AI-Integrated Model for Developing Professional Foreign Language Communicative Competence: An Experimental Study with Law Students

Gulnara Kassymova[1], Madina Samatkyzy[2*], Arailym Beikutova[3]
[1]SDU University, Kazakhstan, [2]Al-Farabi Kazakh National University, Kazakhstan, [3]Kazakh Ablai Khan University of International Relations and World Languages, Kazakhstan

Abstract The integration of artificial intelligence (AI) into higher education has created new opportunities for professionally oriented foreign language instruction requiring specialized linguistic and professional competencies. This study aimed to develop and evaluate an AI-integrated instructional model for enhancing Professional Foreign Language Communicative Competence (PFLCC) among law students and to explore their experiences of AI-supported learning. A mixed-methods quasi-experimental design compared students receiving AI-integrated instruction with those receiving conventional instruction. The model comprised five interconnected components: Linguistic-Professional, Discourse-Genre, Intercultural Professional, Cognitive-Critical and AI-Supported Professional Interaction. The groups demonstrated comparable PFLCC performance before the intervention. Following the intervention, students receiving AI-integrated instruction demonstrated significantly higher overall post-test PFLCC performance and a stronger descriptive shift toward higher competence levels. Qualitative analysis identified professional communication practice, feedback and revision, and critical engagement with AI-generated content as recurrent themes. The findings suggest that combining AI-supported activities with professionally situated legal tasks, critical evaluation, revision, and teacher guidance can effectively support PFLCC development among law students.

Keywords: *Professional foreign language, Technology-enhanced learning, AI-assisted language learning, AI-integrated model, Law students*

[*]**Corresponding Author:**
Madina Samatkyzy
samatkyzy.madina@mail.ru

Received: June 1, 2026
Revised: September 9, 2026
Accepted: October 1, 2026
First Published: October 24, 2026

© 2026 Kassymova, Samatkyzy, and Beikutova.
This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY).

1. Introduction

The rapid digital transformation of higher education and the increasing integration of artificial intelligence (AI) are reshaping foreign language teaching and learning, particularly in professionally oriented education. In university contexts, foreign language competence is increasingly expected to extend beyond general linguistic proficiency and enable students to communicate effectively in discipline-specific academic and professional situations. This requirement is particularly relevant to legal education, where students need to use professional terminology, construct and

https://doi.org/10.22034/ijscl.2026.2090234.4737

evaluate arguments, explain legal issues to different audiences, participate in professional interaction, and communicate with appropriate discourse, pragmatic, and intercultural awareness. AI-supported systems can provide personalized practice, immediate feedback, conversational interaction, and simulations of communicative situations that may be difficult to reproduce repeatedly in conventional classrooms. Recent research has demonstrated the potential of AI and generative AI to support multiple aspects of language learning, including speaking, writing, reading, learner engagement, self-regulation, and communicative competence (Crompton et al., 2024; Du & Daniel, 2024; Karatash et al., 2024; Kohnke et al., 2023; Zhumatayeva et al., 2024).

Within the Kazakhstani higher education context, the development of foreign language communicative competence is also closely connected with the sociocultural environment in which language learning occurs. Recent research with university students in Kazakhstan has shown that integrating locally relevant cultural content into EFL instruction can support communicative development and strengthen the contextual relevance of language learning (Tazhitova et al., 2025). This highlights the importance of designing professionally oriented foreign language instruction that responds not only to linguistic objectives but also to the specific educational and sociocultural context of learners. These developments create a need for instructional approaches that connect AI-supported learning with the linguistic, professional, critical, and sociocultural requirements of students' future professional communication. Existing research has frequently examined AI-supported writing, speaking, feedback, engagement, or self-regulated learning as separate outcomes, whereas less attention has been given to multidimensional professional communicative competence in which linguistic accuracy, professional discourse, intercultural appropriateness, critical reasoning, and responsible AI-supported interaction are developed within a unified instructional model. This gap is particularly relevant in legal education, where effective professional communication requires not only accurate language use but also genre awareness, audience adaptation, argumentation, critical evaluation of information, and responsible engagement with AI-generated content. Furthermore, there is limited empirical evidence on how an AI-integrated model of this kind operates with law students in the Kazakhstani higher education context. There is therefore a need not only to develop such a model but also to evaluate its effectiveness through a controlled instructional intervention and to examine students' experiences of AI-supported professional language learning through complementary qualitative evidence.

Accordingly, the aims of this study are to develop an AI-integrated methodological model for enhancing Professional Foreign Language Communicative Competence (PFLCC) among law students in Kazakhstan; to evaluate its effectiveness through a 10-week quasi-experimental intervention comparing an experimental group with a control group; and to explore students' experiences of AI-supported professional language learning through reflective self-assessments, classroom observations, and student feedback. In particular, the study addresses the following research questions:

1. To what extent does PFLCC change among law students receiving the AI-integrated intervention from pre-test to post-test?
2. How does the distribution of students across the four PFLCC competence levels differ between the pre-test and post-test, and between the experimental and control groups at post-test?
3. Are there statistically significant differences in post-test PFLCC performance between law students receiving AI-integrated instruction and those receiving conventional foreign language instruction?
4. How do students experience and perceive AI-supported professional foreign language learning during the intervention, as reflected in their self-assessments, classroom observations, and feedback?

2. Theoretical Framework

The literature relevant to the present study can be organized around two central variables: PFLCC as the target educational outcome and AI-integrated instruction as the pedagogical condition through which this competence may be developed. A third area concerns the relationship between these two variables in professionally oriented legal education and provides the basis for identifying the research gap and proposing the model examined in this study.

IN PRESS

2.1. Professional Foreign Language Communicative Competence as the Target Construct

The competence-based approach provides the general theoretical foundation for PFLCC. The conceptualization of communicative competence extends language learning beyond grammatical knowledge to the ability to use language appropriately in socially situated communication (Hymes, 1972). This perspective was further developed by Canale and Swain (1980), who emphasized the multidimensional nature of communicative competence. Rather than conceptualizing foreign language learning as the acquisition of isolated grammatical and lexical knowledge, competence-oriented language education emphasizes learners' ability to mobilize linguistic resources to accomplish meaningful communicative purposes (Richards & Rodgers, 2014). From this perspective, successful foreign language learning cannot be reduced to mastery of vocabulary and grammar; it also involves the ability to select and organize linguistic resources according to communicative purpose, audience, and context.

The conceptualization of PFLCC can also be informed by established perspectives on professionally oriented, reflective, and experience-based learning. In English for Specific Purposes, Hutchinson and Waters (1987) emphasized that language instruction should be shaped by learners' specific communicative needs and the contexts in which the target language is expected to be used. This perspective is particularly relevant to legal education, where language learning needs to be connected with professional genres, terminology, communicative purposes, and situations that students may encounter in future practice. Swales (1990) further demonstrated the importance of genre and discourse conventions in academic and professional communication, supporting the inclusion of discourse–genre competence as a distinct dimension of PFLCC. Professional competence also involves reflection, evaluation, and the ability to reconsider one's communicative choices. Dewey (1933) conceptualized reflective thinking as an active process of examining knowledge and the grounds on which it is based, while Schön (1983) emphasized reflection as an important feature of professional practice and decision-making. Mezirow (1991) subsequently highlighted critical reflection as a process through which learners examine and revise their assumptions and interpretations. These perspectives are relevant to the cognitive–critical dimension of the present PFLCC model because students are expected not only to produce professional language but also to evaluate information, justify communicative decisions, identify problematic output, and revise their responses. In an AI-integrated learning environment, such reflective and critical processes become particularly important when students evaluate, verify, and revise AI-generated professional communication rather than accepting generated content uncritically.

