

Developing Linguocognitive Competence in ESP through a TPACK-Based Integrative Digital Methodology: A Multi-Site Quasi-Experimental Study in Uzbekistan

Gofur Ashurovich Allamuratov

Department of English, National Pedagogical University of Uzbekistan named after Nizami, Tashkent, Uzbekistan

E-mail: allamuradovgofur@gmail.com ORCID: 0009-0005-2988-0325

DOI: [https://doi.org/10.63001/tbs.2026.v21.i04.S.I\(4\).pp455-471](https://doi.org/10.63001/tbs.2026.v21.i04.S.I(4).pp455-471)

KEYWORDS

linguocognitive competence; English for Specific Purposes; TPACK; generative AI; task-based learning; quasi-experimental study; higher education; Uzbekistan

Received: 20/09 /2026

Accepted :07/10/2026

Published: 17/10/2026

Abstract

Background: English for Specific Purposes (ESP) courses for non-philology students often focus on terminology and grammar but pay little attention to students' ability to think with professional concepts in a foreign language.

Purpose: This study proposed a five-criterion model of linguocognitive competence and evaluated a methodology for developing it that integrates technological, pedagogical and content knowledge (TPACK), a staged sequence of receptive, productive and creative tasks, structured reflection and AI-supported digital tools.

Method: A quasi-experimental pre-test–post-test design with a non-equivalent control group was conducted with 320 applied psychology students at three pedagogical universities in Uzbekistan (experimental group $n = 162$; control group $n = 158$). Linguistic, cognitive, professional-communicative, technological and creative-reflective competence were assessed with performance tasks and an analytic rubric.

Results: The groups were equivalent at baseline ($\chi^2(2) \leq 0.44$, $V \leq 0.04$). At the post-test, the experimental group outperformed the control group on all criteria ($\chi^2(2) = 44.67$ – 54.01 , $p < 0.001$, Cramér's $V = 0.37$ – 0.41). Between 77.8% and 79.0% of experimental students reached at least the intermediate level, compared with 38.6–41.1% of control students (Mantel–Haenszel OR = 5.01–5.99), with no heterogeneity across institutions.

Conclusion: ESP instruction can develop linguocognitive competence when tasks progress from understanding to creating, every lesson includes reflection, and digital and AI tools are selected for specific learning activities under teacher guidance.

1. Introduction

Students who learn English for Specific Purposes (ESP) in non-philological programmes rarely need English as an end in itself. A future psychologist, historian or chemistry teacher needs to read a professional text, recognise the concepts behind its terms, reason about a case and explain that reasoning to others. In many universities, however, ESP instruction still concentrates on lists of

terminology and controlled grammar practice. Students may recognise a term such as *anxiety* or *cognitive load*, yet be unable to use it to explain why a person behaves in a certain way. The gap lies less in vocabulary size than in the ability to think with professional concepts in a foreign language.

We refer to this ability as *linguocognitive competence*: the integrated

capacity to understand and use language units as carriers of professional concepts, to perform cognitive operations such as classifying, inferring and evaluating through the foreign language, and to reflect on one's own language and thinking. The construct draws on the communicative turn in language education, which extended the notion of competence from knowledge of a language system to its appropriate use in social contexts (Hymes, 1972), and on cognitive linguistics, which treats meaning as conceptualisation grounded in human experience (Evans & Green, 2006; Lakoff & Johnson, 1980). It is also consistent with the view that thought and speech develop in close interaction (Vygotsky, 1986).

At the same time, digital technologies, and generative artificial intelligence (AI) in particular, have changed the conditions of language learning. AI tools can provide instant explanations, model texts and feedback, but they can also encourage students to outsource thinking (Kasneci et al., 2023). International guidance therefore stresses a human-centred approach in which teachers remain responsible for how AI is used in education (UNESCO, 2023). The Technological Pedagogical Content Knowledge (TPACK) framework offers a lens for designing such instruction, because it treats effective technology use as the product of technological, pedagogical and

content knowledge working together (Mishra & Koehler, 2006), and it has recently been reconsidered in the light of generative AI (Mishra et al., 2023).

In Uzbekistan, the expansion of foreign language education has been a national policy priority (President of the Republic of Uzbekistan, 2021), and ESP courses such as “Foreign Language in the Professional Field” are part of non-philological programmes at pedagogical universities. Local studies have explored content and language integration and corpus-based instruction for non-philology students (Pulatova, 2026; Tashmatova, 2025), but empirical evidence on methods that explicitly target the cognitive side of ESP learning, and that combine it with digital and AI-supported tools, remains limited. In addition, results obtained in a single institution make it difficult to judge whether an effect is due to the method or to local conditions.

This study addresses these gaps. It describes a TPACK-based integrative digital methodology for developing linguocognitive competence and evaluates it in a quasi-experimental study involving 320 applied psychology students at three pedagogical universities. The study was guided by three research questions:

RQ1. How can linguocognitive competence be operationalised for ESP instruction in a way that allows

teachers to observe and assess it?

