimensions. *Bhartiyam International Journal of Education & Research,* 14 (3), 58–63. https://www.gangainstituteofeducation.com/wp-content/uploads/2025/07/5.-AI-vs-Traditional-Teaching.pdf

Dahbi, M. (2023). Integrating an intelligent language tutoring system in teaching English grammar. *Arab World English Journal,* 14 (4), 189–196. https://doi.org/10.24093/awej/vol14no4.11

Dhanapal, C., Asharudeen, N., & Alfaruque, S. Y. (2024). Impact of artificial intelligence versus traditional instruction for language learning: A survey. *World Journal of English Language,* 14 (2), 182–193. https://doi.org/10.5430/wjel.v14n2p182

Hashim, A. A. (2025). Innovative approaches in English language teaching: A comparative analysis of traditional and modern methods. *South Asian Research Journal of Arts, Language and Literature,* 7 (1), 1–7. https://doi.org/10.36346/sarjall.2025.v07i01.001

Kovalenko, I., & Baranivska, N. (2024). The impact of social media on foreign language learning: Opportunities for interactive education. *Actual Issues in Modern Science,* 7 (25), 573–582. https://doi.org/10.52058/2786-6300-2024-7(25)-573-582

Kumar, S., Mikayelyan, A., & Vorfolomeyeva, O. (2026). Fluency illusion: A review on influence of ChatGPT in classroom settings. *Information,* 17 (3), 299. https://doi.org/10.3390/info17030299

Li, C., & Long, J. (2025). Comparing AI-assisted and traditional teaching in college English: Pedagogical benefits and learning behaviors. Information, 16 (10), 895. https://doi.org/10.3390/info16100895

Macinska, S., & Vinkler, A. (2024). *The role of AI in language learning: Research evidence and strategies for success. Cambridge Papers in English Language Education.* Cambridge University Press & Assessment. https://www.cambridge.org/sites/default/files/media/documents/The_Role_of_AI_in_Language_Learning_Research_Evidence_and_Strategies.pdf

OECD (2024). *Education Policy Outlook 2024: Reshaping Teaching into a Thriving Profession from ABCs to AI.* https://doi.org/10.1787/dd5140e4-en

Shi, H., & Aryadoust, V. (2024). A systematic review of AI-based automated written feedback research. *ReCALL,* 36 (2), 187–209. https://doi.org/10.1017/S0958344023000265

Van Horn, K. R. (2024). ChatGPT in English language learning: Exploring perceptions and promoting autonomy in a university EFL context. *TESL-EJ,* 28 (1). https://doi.org/10.55593/ej.28109a8

Varsik, S., & Vosberg, L. (2024). The potential impact of artificial intelligence on equity and inclusion in education. *OECD Artificial Intelligence Papers,* 23. https://doi.org/10.1787/15df715b-en

Zhu, M., & Wang, C. (2025). A systematic review of research on AI in language education: Current status and future implications. *Language Learning & Technology,* 29 (1), 1–29. https://doi.org/10.64152/10125/73606

Received 2.06.2026
Accepted 29.10.2026
Published 30.10.2026

ВПЛИВ ШТУЧНОГО ІНТЕЛЕКТУ НА ВИВЧЕННЯ МОВИ: МІЖ РЕАЛЬНИМ ПРОГРЕСОМ ТА ІЛЮЗІЄЮ ВІЛЬНОСТІ ВОЛОДІННЯ МОВОЮ

Олеся Стойка
https://orcid.org/0000-0002-7695-6100
доктор педагогічних наук, професор,
професор кафедри прикладної лінгвістики,
Ужгородський національний університет,
вул. Університетська, 14, 88000, Ужгород, Україна;
головний науковий співробітник
відділу порівняльної педагогіки,
Інститут педагогіки НАПН України,
вул. Артема, 52-Д, 04053, Київ, Україна
olesya.stoyka@uzhnu.edu.ua

© Stoyka Olesia, Dupyn Yelyzaveta, 2026

Єлизавета Дупин
https://orcid.org/0009-0005-0061-2490
магістрантка кафедри прикладної лінгвістики,
факультет міжнародних економічних відносин,
Ужгородський національний університет,
вул. Університетська, 14, 88000, Ужгород, Україна;
dupyn.yelyzaveta@student.uzhnu.edu.ua

Стрімка інтеграція генеративного штучного інтелекту (ШІ) у навчання англійської мови як іноземної (EFL) актуалізувала наукову дискусію щодо того, чи сприяє ШІ реальному розвитку іншомовної компетентності, чи лише створює ілюзію вільного володіння мовою. У сучасних дослідженнях, а також у документах ОЕСР наголошується як на педагогічному потенціалі ШІ для персоналізованого навчання та адаптивного зворотного зв'язку, так і на методологічних ризиках, пов'язаних із надмірною залежністю здобувачів освіти від контенту, згенерованого штучним інтелектом. Метою дослідження є визначення впливу штучного інтелекту на навчання англійської мови порівняно з традиційними методами викладання та встановлення умов, за яких використання ШІ сприяє справжньому розвитку мовної компетентності. Завдання дослідження передбачали аналіз сучасних наукових праць, присвячених застосуванню ШІ у навчанні EFL, порівняння AI-підтримуваного та традиційного навчання, обґрунтування поняття ілюзії вільного володіння мовою. У дослідженні застосовано порівняльний педагогічний дизайн, що поєднав теоретичний аналіз, квазіекспериментальний підхід, до- та післятестування, анкетування студентів, методи описової статистики та критерій виконання завдань самостійно. У дослідженні взяли участь 40 студентів бакалаврату рівня володіння англійською мовою A2–B1 віком 17–22 роки. Експериментальна група використовувала ChatGPT та подібні чат-боти на етапах підготовки, перевірки й доопрацювання завдань, тоді як контрольна група навчалася за традиційною методикою під керівництвом викладача без використання ШІ. Отримані результати свідчать, що ШІ здатний підвищувати ефективність окремих аспектів навчання мови, проте одночасно може формувати завищене уявлення здобувачів освіти про власний рівень мовної компетентності. У зв'язку з цим використання ШІ доцільно впроваджувати в межах поетапної моделі змішаного навчання, у якій він доповнює, а не замінює самостійну мовленнєву діяльність студентів і педагогічний супровід викладача. Перспективи подальших досліджень пов'язані з вивченням феномену ілюзії вільного володіння мовою на основі масштабніших тривалих і міжкультурних досліджень, аналізом ролі ШІ-грамотності як чинника, що впливає на її формування, удосконаленням підходів до оцінювання результатів навчання в умовах використання ШІ, а також дослідженням можливостей застосування штучного інтелекту для розвитку вимови та навичок аудіювання.

Ключові слова: *автономія учня, англійська як іноземна мова, засвоєння мови, змішане навчання, ілюзія мовної компетентності, самооцінка, традиційні методи, ШІ-грамотність, штучний інтелект, ChatGPT.*

© Stoyka Olesia, Dupyn Yelyzaveta, 2026