Professional communication is also multidimensional. Linguistic accuracy and knowledge of professional terminology are necessary but insufficient when communication requires students to structure information coherently, follow genre conventions, adapt language to interlocutors, evaluate arguments, and make contextually appropriate communicative choices. Intercultural competence is particularly relevant when professional communication involves participants with different linguistic and cultural backgrounds. Research on Kazakh–English communication, for example, has emphasized the role of cultural specificity in transferring meanings across languages and communicative contexts (Mukhtarova et al., 2025). Within the multilingual and multicultural context of Kazakhstan, professionally oriented English instruction therefore needs to address not only language form but also discourse, pragmatic appropriateness, cultural awareness, and the ability to communicate effectively across professional contexts.

For the purposes of the present study, these strands of research support a multidimensional understanding of PFLCC encompassing linguistic-professional, discourse-genre, intercultural professional, and cognitive-critical capacities. The increasing use of AI in professional and educational communication adds a further dimension: students must also be able to engage with AI-supported resources purposefully and critically rather than treating generated content as automatically reliable.

2.2. AI as a Pedagogical Mediator of Professional Communication

The second theoretical dimension concerns the role of AI in language learning. The development of AI in education has expanded opportunities for adaptive learning, individualized support, feedback, and learner interaction (Holmes et al., 2019; Luckin et al., 2016). Research in higher education has further

emphasized that the educational value of AI depends not only on technological capabilities but also on pedagogically informed implementation and the relationships established among learners, teachers, and intelligent systems (Ouyang & Jiao, 2021; Zawacki-Richter et al., 2019). This distinction is important because access to an AI tool does not in itself constitute an instructional model; its educational contribution depends on how it is incorporated into learning objectives, tasks, feedback, and teacher guidance.

In language education specifically, AI-supported environments can facilitate personalized practice, immediate feedback, conversational interaction, language production, and learner self-regulation (Crompton et al., 2024; Godwin-Jones, 2018; Huang et al., 2022). Generative AI further enables learners to engage with dynamically produced texts, explanations, scenarios, and simulated dialogue, thereby expanding opportunities for repeated language practice and interaction (Kohnke et al., 2023; Lo et al., 2024). Such affordances are potentially relevant to professionally oriented language education because professional communicative situations can be simulated repeatedly and adapted to different purposes, levels of complexity, and interlocutors. In legal language education, for example, AI-supported practice may involve simulated client consultations, professional correspondence, case-based discussion, argumentation, or comparison of alternative formulations.

At the same time, recent literature cautions against equating increased access to generated language with improved communicative competence. Large language models may produce inaccurate, biased, or contextually inappropriate information, making critical evaluation an essential part of AI-supported learning (Kasneci et al., 2023; Lo et al., 2024). This issue is especially consequential in legal contexts, where inaccurate terminology, unsupported claims, or inappropriate professional formulations may substantially alter meaning. Students therefore need to verify information, compare alternatives, identify problematic output, revise generated language, and justify their communicative choices.

The role of the teacher consequently remains central in AI-integrated instruction. Rather than replacing teacher-mediated learning, AI can extend opportunities for practice, simulation, feedback, and revision when these activities are structured around explicit professional and linguistic objectives. Teacher guidance is required to establish task purposes, model appropriate professional communication, support language development, and guide students in evaluating AI-generated content (Li & Han, 2024; Miao & Holmes, 2023). This pedagogical position shifts attention from "using AI" as an isolated classroom activity to integrating AI within a structured cycle of professional task performance, feedback, critical evaluation, and revision.

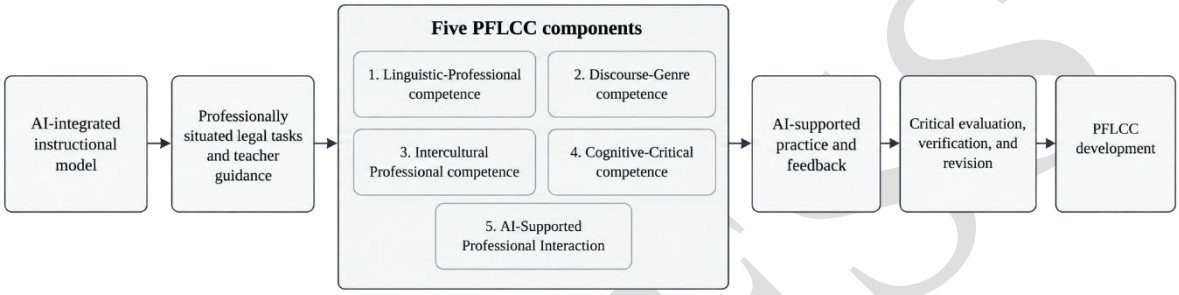
2.3. Research Gap and the Proposed AI-Integrated PFLCC Model

Although previous research provides evidence of the potential of AI for language practice, feedback, interaction, writing, speaking, and learner self-regulation, several issues remain insufficiently addressed when these findings are considered in relation to professional foreign language education. First, much of the literature examines particular language skills or individual AI functions rather than PFLCC as an integrated, multidimensional outcome. Second, studies of AI-supported language learning do not always connect technological activities systematically with the discourse, intercultural, cognitive-critical, and professional requirements of a specific disciplinary context. Third, the literature provides comparatively limited evidence on AI-integrated professional foreign language instruction for law students, particularly within the Kazakhstani higher education context. Finally, the need for learners to evaluate, verify, and revise AI-generated professional communication suggests that responsible AI-supported interaction should be treated as part of competence development rather than merely as access to a technological tool. These limitations indicate a gap between research on AI-assisted language activities and the development of an integrated pedagogical model in which AI-supported learning is systematically aligned with the components of professional communicative competence. The present study addresses this gap by proposing an AI-integrated methodological model specifically oriented toward the development of PFLCC among law students.

The proposed model conceptualizes PFLCC through five interconnected components: (1) Linguistic-Professional competence, concerning accurate and contextually appropriate use of legal English; (2) Discourse-Genre competence, concerning the organization of communication according to professional

genres and communicative purposes; (3) Intercultural Professional competence, concerning pragmatic and culturally appropriate professional interaction; (4) Cognitive-Critical competence, involving analysis, verification, evaluation, and reasoned communicative decision-making; and (5) AI-Supported Professional Interaction, concerning the purposeful, critical, and responsible use of AI in professional communication. These components are treated as interdependent rather than isolated abilities. The model links these components through a pedagogical sequence in which professionally situated legal tasks and teacher guidance provide the instructional context; AI-supported practice and feedback extend opportunities for interaction and revision; and students critically evaluate, verify, and revise both their own and AI-generated communication. Figure 1 illustrates the conceptual relationships.