RQ2. Do students taught with the TPACK-based integrative methodology reach higher levels of linguocognitive competence than students taught with conventional ESP methods?

RQ3. Is the effect consistent across competence criteria and across institutions?

2. Literature review

2.1. From linguistic to linguocognitive competence

The concept of competence in language education has expanded considerably since Chomsky (1965) described linguistic competence as a speaker's internal knowledge of a language system. Hymes (1972) argued that knowing a language also means knowing when, where and how to use it, and introduced the notion of communicative competence. The Common European Framework of Reference for Languages and its Companion Volume further describe language users as social agents who use strategies and mediate meaning in real tasks (Council of Europe, 2020). These developments shifted attention from knowledge about language to the use of language for purposeful action.

Cognitive linguistics contributed a different but compatible perspective. It

treats meaning as conceptualisation and grammar as a means of construing experience (Evans & Green, 2006; Langacker, 2008). Lakoff and Johnson (1980) showed that much of everyday thinking is organised through conceptual metaphors, which suggests that learning a word in a professional field involves entering a network of related concepts rather than memorising an isolated label. From a pedagogical standpoint, this means that ESP learners need tasks in which terms are categorised, related to one another and used to explain professional phenomena.

A third strand concerns metacognition and self-regulation. Flavell (1979) described metacognition as knowledge about and regulation of one's own cognitive activity, and Zimmerman (2002) showed that self-regulated learners plan, monitor and evaluate their learning. Formative assessment research demonstrates that feedback during learning, when students are actively involved in judging their own work, improves achievement (Black & Wiliam, 1998). These ideas suggest that a reflective component should be part of any model of competence that aims to connect language and thinking.

Taken together, these perspectives support a multi-component view of linguocognitive competence. In this study it is described by five criteria: linguistic

(accurate and context-appropriate use of professional language), cognitive (understanding and relating concepts, analysing cases), professional-communicative (expressing and justifying ideas in professional situations), technological (purposeful use of digital tools for learning and communication) and creative-reflective (proposing alternatives and evaluating one's own performance).

2.2. ESP, content integration and task-based learning

ESP has traditionally been defined as an approach in which content and method are based on the learner's reasons for learning (Hutchinson & Waters, 1987). Course design therefore starts from an analysis of the target situations and texts that learners will face (Basturkmen, 2010). Content and language integrated learning (CLIL) shares this orientation and adds an explicit concern with cognition: its 4Cs framework links content, communication, cognition and culture (Coyle et al., 2010). Task-based language teaching contributes a further principle, namely that tasks with a primary focus on meaning and a clear outcome create favourable conditions for language development (Ellis, 2003).

The revised version of Bloom's taxonomy offers a practical way of sequencing cognitive demand, from remembering and understanding through

applying and analysing to evaluating and creating (Anderson & Krathwohl, 2001). In ESP practice, however, tasks often remain at the lower levels of this hierarchy. Students identify and translate terms, but are rarely asked to evaluate competing explanations of a professional case or to design a solution. One aim of the present methodology was to move systematically through the whole sequence within each lesson.

2.3. TPACK and AI-supported language learning

The TPACK framework describes the knowledge teachers need to integrate technology into teaching, focusing on the intersections of technological, pedagogical and content knowledge (Koehler & Mishra, 2009; Mishra & Koehler, 2006). Reviews show that TPACK has been widely used to design and evaluate technology-enhanced instruction, although its components are interpreted and measured in different ways (Chai et al., 2013; Voogt et al., 2013). Harris et al. (2009) recommended planning technology integration around learning activity types, so that the curriculum goal determines the choice of tool. In ESP, Tai et al. (2015) applied TPACK to design an online English writing course for nursing students, illustrating how the framework can link professional content with language skills and technology.

Research on AI in language education has grown rapidly. A review by Huang et al. (2023) mapped the trends and applications of AI in language learning, while an earlier systematic review of AI in higher education observed that many studies paid little attention to pedagogical perspectives and to the risks involved (Zawacki-Richter et al., 2019). With the arrival of large language models, researchers have emphasised both opportunities, such as personalised feedback and text generation, and challenges, such as over-reliance and inaccurate output (Kasneci et al., 2023). Jeon and Lee (2023) argue that the roles of human teachers and ChatGPT are complementary and that teachers remain essential in orchestrating learning. Mishra et al. (2023) suggest that generative AI requires teachers to develop new contextual knowledge about when and how such tools serve learning.

Despite this growing literature, few studies have combined a cognitively oriented model of ESP competence with a TPACK-based design that includes AI tools and tested it across several institutions. The present study was designed to contribute evidence of this kind.

3. Methodology

3.1. Design

The study used a quasi-experimental pre-test–post-test design with a non-equivalent control group (Shadish et al., 2002). Intact groups were assigned to the experimental condition (EG) or the control condition (CG) at each of three institutions. Groups were not randomised; instead, they were formed so that their previous-year ESP grades and academic activity were comparable. Baseline equivalence was then tested statistically before the intervention.