Figure 1
The Relationship among the Theoretical Constructs and Instructional Intervention



The proposed model conceptualizes AI-integrated instruction as a teacher-guided and iterative approach to developing PFLCC through professionally meaningful communicative tasks. Rather than treating AI primarily as a tool for generating or correcting language, the model positions it as a supplementary resource for professional practice, simulation, feedback, and revision. Critical evaluation of AI-generated content and teacher-guided reflection are integrated into the instructional process as pedagogical principles supporting responsible professional communication. The model informed the design of the 10-week intervention, while its effectiveness was evaluated using the overall PFLCC performance measures described in the Methodology and Results sections. The empirical analysis examines changes in overall PFLCC performance and students' reported learning experiences; it does not independently test the sequential or causal relationships among the model's pedagogical processes.

3. Methodology

3.1. Research Design

The study was conducted within a pragmatic research paradigm, which allows quantitative and qualitative evidence to be combined when examining educational interventions in authentic instructional settings. The quantitative component employed a quasi-experimental pre-test/post-test nonequivalent-groups design involving an experimental group (EG) and a control group (CG). Since individual random assignment was not employed, the research design was quasi-experimental rather than a randomized controlled trial. The qualitative component included students' reflective self-assessments, classroom observations, and feedback concerning their learning experiences. The qualitative data were used to examine how students experienced AI-supported professional foreign language learning during the intervention, with particular attention to perceived benefits, difficulties, use of AI-supported feedback, critical evaluation of AI-generated content, and changes in professional communicative practice.

For the purposes of the present study, PFLCC was operationally defined as the ability of law students to use English accurately, appropriately, coherently, critically, and interactively to accomplish professionally relevant communication tasks. PFLCC was conceptualized and assessed as a multidimensional construct comprising five interrelated sub-competences: (1) Linguistic–Professional, (2) Discourse–Genre, (3) Intercultural Professional, (4) Cognitive–Critical, and (5) AI-Supported Professional Interaction. The same five dimensions informed the theoretical framework, instructional

intervention, diagnostic and post-test assessment, and component-level analysis. To ensure direct alignment between the theoretical structure of PFLCC, instructional activities, and assessment, each of the five components was operationalized through specific professionally oriented learning and assessment tasks. The same component framework guided both the 10-week instructional intervention and the evaluation of student performance. Linguistic-Professional competence focused on accurate and contextually appropriate use of legal English; Discourse-Genre competence concerned the organization of communication according to professional legal genres; Intercultural Professional competence addressed pragmatic and culturally appropriate professional interaction; Cognitive-Critical competence involved the analysis, verification, and reasoned evaluation of professional and AI-generated information; and AI-Supported Professional Interaction concerned the purposeful and critical use of AI in simulated legal communication. Table 1 presents the alignment between each component, the instructional activities used to develop it, the corresponding assessment tasks, and the criteria applied to student performance.

Table 1

Operationalization, Instruction, and Assessment of PFLCC Components

PFLCC component	What it means in this study	Activities used to develop it	Assessment task	Assessment criteria
Linguistic-Professional	Ability to use English accurately and appropriately in legal contexts, including legal terminology and grammatical structures	Legal vocabulary exercises; terminology matching; case-based vocabulary tasks; drafting and AI-assisted revision of legal texts	Students explained a legal issue/case using appropriate professional vocabulary	Accuracy; appropriate use of legal terminology; grammatical control
Discourse-Genre	Ability to organize spoken and written communication according to conventions of legal and professional genres	Legal correspondence; case summaries; client-oriented explanations; structured arguments; AI-assisted drafting and revision	Students produced a professional legal text or structured oral response	Organization; coherence; genre appropriateness; task completion
Intercultural Professional	Ability to communicate appropriately in professional situations involving different cultural and communicative expectations	Client consultations; role-plays; comparison of professional communication conventions; adaptation of messages for different audiences	Students responded to a simulated intercultural professional situation	Pragmatic appropriateness; audience awareness; intercultural sensitivity
Cognitive-Critical	Ability to analyse legal information, evaluate arguments, identify inaccuracies, and justify communicative choices	Case analysis; comparison of alternative arguments; verification of AI-generated information; identification and correction of problematic AI output	Students evaluated an AI-generated legal/professional response and justified revisions	Critical evaluation; reasoning; identification of problems; justification of revisions
AI-Supported Professional Interaction	Ability to use AI critically and purposefully to support professional communication rather than accept generated output uncritically	AI-supported client simulations; prompt-based legal discussions; drafting and revising professional correspondence; comparison of student and AI responses	Students completed an AI-supported professional communication task and critically revised the generated output	Appropriate AI use; interaction quality; evaluation and revision of AI output; professional appropriateness

IN PRESS

3.2. Participants

The participants were 47 first-year undergraduate law students from Al-Farabi Kazakh National University, Kazakhstan, aged 17–18 years. Two intact classes participated in the study. The EG consisted of 25 students, whereas the CG consisted of 22 students. All participants had approximately B1 English proficiency according to the Common European Framework of Reference for Languages (CEFR).

The inclusion criteria were enrolment as a first-year law student in the participating university, participation in the relevant foreign language course during the study period, approximately B1 English proficiency, completion of the diagnostic assessment, and informed consent to participate in the study. Only participants with the required diagnostic and post-test data were included in the reported analysis.

The EG and CG were comparable in terms of academic specialization, year of study, age range, approximate initial English proficiency, and instructional time. Both groups completed the same PFLCC assessment framework before and after the intervention. Baseline competence distributions were also examined before the formative stage. In the EG, 32.0% of students were at the Low level, 44.0% at Medium, 20.0% at Sufficient, and 4.0% at High. The corresponding proportions in the CG were 27.3%, 40.9%, 22.7%, and 9.1%, respectively. These distributions indicated a broadly similar baseline pattern, with Low and Medium levels predominating in both groups. Because individual randomization was not performed, however, baseline similarity was not treated as evidence of complete experimental equivalence.

3.3. Instruments

Several AI-supported tools were incorporated according to specific pedagogical functions. Grammarly was used to support revision of grammatical accuracy, lexical choice, mechanics, and professional style in written communication. Students were expected to evaluate suggestions rather than accept corrections automatically. DeepL Translator was used to compare alternative translations and formulations of professional terminology and to examine contextual appropriateness. ELSA Speak supported pronunciation, intelligibility, and repeated oral practice. TalkPal AI was employed for simulated dialogue and role-play activities, providing opportunities for repeated professional interaction. Canva Magic Write supported preparation and revision of multimodal professional presentations. Otter.ai provided transcripts of oral production that could subsequently be analysed for fluency, organization, and recurring language problems. QuillBot was used for controlled paraphrasing and revision activities in which students compared alternative formulations and selected or revised language according to professional purpose.

3.3. Procedure

The study consisted of three consecutive stages: diagnostic, formative, and control-reflective. The diagnostic stage was conducted before the intervention to establish students' initial PFLCC levels and provide a baseline for subsequent comparison. Both groups completed assessment activities targeting the same five PFLCC dimensions. The diagnostic assessment combined professionally contextualized language-use tasks, written communication, simulated oral professional interaction, analytical and argumentative activities, and tasks designed to examine students' ability to engage critically with AI-supported resources. Supplementary self-report information concerning students' attitudes toward language learning, digital technologies, and AI-supported learning was also collected. These self-report data were used to contextualize students' experiences and were not treated as substitutes for performance-based measures of PFLCC. Examples of open-ended reflective prompts included: "Which AI-supported activity was most useful for practicing professional legal English, and why?", "Describe an example in which you revised an AI-generated response because it was inaccurate or inappropriate", "How did AI-supported feedback influence the way you revised your professional communication?", and "In which situations did you need teacher guidance when using or evaluating AI-generated content?"

The formative intervention lasted 10 weeks and comprised 20 instructional sessions. Two sessions were conducted each week, and each session lasted two hours, resulting in a total instructional duration of 40

hours. Following completion of the 10-week formative stage, both groups completed the post-test assessment. The post-test was designed to measure the same five PFLCC dimensions assessed during the diagnostic stage. Parallel rather than identical tasks were used at the two measurement points to reduce direct memorization effects while preserving comparability in terms of targeted competence, professional context, task type, and expected level of difficulty.

These materials provided supplementary information about students' perceptions of their development, difficulties, use of feedback, and experience of AI-supported professional language learning. Reflective data were treated as complementary qualitative evidence and were not substituted for performance-based PFLCC outcomes.

3.3.1. Instruction

3.3.1.1. EG Instruction

During the 10-week formative stage, the EG received instruction based on the developed AI-integrated PFLCC model (see Table 2). The intervention consisted of 20 two-hour sessions and combined professionally oriented legal communication, AI-supported practice, multimodal resources, collaborative activities, critical evaluation, feedback, and reflection. The intervention was organized around communication relevant to law students rather than general-purpose AI activities. Professional tasks included development and use of legal-professional vocabulary, analysis of professionally relevant texts, legal case discussions, client-oriented explanations, professional correspondence, simulated consultations, argumentation, negotiation, presentations, critical evaluation of AI-generated content, and revision of spoken and written professional communication.

Table 2
Ten-Week Instructional Sequence

Week	Instruction
1	Introduction to PFLCC and professional legal communication: Students were introduced to the requirements of professional foreign language communication, legal-professional vocabulary, communicative purposes, and principles of responsible AI-supported learning.
2	Professional vocabulary and linguistic accuracy: Activities focused on professional terminology, explanation of legal concepts, contextual vocabulary use, pronunciation, and AI-supported comparison and revision of language.
3	Professional discourse and written genres: Students worked with professional correspondence, structured written communication, case summaries, organization, coherence, and appropriate register.
4	Client-oriented communication: Students practised explaining professionally relevant information to a simulated client, adapting terminology and register, asking and answering questions, and participating in AI-supported role-play.
5	Case discussion and critical reasoning: Students analysed professionally relevant cases, identified key information, formulated positions, constructed arguments, and critically evaluated alternative and AI-generated interpretations.
6	Professional consultation and interaction: Activities developed questioning, clarification, explanation, turn management, responsiveness, and interactional strategies through simulated professional consultations.
7	Argumentation and negotiation: Students presented and defended positions, supported claims, responded to counterarguments, negotiated meaning, and compared alternative formulations for professional appropriateness.
8	Intercultural and pragmatic professional communication: Activities focused on register, politeness, professional etiquette, audience awareness, pragmatic appropriateness, and adaptation of communication to different interlocutors and contexts
9	Multimodal professional communication: Students prepared and delivered professionally oriented presentations or case briefings using multimodal resources. AI-supported tools were used for preparation, revision, feedback, and reflection.
10	Integrated professional communication and reflection: Students completed integrated tasks combining linguistic-professional, discourse–genre, intercultural, cognitive-critical, and AI-supported dimensions, followed by self-assessment, revision, and reflection.

IN PRESS

A recurring instructional sequence was followed. First, the teacher introduced the professional communication objective and relevant linguistic, discourse, or pragmatic features. Second, students analysed examples or observed teacher modelling of the targeted performance. Third, guided practice was conducted to establish the necessary linguistic and professional resources. Fourth, students applied these resources in professionally situated tasks with AI-supported practice, simulation, or feedback. Fifth, students evaluated the generated output and revised their performance rather than accepting AI suggestions automatically. Finally, the activity concluded with reflection and teacher feedback.

The first instructional session of each week generally emphasized introduction, modelling, analysis, and guided practice of the targeted professional communication skills. The second session emphasized application through professionally situated communication, AI-supported interaction or feedback, revision, collaborative work, and reflection. This sequence enabled students to progress from controlled practice toward increasingly independent professional communication. The teacher acted as instructional designer, facilitator, language guide, and evaluator. The teacher introduced the objectives of professional tasks, explained relevant language and discourse features, demonstrated responsible use of AI tools, monitored interaction, provided formative feedback, and guided students in evaluating AI-generated output. Students were expected to analyse professional material, participate actively in individual and collaborative tasks, interact in simulated professional situations, evaluate AI-generated responses, revise their own language production, and reflect on their communicative decisions.

3.3.1.2. CG Instruction

The CG completed the same number of instructional sessions as the experimental group: 20 sessions over 10 weeks, with two two-hour sessions per week, resulting in 40 hours of instruction. The control group followed conventional foreign language instruction without systematic implementation of the developed AI-integrated PFLCC model. Instruction included teacher explanation, reading and vocabulary activities, grammar practice, pair and group speaking tasks, written activities, classroom discussion, and conventional teacher feedback.

To maintain comparability, the CG worked toward broadly corresponding language and professional communication objectives and received the same amount of instructional time. Where the experimental group completed an AI-supported activity, the control group completed a comparable conventional activity targeting the same general communicative function whenever possible. For example, AI-supported dialogue or professional simulation in the EG corresponded to peer role-play and teacher-guided interaction in the CG. AI-supported writing revision in the EG corresponded to conventional self-editing, peer review, or teacher correction in the CG. Both groups were assessed using the same PFLCC dimensions, four-level performance framework, and pre-test/post-test procedures.

3.3.2. Assessment and Scoring

Several procedures were used to support the consistency and interpretability of the PFLCC assessment. Because PFLCC was conceptualized as a multidimensional performance-based construct, evidence concerning measurement quality was considered in relation to the alignment between the theoretical construct, assessment tasks, scoring criteria, and instructional objectives rather than being inferred solely from a general claim of instrument validity. Content validity was addressed through construct-to-task alignment. The assessment framework was developed around the same five dimensions specified in the theoretical model: Linguistic-Professional, Discourse-Genre, Intercultural Professional, Cognitive-Critical, and AI-Supported Professional Interaction. Each dimension was represented by observable performance indicators and corresponding task types. Linguistic accuracy, professional vocabulary, comprehensibility, and fluency were associated with Linguistic-Professional competence; coherence, organization, genre appropriateness, and task completion with Discourse-Genre competence; register, politeness, pragmatic appropriateness, and audience adaptation with Intercultural Professional competence; analysis, argumentation, comparison, verification, and justification with Cognitive-Critical competence; and purposeful AI interaction, evaluation of generated output, feedback use, and revision with AI-Supported Professional Interaction. The four-level analytic scoring framework (1 = Low, 2 = Medium, 3 = Sufficient, 4 = High) was defined using explicit performance descriptors. Each of the five PFLCC components was rated on a four-point scale: 1 = Low, indicating

limited performance and substantial difficulty completing the professional communicative task; 2 = Medium, indicating partial performance with noticeable linguistic, organizational, pragmatic, critical, or interactional limitations; 3 = Sufficient, indicating generally effective and appropriate performance with minor limitations; and 4 = High, indicating consistently accurate, appropriate, coherent, critical, and professionally effective performance (see Appendix B).