3.2. Participants and context

Participants were 320 third-year undergraduates majoring in applied psychology at the National Pedagogical University of Uzbekistan named after Nizami (NPUU, Tashkent), Bukhara State Pedagogical Institute (BSPI) and Kokand State Pedagogical University (KSPU). The intervention, including the pre-test and post-test, was conducted in the 2022–2023 academic year as part of a doctoral research project; all participants were in their third year of study throughout the experiment. The EG comprised 162 students and the CG 158 students (Table 1). All participants were taking the compulsory course “Foreign Language in the Professional Field”. Their English proficiency was mainly at CEFR levels A2–B1.

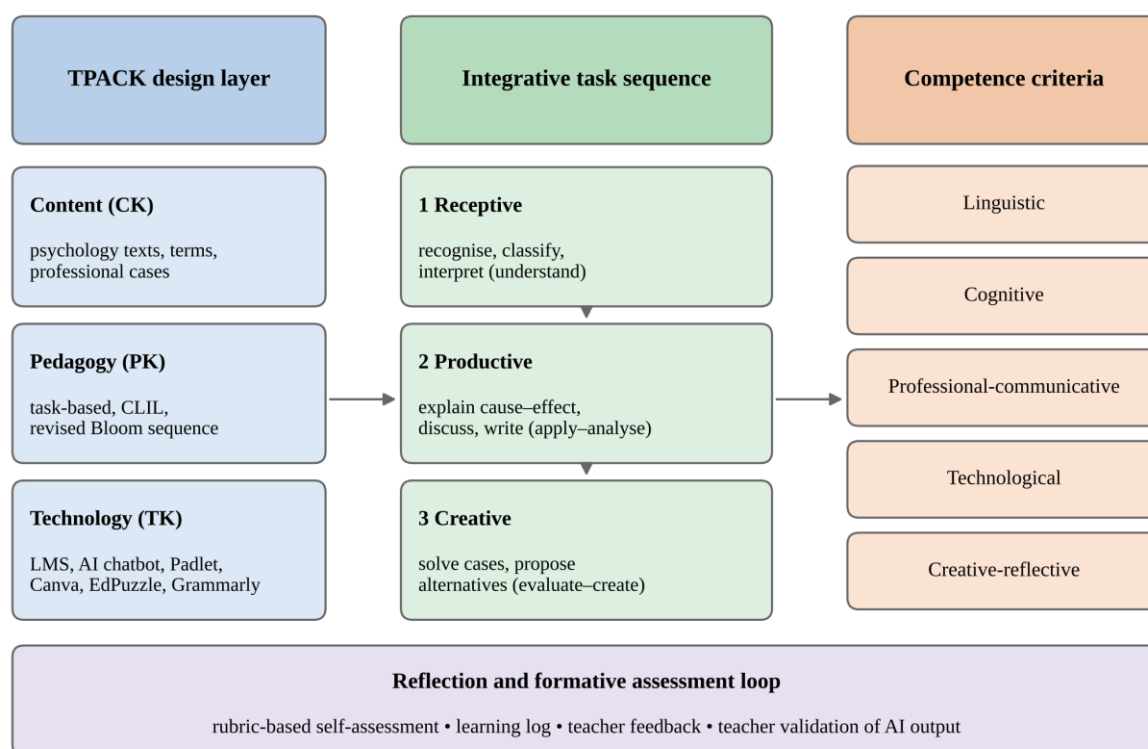
Table 1. *Participants by institution and group*

Institution	EG	CG	Total
NPUU (Tashkent)	54	53	107
BSPI (Bukhara)	55	54	109
KSPU (Kokand)	53	51	104
Total	162	158	320

3.3. Intervention

The EG was taught with the TPACK-based integrative digital methodology (Figure 1). Its design layer combined three types of knowledge. Content knowledge was represented by psychology-related texts, terminology and professional cases. Pedagogical knowledge was represented by task-based and CLIL principles and by a sequence of cognitive demand based on the

revised Bloom's taxonomy. Technological knowledge was represented by tools selected for specific activities: a learning management system, collaborative boards (Jamboard, Padlet), visual design tools (Canva), interactive video (EdPuzzle), writing and pronunciation support (Grammarly, ELSA Speak) and an AI chatbot.

**Figure 1.** *TPACK-based integrative methodology for developing linguocognitive competence*

Each lesson followed a three-stage task sequence. In the receptive stage, students worked with authentic or adapted

professional input and categorised language units according to the concepts they expressed. In the productive stage, they

used these units to explain cause–effect relations, discuss cases and write short texts. In the creative stage, they solved problems, proposed alternative solutions and presented them to peers. Every lesson ended with structured reflection, in which students assessed their work with a rubric and wrote a short learning-log entry. When AI tools were used, students first produced their own ideas and then compared them with the tool's output; the teacher checked the AI output for accuracy and adjusted its language to the students' level.