For each student, the five component scores were averaged to obtain an overall PFLCC score ranging from 1.00 to 4.00. For interpretation, the overall scores were classified into four equal-width ranges: 1.00-1.75 = Low, 1.76-2.50 = Medium, 2.51-3.25 = Sufficient, 3.26-4.00 = High. The same scoring procedure and cut-off ranges were applied at the diagnostic and post-test stages. Scores were assigned according to demonstrated communicative performance rather than students' self-reported confidence or frequency of technology use. For example, in the AI-Supported Professional Interaction task, a student who accepted an AI-generated legal response without identifying inaccuracies or revising inappropriate language was rated at the Low level (1). A student who identified some problems but provided only limited revision or justification was rated at the Medium level (2). A student who identified the main problems, revised the response appropriately, and provided a generally adequate justification was rated at the Sufficient level (3). A student who systematically identified linguistic and professional problems, critically evaluated the generated content, produced an appropriate revision, and clearly justified the changes was rated at the High level (4). Sample assessment tasks and scoring examples are provided in Appendix A.

3.3.3. Data Analysis

Quantitative analysis was conducted at both descriptive and inferential levels. Frequencies and percentages were used to describe the distribution of students across the four PFLCC performance levels (Low, Medium, Sufficient, and High). Means and standard deviations were calculated for continuous/ scale-based overall PFLCC scores.

To examine within-group change from pre-test to post-test, paired-samples t tests were conducted separately for the EG and CG (RQ1). For each group, the mean pre-test PFLCC score was compared with the corresponding mean post-test score. Mean differences, t values, degrees of freedom, and p-values were used to evaluate whether the observed pre-test-post-test changes were statistically significant. Statistical significance was evaluated at an alpha level of .05.

Baseline data were examined before evaluating post-intervention differences. Because the study employed intact classes rather than individual random assignment, baseline comparability was considered important for the interpretation of subsequent between-group differences. Categorical distributions across the four competence levels were examined separately using a chi-square test of independence. For RQ2, frequencies and percentages were used to describe the distribution of students across the four PFLCC competence levels at pre-test and post-test in each group. The post-test distributions of the experimental and control groups were compared using a chi-square test of independence. Because the pre-test and post-test classifications were obtained from the same students, the descriptive distributions were not interpreted as evidence of statistically significant within-group categorical changes. Statistical significance was evaluated at an alpha level of .05. Statistical significance was evaluated using an alpha level of .05. Effect sizes and confidence intervals were considered together with p-values in order to avoid interpreting statistical significance as the sole indicator of educational importance. For the between-group comparison of overall post-test PFLCC scores, an independent-samples *t* test was used (RQ3). Cohen's *d*, calculated using the pooled standard deviation, was reported to quantify the standardized magnitude of the between-group difference, together with a 95% confidence interval for the unstandardized mean difference.

Assumptions corresponding to each inferential procedure were examined before interpretation. For the paired-samples t tests, the distribution of the pre-test-post-test difference scores within each group was examined for approximate normality and substantial outliers. For the independent-samples *t* test, these included independence of observations, approximate normality of the outcome within groups, and homogeneity of variance. For categorical analysis, expected cell frequencies were inspected when evaluating the suitability of the chi-square test.

IN PRESS

To address RQ4, the qualitative data obtained from students' reflective self-assessments, classroom observations, and feedback were analysed using thematic analysis. The analysis followed a systematic process of familiarization, initial coding, theme development, and review (Braun & Clarke, 2006). First, the qualitative records were reviewed repeatedly to identify statements and observations relevant to RQ4. Initial codes were then assigned to recurring ideas concerning professional communication practice, AI-supported feedback, difficulties encountered during the intervention, critical evaluation of AI-generated content, revision of language output, and perceived learning benefits. Related codes were subsequently grouped into broader themes. The themes were reviewed against the original qualitative records to ensure that they adequately represented the students' reported and observed experiences. Candidate themes were then compared with the coded extracts and the original qualitative records, refined where necessary, and defined according to the recurring patterns represented in the data. The qualitative findings were used to complement, rather than statistically validate, the quantitative PFLCC outcomes and to provide contextual insight into how students experienced the AI-integrated instructional model.

4. Results

The results are presented according to the four research questions. Before addressing the research questions, baseline equivalence between the EG and CG was examined to determine whether the groups were statistically comparable before the intervention. RQ1 then examined within-group changes in overall PFLCC performance from pre-test to post-test; RQ2 addressed changes in the distribution of students across the four PFLCC competence levels; RQ3 examined between-group differences in overall post-test PFLCC performance; and RQ4 explored students' experiences of AI-supported professional foreign language learning.

4.1. Diagnostic Findings

Before examining within-group changes (RQ1), baseline equivalence between the EG and CG was tested using an independent-samples t test (Table 3).

Table 3
Baseline Comparison of Overall PFLCC Scores

Outcome	EG, M (SD)	CG, M (SD)	Mean difference	t(45)	p
Pre-test overall PFLCC	1.96 (0.84)	2.14 (0.89)	−0.18	−0.70	.49

The baseline comparison showed that the EG (M = 1.96, SD = 0.84) and CG (M = 2.14, SD = 0.89) did not differ significantly in overall pre-test PFLCC performance (t(45) = −0.70, p = .49). Thus, no statistically significant baseline difference was detected between the two groups, supporting their comparability before the intervention. After baseline comparability had been examined, paired-samples t tests were conducted separately for the EG and CG to address RQ1 (Table 4). Each student's overall pre-test PFLCC score was paired with the same student's post-test score.

Table 4
Diagnostic Assessment of PFLCC

Group	Pre-test M (SD)	Post-test M (SD)	Mean change	t	df	p	95% CI for mean change
EG	1.96 (0.84)	2.96 (0.88)	+1.00	5.21	24	< .001	[0.60, 1.40]
CG	2.14 (0.89)	2.41 (0.96)	+0.27	1.42	21	.170	[−0.13, 0.67]

In this analysis, the EG showed an increase in overall PFLCC performance from pre-test (M = 1.96, SD = 0.84) to post-test (M = 2.96, SD = 0.88). The paired-samples *t* test indicated a statistically significant increase (t(24) = 5.21, p < .001), with a mean change of 1.00 points. In contrast, the CG showed a smaller increase from pre-test (M = 2.14, SD = 0.89) to post-test (M = 2.41, SD = 0.96). In this table, the within-group change was not statistically significant (t(21) = 1.42, p = .170), with a mean change of 0.27 points.

4.2. Change in the Overall PFLCC Attainment

To address RQ2, students were classified into four PFLCC performance levels: Low, Medium, Sufficient, and High. Scores classified as Sufficient or High were treated as attainment of the target PFLCC threshold. Table 5 compares the pre-test and post-test distributions for both groups.