Two lessons illustrate the design. In the lesson “Daily Routines and Cognitive Triggers”, students read a short text about a student's morning and sorted its verbs and adjectives into physical actions (*wakes up, rushes*), cognitive actions (*reflects, plans*) and emotional states (*anxious, stressed*). They then listened to an audio recording and predicted the emotional consequences of each event, worked in pairs to formulate cause–effect statements with *because, so* and *when* (“If someone skips breakfast, how might they feel later?”), described a visual timetable (“When I miss the bus, I feel anxious and start worrying about being late”) and finally wrote a reflection on which part of their own routine causes the most stress. In the lesson “Planning the Day and Handling Challenges”, students brainstormed on a shared board, read a text on time management via a QR code,

watched an interactive video, created a “Plan – Obstacle – Strategy” diagram in Canva, developed a group plan on Padlet, wrote a short essay and presented their plan orally using an online calendar. Both lessons thus moved from recognising and classifying language to applying it and, finally, to creating and evaluating solutions.

Each group took part in eight lessons on the same topics. The CG studied the same topics with conventional ESP methods, in which language units were presented mainly at the level of form and meaning, with little attention to their cognitive and contextual content, and without the staged task sequence, the structured reflection or the digital tools described above. The volume of material and the assessment tasks were identical in both groups.

3.4. Instruments

Linguocognitive competence was assessed with three instruments. First, a diagnostic set of performance tasks was used: (a) semantic analysis of an authentic psychology text (identifying the main idea, relating concepts and summarising); (b) a situational dialogue based on a counselling scenario; (c) a cognitive mapping task in which students interpreted data presented in a table or chart; and (d) an alternative-solution task in which students proposed and compared solutions to a psychological

situation. The use of digital tools during these tasks provided evidence for the technological criterion. Second, a self-assessment questionnaire asked students to rate their own performance on each criterion on a five-point Likert scale. Third, teachers recorded observations during lessons.

On the basis of this evidence, each student was assigned to one of three levels (high, intermediate or low) on each of the five criteria using an analytic rubric (Table 2). The rubric was aligned with CEFR descriptors for levels A2–B2 and adapted to the professional content of psychology.

Table 2. *Analytic rubric for assessing linguocognitive competence*

Criterion	Indicators	High	Intermediate	Low
Linguistic	range and accuracy of vocabulary and grammar; choice of language suited to the psychological context	uses rich, accurate and topic-appropriate language	some errors; limited lexical range	frequent lexical and grammatical errors
Cognitive	understanding terms; relating concepts; analysing psychological situations	analyses with justification; clearly explains relations between concepts	partial understanding; superficial analysis	confuses concepts; vague explanations
Professional-communicative	communicating in professional situations; justifying ideas; correct use of terms	communicates freely in professional settings	expresses ideas, but errors affect communication	has difficulty communicating; ignores context
Technological	purposeful use of AI tools, online platforms and multimodal resources	uses digital tools independently and effectively	uses tools but needs guidance	rarely or ineffectively uses tools
Creative-reflective	creative solutions to problems; transfer to new contexts; analysis of own performance	thinks independently; develops new approaches	relies on models; partial reflection	reproductive performance; little reflection

Note. Levels were assigned on the basis of performance tasks, supported by self-assessment and lesson observation. Descriptors were aligned with CEFR levels A2–B2.

3.5. Data analysis

Because the outcome for each criterion was an ordinal level, the main analyses were based on frequency data. Differences between the EG and the CG in the distribution of levels were tested with Pearson's χ^2 test ($df = 2$), and effect sizes were expressed as Cramér's V , interpreted, because each comparison involved two groups ($\min(r - 1, c - 1) = 1$), with Cohen's (1988) benchmarks of 0.10 (small), 0.30

(medium) and 0.50 (large). As a complementary ordinal analysis, the Mann–Whitney U test was applied, and the effect size r was computed as $z/\sqrt{N}$. To estimate a practically meaningful contrast, the outcome was also dichotomised into “at least intermediate” versus “low”. For this outcome, we report proportions, risk differences with Newcombe hybrid-score 95% confidence intervals and odds ratios with 95% confidence intervals. To account

for clustering by institution, the Cochran–Mantel–Haenszel (CMH) test and the Mantel–Haenszel pooled odds ratio were computed, and homogeneity of odds ratios across institutions was examined with the Woolf test (Agresti, 2013). For comparability with national grading practice, the three levels were also coded as 3 (low), 4 (intermediate) and 5 (high), and group means and standard deviations were calculated. With five criteria, a Bonferroni-adjusted significance level of $\alpha = 0.01$ was applied.

3.6. Ethical considerations

The study was conducted in accordance with the ethical principles of the Declaration of Helsinki as part of a doctoral research project. The study protocol was reviewed and approved by the ethics committee of the National Pedagogical University of Uzbekistan named after Nizami, and permission to collect data was obtained from the participating institutions. All participants were informed of the aims and procedures of the study and gave informed consent. Participation was voluntary and did not affect course grades,

and all data were anonymised before analysis.