Table 5

Changes in PFLCC Competence-Level Distribution from Pre-Test to Post-Test

Competence level	EG Pre-test n (%)	EG Post-test n (%)	CG Pre-test n (%)	CG Post-test n (%)
Low	8 (32.0)	2 (8.0)	6 (27.3)	5 (22.7)
Medium	11 (44.0)	5 (20.0)	9 (40.9)	7 (31.8)
Sufficient	5 (20.0)	10 (40.0)	5 (22.7)	6 (27.3)
High	1 (4.0)	8 (32.0)	2 (9.1)	4 (18.2)
Sufficient + High	6 (24.0)	18 (72.0)	7 (31.8)	10 (45.5)
Total	25 (100.0)	25 (100.0)	22 (100.0)	22 (100.0)

At pre-test, 24.0% (6/25) of students in the EG and 31.8% (7/22) in the CG were classified at the Sufficient or High level. At post-test, these proportions increased to 72.0% (18/25) and 45.5% (10/22), respectively. Thus, the proportion reaching the target PFLCC threshold increased by 48.0 percentage points in the EG and 13.7 percentage points in the CG. The distributional change was particularly pronounced in the EG. The proportion of students at the Low level decreased from 32.0% to 8.0%, while the proportion at the High level increased from 4.0% to 32.0%. In the CG, the Low category decreased from 27.3% to 22.7%, whereas the High category increased from 9.1% to 18.2%.

To complement the descriptive distributional analysis, a chi-square test of independence was used to compare the distribution of students across the four PFLCC categories between the experimental and control groups at post-test. Although a descriptively greater proportion of students in the experimental group reached the Sufficient and High categories, the between-group difference in the post-test distribution was not statistically significant ($\chi^2(3) = 3.78$, $p = .287$). These findings indicate a descriptive shift toward higher competence categories in the experimental group; however, the post-test comparison did not establish a statistically significant difference in categorical distributions between the groups. As no within-group test of pre-post categorical transitions was conducted, changes in the distribution of PFLCC categories over time cannot be inferred statistically from this analysis alone.

4.3. Post-Test Comparisons

To address RQ3, an independent-samples t test was used to compare overall post-test PFLCC scores between the EG and CG.

Table 6

Between-Group Comparison of Overall Post-Test PFLCC Scores

Outcome	EG, M (SD)	CG, M (SD)	Mean difference	t(45)	p	Cohen's d	95% CI
Overall PFLCC	2.96 (0.88)	2.41 (0.96)	0.55	2.05	.046	0.60	[0.01, 1.09]

The EG demonstrated significantly higher overall post-test PFLCC performance ($M = 2.96$, $SD = 0.88$) than the CG ($M = 2.41$, $SD = 0.96$; $t(45) = 2.05$, $p = .046$). The mean difference was 0.55 points, and Cohen's $d = 0.60$ indicated a moderate standardized between-group difference. Thus, in response to RQ3, the students who received the AI-integrated instructional model demonstrated significantly higher overall post-test PFLCC performance than those who received conventional foreign language instruction.

4.4. Students' Experiences of AI-Supported Professional Foreign Language Learning

To address RQ4 and represent the qualitative strand of the mixed-methods study, students' reflective self-assessments, classroom observations, and feedback were analysed thematically. Three recurrent themes were identified: (1) professional communication practice, (2) feedback and revision, and (3)

critical engagement with AI-generated content. Three recurrent themes were identified through thematic analysis: professional communication practice, feedback and revision, and critical engagement with AI-generated content (Table 7). In this illustrative distribution, professional communication practice was the most frequently represented theme, followed by feedback and revision and critical engagement with AI-generated content. The pattern is consistent with the qualitative interpretation presented above, which emphasizes students' opportunities to practise legal English, respond to feedback, revise their work, and evaluate AI-generated responses critically. However, the numerical frequencies must be verified against the original qualitative records before being reported as empirical findings.

Table 7

Frequency of Themes Identified in the Qualitative Analysis

Theme	Illustrative evidence from the study	Frequency, n	Percentage, %
Professional communication practice	Practice of legal English, case discussions, client-oriented communication, and professional correspondence	20	80.0
Feedback and revision	Use of AI-supported feedback to improve vocabulary, grammar, organization, and professional appropriateness	18	72.0
Critical engagement with AI-generated content	Verification, evaluation, and revision of AI-generated responses	16	64.0

The first theme, professional communication practice, concerned students' experiences of using English in simulated professional legal situations. Students' reflections suggested that AI-supported activities provided additional opportunities to practise legal English through client-oriented communication, case discussion, and professional correspondence. One student highlighted the practical value of these activities: *"It helped me understand how to explain a legal problem in English, especially when I had to respond as if I were talking to a real client"* (AK1). Similarly, another participant emphasized the importance of practising professional communication: *"Before these activities, I knew some legal words, but I was not always confident about how to use them in a professional conversation"* (AF2). These reflections illustrate the perceived value of contextualized language practice. Classroom observations also suggested that repeated practice supported students in formulating professional responses and using legal terminology in context. However, producing extended and appropriately structured professional responses remained more demanding than completing isolated vocabulary or language exercises.

The second theme, feedback and revision, reflected students' experiences of receiving and responding to AI-supported feedback. Students generally perceived such feedback as useful for identifying language problems and considering alternative formulations. One participant explained: *"The AI feedback helped me notice mistakes in my grammar and choose more appropriate words for my legal response"* (AF2). Another student emphasized the importance of revising initial answers: *"When I compared my first answer with the suggested version, I understood that I needed to organize my ideas more clearly and make my response sound more professional"* (LF). These comments illustrate how AI-supported feedback could encourage students to reconsider vocabulary, grammatical accuracy, organization, and professional appropriateness. Classroom observations further indicated that teacher guidance remained important when students were uncertain whether an AI-generated suggestion was suitable for a particular legal communicative context.

The third theme, critical engagement with AI-generated content, concerned students' evaluation and revision of generated responses. Students' reflections suggested growing awareness of the need to verify and revise AI-generated output rather than accept it automatically. One student observed: *"Sometimes the AI answer sounds correct, but I still need to check whether the information is accurate and relevant to the legal situation"* (LF). Another participant described the importance of independent evaluation: *"I learned that I should not copy the AI response directly. I need to check the information and change the wording if it does not fit the situation"* (AK1). These reflections illustrate the distinction between linguistic plausibility and professional accuracy. Activities involving comparison, verification, and

revision may therefore support the development of critical evaluation alongside professional language practice. Nevertheless, the extent to which students developed these skills should be established using the actual reflection data and classroom observation records.

Overall, the qualitative findings addressing RQ4 indicate that students experienced the AI-integrated model through expanded opportunities for professional communication practice, iterative feedback and revision, and critical evaluation of AI-generated content. Teacher guidance remained relevant across these processes. These qualitative findings complement the quantitative results by providing contextual insight into how students engaged with the instructional model during the intervention.

5. Discussion

To examine changes in PFLCC during the intervention, pre-test and post-test results were analyzed separately for the EG and CG. This analysis focused on changes in overall PFLCC scores over the 10-week intervention and provided an initial assessment of whether students' performance improved in each group. Within-group comparisons were used to identify changes over time, while the between-group comparison was considered separately to determine whether the groups differed at the post-test stage. Regarding RQ1, the EG demonstrated a statistically significant increase in overall PFLCC scores, whereas the CG showed a smaller, non-significant change. The improvement observed in the EG may be associated with the combination of professionally oriented tasks, AI-supported language practice, feedback, and revision. However, within-group findings alone do not establish that the improvement in the EG was significantly greater than that in the CG. This interpretation is consistent with research highlighting the potential of AI to support language practice and teaching (Crompton et al., 2024; Kohnke et al., 2023; Lo et al., 2024). Du and Daniel (2024), in a systematic review of 24 empirical studies, found that AI-powered chatbots were associated with benefits for EFL speaking practice, including improved speaking performance, confidence, engagement, and motivation, as well as reduced speaking anxiety. Similarly, Karataş et al. (2024), in a qualitative case study involving 13 university preparatory-class students over four weeks, reported that ChatGPT supported students' learning experiences, particularly in writing, grammar, and vocabulary, while also enhancing motivation and engagement.