4. Results

4.1. Baseline equivalence

At the pre-test, about three quarters of the students in both groups were at the low level on every criterion (EG: 73.5–74.7%; CG: 73.4–74.1%), and only 4.3–5.7% reached the high level. The distribution of levels did not differ between groups on any criterion ($\chi^2(2) = 0.00–0.44$, $p \geq 0.80$, Cramér's $V \leq 0.04$). The mean coded score was 3.30–3.32 in both groups. The groups can therefore be regarded as equivalent at baseline with respect to the outcome measures (Figure 2, left panel).

4.2. Post-test differences between groups

After the intervention, the distributions changed markedly in the EG and only slightly in the CG (Figure 2, right panel). In the EG, the proportion of low-level students fell to 21.0–22.2%, while 56.8–58.0% reached the intermediate level and 20.4–21.6% the high level. In the CG, 58.9–61.4% of students remained at the low level, and 9.5–10.8% reached the high level.

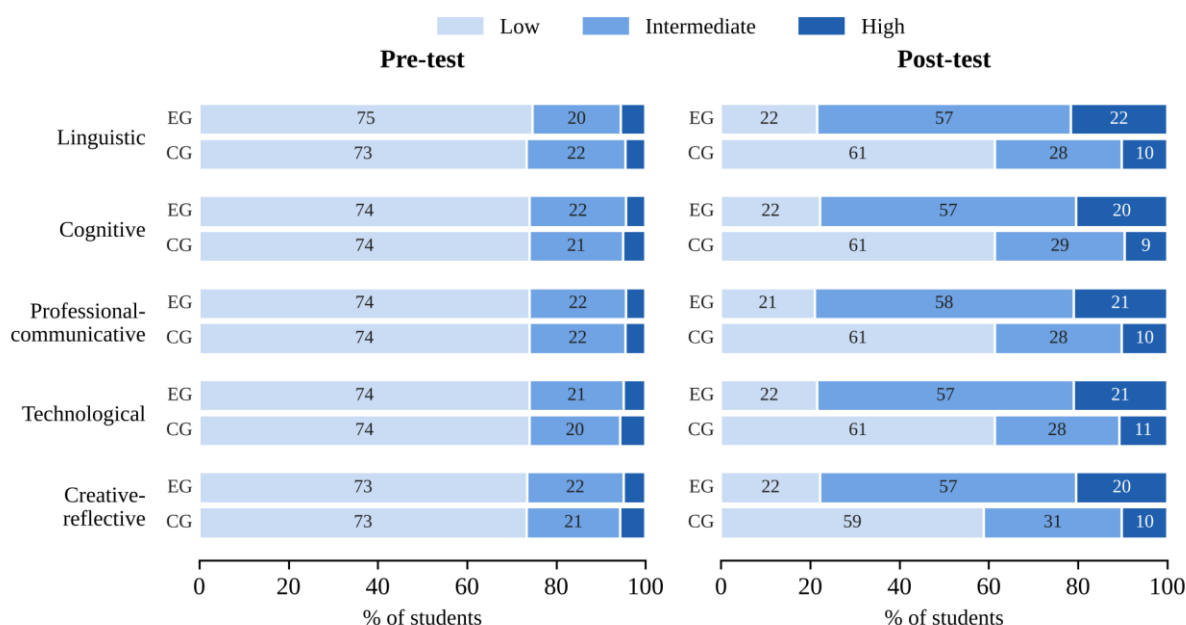


Figure 2. Distribution of competence levels in the experimental (EG, $n = 162$) and control (CG, $n = 158$) groups before and after the intervention

Table 3 summarises the post-test comparisons. Differences were significant on all five criteria ($\chi^2(2) = 44.67$ – 54.01 , all $p < 0.001$), and they remained significant after Bonferroni adjustment. Effect sizes were medium (Cramér's $V = 0.37$ – 0.41). The Mann–Whitney tests led to the same

conclusion ($z = 6.25$ – 6.81 , $r = 0.35$ – 0.38). Expressed as probability of superiority, a randomly chosen EG student had a level higher than a randomly chosen CG student in about 70% of comparisons (ties counted as half).

Table 3. Post-test distribution of competence levels by group

Criterion	Group	High n (%)	Intermediate n (%)	Low n (%)	M (SD)	$\chi^2(2)$	V
Linguistic	EG	35 (21.6)	92 (56.8)	35 (21.6)	4.00 (0.66)	52.28***	0.40
	CG	16 (10.1)	45 (28.5)	97 (61.4)	3.49 (0.67)		
Cognitive	EG	33 (20.4)	93 (57.4)	36 (22.2)	3.98 (0.65)	50.58***	0.40
	CG	15 (9.5)	46 (29.1)	97 (61.4)	3.48 (0.67)		
Professional-communicative	EG	34 (21.0)	94 (58.0)	34 (21.0)	4.00 (0.65)	54.01***	0.41
	CG	16 (10.1)	45 (28.5)	97 (61.4)	3.49 (0.67)		
Technological	EG	34 (21.0)	93 (57.4)	35 (21.6)	3.99 (0.65)	52.27***	0.40
	CG	17 (10.8)	44 (27.8)	97 (61.4)	3.49 (0.68)		
Creative-reflective	EG	33 (20.4)	93 (57.4)	36 (22.2)	3.98 (0.65)	44.67***	0.37
	CG	16 (10.1)	49 (31.0)	93 (58.9)	3.51 (0.67)		

Note. EG $n = 162$, CG $n = 158$. M (SD) computed with levels coded 3 = low, 4 = intermediate, 5 = high. V = Cramér's V. *** $p < 0.001$.

4.3. Practical significance and consistency across institutions

The dichotomised analysis shows the practical size of the effect (Table 4). On each criterion, 77.8–79.0% of EG students reached at least the intermediate level,

compared with 38.6–41.1% of CG students. The risk difference was 36.6–40.4 percentage points, and the lower bound of each 95% confidence interval exceeded 26 points. Odds ratios ranged from 5.01 to 5.99.

Table 4. Proportion of students reaching at least the intermediate level at the post-test

Criterion	EG ≥ intermediate, %	CG ≥ intermediate, %	Risk difference, pp [95% CI]	OR [95% CI]	MH OR [95% CI]	Woolf Q (p)
Linguistic	78.4	38.6	39.8 [29.4, 49.0]	5.77 [3.53, 9.44]	5.83 [3.55, 9.56]	0.05 (0.98)
Cognitive	77.8	38.6	39.2 [28.7, 48.4]	5.57 [3.41, 9.08]	5.57 [3.41, 9.10]	0.04 (0.98)
Professional-communicative	79.0	38.6	40.4 [30.0, 49.5]	5.99 [3.65, 9.83]	5.99 [3.65, 9.83]	0.00 (1.00)
Technological	78.4	38.6	39.8 [29.4, 49.0]	5.77 [3.53, 9.44]	5.79 [3.54, 9.48]	0.06 (0.97)
Creative-reflective	77.8	41.1	36.6 [26.2, 46.0]	5.01 [3.08, 8.15]	5.01 [3.08, 8.15]	0.06 (0.97)

Note. pp = percentage points; CI for the risk difference: Newcombe hybrid score method; MH OR = Mantel–Haenszel odds ratio stratified by institution; Woolf Q tests homogeneity of odds ratios across the three institutions (df = 2).

Institution-level analyses produced similar results. At each of the three universities, EG–CG differences were significant on every criterion ($\chi^2(2) = 14.35$ – 19.14 , all $p < 0.001$; Cramér's $V = 0.36$ – 0.42). After stratification by institution, the CMH tests remained highly significant ($\chi^2(1) = 44.20$ – 53.51 , $p < 0.001$),

and the Mantel–Haenszel pooled odds ratios (5.01–5.99) were almost identical to the crude estimates. The Woolf tests showed no evidence of heterogeneity across institutions ($Q = 0.00$ – 0.06 , $p \geq 0.97$). Figure 3 displays the institution-specific and pooled odds ratios.

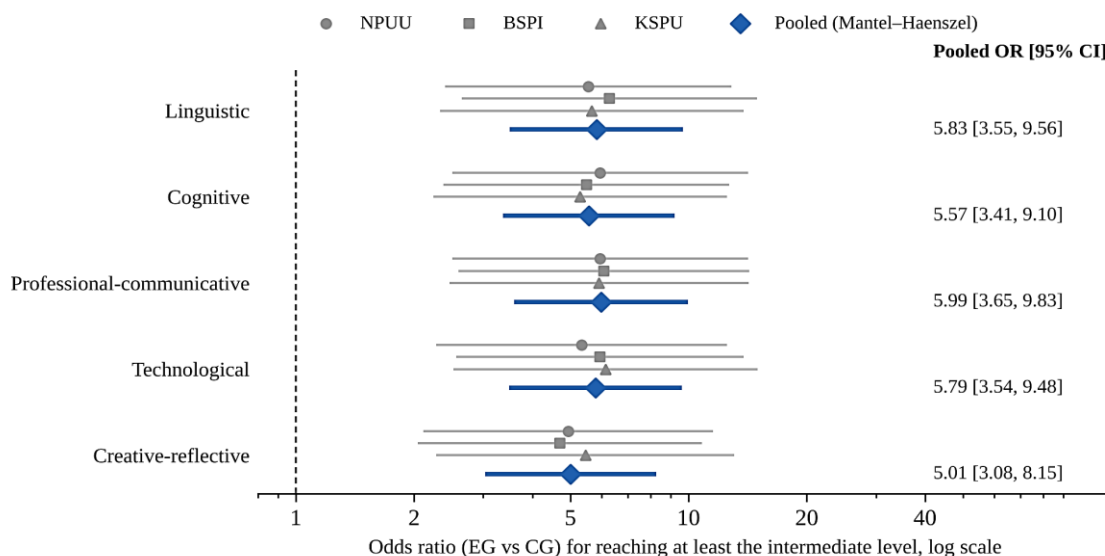


Figure 3. Institution-specific and pooled odds ratios for reaching at least the intermediate level

The effect was slightly smaller for the creative-reflective criterion ($V = 0.37$; pooled OR = 5.01) than for the linguistic and professional-communicative criteria ($V = 0.40$ – 0.41 ; pooled OR = 5.83–5.99). In terms of the coded scores, EG means at the post-test were 3.98–4.00 and CG means 3.48–3.51.