Regarding RQ2, the distribution of students across the four PFLCC competence levels was examined at pre-test and post-test in both the EG and CG. The proportion of students classified at the Sufficient or High levels increased from 24.0% to 72.0% in the EG and from 31.8% to 45.5% in the CG. The EG also showed a decrease in the proportion of students at the Low level and an increase in the proportion reaching the High level. At post-test, the distribution of students across the four competence levels was also compared between the EG and CG to examine differences in their competence profiles. These descriptive changes suggest a shift toward higher PFLCC attainment following the intervention. Professionally contextualized tasks may have helped students apply language resources to communicative purposes, consistent with communicative competence theory (Canale & Swain, 1980) and the emphasis on specific learning needs in English for Specific Purposes (Hutchinson & Waters, 1987). AI-supported practice and feedback may have further supported this process (Huang et al., 2022; Kohnke et al., 2023). However, the post-test comparison of the four competence-level distributions was not statistically significant. The findings should therefore be interpreted as descriptive rather than conclusive evidence of differences in categorical attainment.

Regarding RQ3, the EG achieved a higher overall post-test PFLCC score than the CG. The difference was statistically significant, with the EG demonstrating better overall performance and a moderate effect size. These findings indicate that participation in the AI-integrated intervention was associated with higher post-test PFLCC performance among law students. This result supports the potential value of integrating AI-supported activities into structured professional language instruction. The model combined professional communication tasks with opportunities for simulated interaction, feedback, and revision, reflecting the pedagogical importance of purposeful AI integration rather than technology use alone (Ouyang & Jiao, 2021; Zawacki-Richter et al., 2019). The result also corresponds with research on AI affordances and challenges in language teaching (Crompton et al., 2024; Kohnke et al., 2023; Lo et al., 2024). However, because the intervention included several connected instructional elements, the

observed difference cannot be attributed exclusively to AI. The relatively small sample and p-value close to .05 also warrant cautious interpretation.

Regarding RQ4, the qualitative analysis of students' self-assessments, classroom observations, and feedback identified three main themes: (1) professional communication practice, (2) feedback and revision, and (3) critical engagement with AI-generated content. These themes reflect students' experiences with AI-supported professional foreign language learning during the intervention. The first theme suggests that simulated legal communication offered opportunities to practise professional vocabulary and respond to context-specific communicative demands. This is consistent with the focus on purposeful language use in communicative approaches (Canale & Swain, 1980) and research on AI-supported speaking practice (Du & Daniel, 2024).

The second theme concerned students' use of feedback to revise vocabulary, grammar, organization, and professional tone. This finding aligns with research examining AI-supported language teaching and feedback (Crompton et al., 2024; Kohnke et al., 2023; Li & Han, 2024). The third theme highlighted the importance of evaluating and revising AI-generated responses instead of accepting them uncritically. This is particularly relevant to legal communication, where linguistic fluency does not guarantee factual accuracy or professional appropriateness. Previous work has similarly emphasized the opportunities and risks of generative AI in education (Kasneci et al., 2023; Miao & Holmes, 2023). Taken together, the themes illustrate how students experienced the intervention's practice, feedback, and critical evaluation activities. They do not, however, independently demonstrate improvement in each of the five PFLCC components or a measured increase in critical AI literacy.

The findings have several practical implications for foreign language instruction in legal education. First, professionally oriented AI-supported activities can be incorporated into language courses to provide students with opportunities to practice legal communication in simulated professional situations, such as client consultations, case discussions, and formal correspondence. Second, AI-generated feedback can be used as a resource for revising vocabulary, grammar, discourse organization, and professional tone, provided that students critically review suggestions rather than accept them automatically. Third, teachers should explicitly guide students in evaluating the accuracy, relevance, and appropriateness of AI-generated content, particularly in legal contexts where linguistic quality does not necessarily ensure factual or professional reliability. These activities should be accompanied by teacher supervision and clear guidelines for responsible AI use. Although the qualitative findings do not establish measurable gains in each PFLCC component or in AI literacy, they suggest practical ways to integrate communication practice, feedback-based revision, and critical evaluation into foreign language teaching for law students.

Overall, the findings should be interpreted in light of a few limitations. The study involved 47 first-year law students at one university and used intact classes rather than individual random assignment. Furthermore, the results report overall PFLCC performance rather than separate statistical changes for all five components, and no delayed post-test was conducted. Future studies should include larger, multi-institutional samples, follow-up assessments, and component-level analyses to investigate the durability and specific dimensions of PFLCC development. In sum, the findings suggest that an AI-integrated, teacher-guided instructional model can support professional foreign language learning, while further research is needed to establish its broader applicability and longer-term effects.

Acknowledgments

The authors sincerely thank the law students of Al-Farabi Kazakh National University who participated in the study and contributed to the completion of the experimental intervention.

AI-assisted tools were used during the preparation of the manuscript for language polishing, improvement of grammatical clarity, and assistance with the organization and editing of selected passages. The authors critically reviewed, revised, and approved all AI-assisted text and take full responsibility for the accuracy, interpretation, arguments, citations, and final content of the manuscript.

Disclosure Statement

The authors claim no conflict of interest.

Funding

The research did not receive any specific grants from funding agencies.

Data Availability

The data supporting the findings of this study are available from the corresponding author upon reasonable request. The data are not publicly available due to considerations related to participant privacy and confidentiality.

Ethics Declaration

Participation in the study was voluntary, and informed consent was obtained from all participants. Students were informed about the purpose and procedures of the research and their right to withdraw without academic consequences. Participant data were anonymized using identification codes during data processing and analysis. Particular attention was given to privacy during the use of third-party AI-supported platforms.

References

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, *3*(2), 77–101. https://doi.org/10.1191/1478088706qp063oa

Canale, M., & Swain, M. (1980). Theoretical bases of communicative approaches to second language teaching and testing. *Applied Linguistics*, *1*(1), 1–47. https://doi.org/10.1093/applin/I.1.1

Crompton, H., Edmett, A., Ichaporia, N., & Burke, D. (2024). AI and English language teaching: Affordances and challenges. *British Journal of Educational Technology*, *55*(6), 2503–2529. https://doi.org/10.1111/bjet.13460

Dewey, J. (1933). *How we think: A restatement of the relation of reflective thinking to the educative process*. D. C. Heath.

Du, J., & Daniel, B. K. (2024). Transforming language education: A systematic review of AI-powered chatbots for English as a foreign language speaking practice. *Computers and Education: Artificial Intelligence*, *6*, Article 100230. https://doi.org/10.1016/j.caeai.2024.100230

Godwin-Jones, R. (2018). Emerging technologies. In J. I. Liontas & M. DelliCarpini (Eds.), *The TESOL encyclopedia of English language teaching* (pp. 1–6). Wiley. https://doi.org/10.1002/9781118784235.eelt0406

Holmes, W., Bialik, M., & Fadel, C. (2019). *Artificial intelligence in education: Promises and implications for teaching and learning*. Center for Curriculum Redesign.