5. Discussion

5.1. Operationalising linguocognitive competence (RQ1)

The first question concerned how linguocognitive competence could be made observable. The five-criterion rubric proved workable for teachers: it described each level in terms of what a student does with professional language, and it could be applied to ordinary classroom tasks. The four diagnostic task types correspond to different points on the revised Bloom's taxonomy, from understanding a text to

proposing alternative solutions (Anderson & Krathwohl, 2001), so the instrument captured both language use and the cognitive operations performed through language. This responds to the concern that competence frameworks remain abstract unless they are linked to tasks and evidence (Council of Europe, 2020).

The almost identical results on the five criteria are noteworthy. They suggest that the components develop together rather than separately, which is consistent with the integrated view of content, communication and cognition in CLIL (Coyle et al., 2010). In the semantic categorisation task, for example, students could not succeed linguistically without understanding the concepts behind the words, and they could not explain their categorisation without professional-communicative skills. Integrated tasks of this kind may be one reason why the effect was spread evenly

across criteria. An alternative explanation is that a teacher's rating on one criterion influenced the ratings on others (a halo effect); this possibility cannot be excluded without independent raters.

5.2. Effectiveness of the methodology (RQ2)

The second question asked whether the TPACK-based methodology led to higher levels of competence than conventional teaching. The answer is positive. The EG outperformed the CG on all criteria with medium effect sizes, and about four out of five EG students reached at least the intermediate level, compared with about two out of five CG students. Because baseline distributions were virtually identical, these differences cannot be explained by initial group differences in the measured outcomes.

Several features of the methodology may account for the effect. First, the three-stage sequence ensured that each language unit was used repeatedly at increasing levels of cognitive demand within a single lesson, which aligns with task-based principles that emphasise meaning and outcome (Ellis, 2003). Second, technology was selected for specific activities rather than added as decoration, following the activity-type approach of Harris et al. (2009). Third, every lesson ended with structured reflection, which supports metacognitive

monitoring (Flavell, 1979) and self-regulated learning (Zimmerman, 2002) and provides the kind of formative feedback shown to improve achievement (Black & Wiliam, 1998). Finally, the rule that students formulate their own ideas before consulting an AI chatbot was intended to keep cognitive work with the learner, addressing the risk of over-reliance identified by Kasneci et al. (2023) and reflecting the complementary roles of teacher and AI described by Jeon and Lee (2023).

5.3. Consistency across criteria and institutions (RQ3)

The third question concerned consistency. The effect was observed at all three institutions, with nearly identical odds ratios and no evidence of heterogeneity. Since the institutions are located in different regions of the country, this consistency strengthens the argument that the effect is linked to the methodology rather than to a single teacher or group.

Across criteria, the creative-reflective component showed the smallest, though still substantial, effect. This is plausible: proposing and justifying alternative solutions in a foreign language demands both linguistic resources and metacognitive skills, which may develop more slowly than vocabulary and comprehension, particularly for learners at A2–B1. Teachers

may therefore need to provide more scaffolding for creative tasks, such as sentence starters, model answers and regular practice in reflective writing.

5.4. Implications

For ESP practice, the findings suggest three implications. First, ESP courses for non-philology students should include tasks at all levels of cognitive demand and should not stop at term recognition. Second, digital and AI tools are most useful when they are tied to a specific learning activity and competence criterion, and when students are required to evaluate rather than simply accept AI output. Third, teacher education should prepare ESP teachers to design such lessons, which requires the integrated knowledge described by TPACK and, in the age of generative AI, the ability to validate AI-generated content (Mishra et al., 2023; UNESCO, 2023).

6. Limitations and future research

The study has several limitations. First, groups were not randomised, so unmeasured differences between them cannot be fully excluded, although baseline equivalence on all outcome criteria was confirmed. Second, all participants were applied psychology students at pedagogical universities, which limits generalisation to other disciplines and institution types. Third, competence levels were assigned on the basis of performance tasks supported by

self-assessment and observation, but inter-rater reliability was not formally estimated. Fourth, the intervention combined several elements, including the task sequence, reflection and digital tools, so their separate contributions cannot be determined. Finally, outcomes were measured immediately after the intervention, and retention was not assessed.

Future research could use randomised or cluster-randomised designs, estimate inter-rater reliability, and compare versions of the methodology with and without AI tools in order to isolate their effect. Studies with students from other non-philological fields and delayed post-tests would also help to establish the generality and durability of the findings.