Huang, W., Hew, K. F., & Fryer, L. K. (2022). Chatbots for language learning—Are they really useful? A systematic review of chatbot-supported language learning. *Journal of Computer Assisted Learning*, *38*(1), 237–257. https://doi.org/10.1111/jcal.12610

Hutchinson, T., & Waters, A. (1987). *English for specific purposes: A learning-centred approach*. Cambridge University Press.

Hymes, D. (1972). On communicative competence. In J. B. Pride & J. Holmes (Eds.), *Sociolinguistics* (pp. 269–293). Penguin.

Karatash, F., Abedi, F. Y., Ozek Gunyel, F., Karadeniz, D., & Kuzgun, Y. (2024). Incorporating AI in foreign language education: An investigation into ChatGPT's effect on foreign language learners. *Education and Information Technologies*, *29*, 19343–19366. https://doi.org/10.1007/s10639-024-12574-6

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

IN PRESS

Kohnke, L., Moorhouse, B. L., & Zou, D. (2023). ChatGPT for language teaching and learning. *RELC Journal*, *54*(2), 537–550. https://doi.org/10.1177/00336882231162868

Lo, C. K., Yu, P. L. H., Xu, S., Ng, D. T. K., & Jong, M. S. Y. (2024). Exploring the application of ChatGPT in ESL/EFL education and related research issues: A systematic review of empirical studies. *Smart Learning Environments*, *11*, Article 50. https://doi.org/10.1186/s40561-024-00342-5

Li, M., & Han, J. (2024). Exploring ChatGPT-supported teacher feedback in the EFL context. *System, 126*, Article 103502. https://doi.org/10.1016/j.system.2024.103502

Luckin, R., Holmes, W., Griffiths, M., & Forcier, L. B. (2016). *Intelligence unleashed: An argument for AI in education*. Pearson.

Mezirow, J. (1991). *Transformative dimensions of adult learning*. Jossey-Bass.

Miao, F., & Holmes, W. (2023). *Guidance for generative AI in education and research.* UNESCO. https://doi.org/10.54675/EWZM9535

Mukhtarova, S., Karagulova, B., Yergazina, A., Imangazina, A., & Kyyakhmetova, S. (2025). Bridging cultural boundaries: Translating untranslatable vocabulary in Kazakh-English literary texts. *International Journal of Society, Culture & Language*, *13*(1), 26–40. https://doi.org/10.22034/ijscl.2024.2034397.3554

Ouyang, F., & Jiao, P. (2021). Artificial intelligence in education: The three paradigms. *Computers and Education: Artificial Intelligence, 2*, Article 100020. https://doi.org/10.1016/j.caeai.2021.100020

Richards, J. C., & Rodgers, T. S. (2014). *Approaches and methods in language teaching* (3rd ed.). Cambridge University Press.

Schon, D. A. (1983). *The reflective practitioner: How professionals think in action*. Basic Books.

Swales, J. M. (1990). *Genre analysis: English in academic and research settings*. Cambridge University Press.

Tazhitova, G., Kurmanayeva, D., Zhalelova, G., Kassymbekova, N., Shegenova, Z., & Kadirsizova, S. (2025). Impact of local culture-based materials on the EFL communication skills of university students in Kazakhstan. *International Journal of Society, Culture & Language*, *13*(1), 121–131. https://doi.org/10.22034/ijscl.2024.2043500.3750

Zawacki-Richter, O., Marín, V. I., Bond, M., & Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education—Where are the educators? *International Journal of Educational Technology in Higher Education*, *16*, Article 39. https://doi.org/10.1186/s41239-019-0171-0

Zhumatayeva, Z. N., Mametkarim, Z. M., & Dosanova, A. M. (2024). The role of artificial intelligence in the formation of communicative competence in foreign language lessons. *Bulletin of the National Academy of Sciences of the Republic of Kazakhstan*, *6*(412), 119–130. https://doi.org/10.32014/2024.2518-1467.858

Appendix A

Examples of Diagnostic PFLCC Assessment Tasks

The following sample tasks illustrate the types of activities used in the diagnostic assessment. The tasks were designed to elicit performance corresponding to the five components PFLCC.

Task 1. Linguistic-Professional Competence

Read the short description of a legal dispute provided by the teacher. Explain the main legal issue in English using appropriate professional terminology. Define at least three relevant legal terms in a way that would be understandable to a non-specialist client.

Assessment focus: accuracy, appropriate use of legal terminology, grammatical control, and clarity of explanation.

Task 2. Discourse-Genre Competence

Write a short professional email to a client explaining the outcome of a legal consultation and outlining the next steps. Organize the information clearly and use an appropriate professional register.

Assessment focus: organization, coherence, genre conventions, professional register, and task completion.

Task 3. Intercultural Professional Competence

Imagine that you are explaining a legal procedure to an international client who is unfamiliar with the local legal and communicative conventions. Conduct a short simulated consultation in which you explain the procedure, respond to the client's questions, and adapt your language to the interlocutor.

Assessment focus: pragmatic appropriateness, audience awareness, politeness, clarity, and intercultural sensitivity.

Task 4. Cognitive-Critical Competence

Read two short arguments concerning the same legal issue. Identify the stronger argument and explain your choice. Identify any unsupported claims, inconsistencies, or information that would require further verification.

Assessment focus: analysis, comparison, identification of problems, reasoning, and justification.

Task 5. AI-Supported Professional Interaction

Review an AI-generated response to a simulated client's legal question. Identify any language, terminology, reasoning, or professional communication that may be inaccurate, unclear, or inappropriate. Revise the response and briefly explain the reasons for your changes.

Assessment focus: critical evaluation of AI-generated output, appropriate AI use, quality of revision, justification, and professional appropriateness.

Note. The diagnostic and post-test assessments targeted the same five PFLCC dimensions. Parallel rather than identical tasks were used at the two measurement points to maintain comparability while reducing direct memorization effects.

Appendix B

Table 1A

General Performance Levels and Descriptors

Score	Level	General performance descriptor
1	**Low**	Performance is substantially incomplete or inappropriate; frequent problems interfere with successful professional communication; little evidence of independent evaluation or revision.
2	**Medium**	Performance partially meets task requirements but contains noticeable linguistic, discourse, pragmatic, critical, or interactional limitations; evaluation and revision are inconsistent.
3	**Sufficient**	Performance generally meets task requirements; communication is mostly accurate, coherent, and professionally appropriate; relevant problems are identified and adequately addressed, although minor limitations remain.
4	**High**	Performance consistently meets task requirements; communication is accurate, coherent, contextually and professionally appropriate; evaluation, reasoning, and revision are systematic and well justified.

Sample scoring illustration: Suppose a student received the following scores: Linguistic-Professional = 3, Discourse-Genre = 3, Intercultural Professional = 2, Cognitive-Critical = 3, and AI-Supported Professional Interaction = 4. The overall PFLCC score was calculated as (3 + 3 + 2 + 3 + 4) / 5 = 3.00.

According to the predefined ranges, an overall score of 3.00 was classified as Sufficient (2.51–3.25). The same procedure was used to classify students' overall PFLCC performance at the diagnostic and post-test stages.

IN PRESS