7. Conclusion

This study proposed a five-criterion model of linguocognitive competence and a TPACK-based integrative digital methodology for developing it in ESP courses, and evaluated the methodology with 320 applied psychology students at three Uzbek pedagogical universities. Groups that were equivalent at baseline differed markedly after the intervention: about 78–79% of experimental students reached at least the intermediate level, compared with about 39–41% of control students, with medium effect sizes and consistent results across institutions and

criteria. The findings indicate that ESP instruction can develop students' ability to think with professional concepts in a foreign language when tasks move systematically from understanding to creating, when every lesson includes structured reflection, and when digital and AI tools are chosen for specific learning purposes and used under teacher guidance.

Declarations

Funding. This research received no specific grant from any funding agency in the public, commercial or not-for-profit sectors. It was carried out within the research plan of the National Pedagogical University of Uzbekistan named after Nizami (2020–2024) as part of the PhD research of G. A. Allamuratov, “Methodology for developing students’ linguocognitive competencies” (13.00.02 – Theory and methodology of education and upbringing (English); registered with the Higher Attestation Commission of the Republic of Uzbekistan under No. B2025.2PhD/Ped9949), supervised by G. O. Maqsudova.

Conflict of interest. The authors declare no conflict of interest. G. O. Maqsudova was the doctoral supervisor of G. A. Allamuratov.

Ethics approval and consent to participate. See Section 3.6.

Data availability. Aggregated frequency data underlying Tables 3 and 4 are reported in the article. Anonymised student-level data are available from the authors on reasonable request.

Use of AI tools. During the preparation of this manuscript, the authors used Claude (Anthropic), ChatGPT (OpenAI) and Gemini (Google) to support language editing, translation and the checking of statistical calculations. The authors reviewed and edited

all output and take full responsibility for the content of the publication.

References

1. Agresti, A. (2013). *Categorical data analysis* (3rd ed.). Wiley.
2. Anderson, L. W., & Krathwohl, D. R. (Eds.). (2001). *A taxonomy for learning, teaching, and assessing: A revision of Bloom’s taxonomy of educational objectives*. Longman.
3. Basturkmen, H. (2010). *Developing courses in English for specific purposes*. Palgrave Macmillan.
4. Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy & Practice*, 5(1), 7–74. <https://doi.org/10.1080/0969595980050102>
5. Chai, C. S., Koh, J. H. L., & Tsai, C.-C. (2013). A review of technological pedagogical content knowledge. *Educational Technology & Society*, 16(2), 31–51.
6. Chomsky, N. (1965). *Aspects of the theory of syntax*. MIT Press.
7. Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Lawrence Erlbaum Associates.
8. Council of Europe. (2020). *Common European Framework of Reference for Languages: Learning, teaching, assessment – Companion volume*. Council of Europe Publishing.
9. Coyle, D., Hood, P., & Marsh, D. (2010). *CLIL: Content and language integrated learning*. Cambridge University Press.
10. Ellis, R. (2003). *Task-based language learning and teaching*. Oxford University Press.
11. Evans, V., & Green, M. (2006). *Cognitive linguistics: An introduction*. Edinburgh University Press.
12. Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new area of cognitive–developmental inquiry.

- American Psychologist*, 34(10), 906–911. <https://doi.org/10.1037/0003-066X.34.10.906>
13. Harris, J., Mishra, P., & Koehler, M. (2009). Teachers' technological pedagogical content knowledge and learning activity types: Curriculum-based technology integration reframed. *Journal of Research on Technology in Education*, 41(4), 393–416. <https://doi.org/10.1080/15391523.2009.10782536>
14. Huang, X., Zou, D., Cheng, G., Chen, X., & Xie, H. (2023). Trends, research issues and applications of artificial intelligence in language education. *Educational Technology & Society*, 26(1), 112–131. [https://doi.org/10.30191/ETS.202301_26\(1\).0009](https://doi.org/10.30191/ETS.202301_26(1).0009)
15. Hutchinson, T., & Waters, A. (1987). *English for specific purposes: A learning-centred approach*. Cambridge University Press.
16. Hymes, D. (1972). On communicative competence. In J. B. Pride & J. Holmes (Eds.), *Sociolinguistics* (pp. 269–293). Penguin.
17. Jeon, J., & Lee, S. (2023). Large language models in education: A focus on the complementary relationship between human teachers and ChatGPT. *Education and Information Technologies*, 28, 15873–15892. <https://doi.org/10.1007/s10639-023-11834-1>
18. Kasneci, E., Seßler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Günemann, 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, 102274. <https://doi.org/10.1016/j.lindif.2023.102274>
19. Koehler, M. J., & Mishra, P. (2009). What is technological pedagogical content knowledge? *Contemporary Issues in Technology and Teacher Education*, 9(1), 60–70.
20. Lakoff, G., & Johnson, M. (1980). *Metaphors we live by*. University of Chicago Press.
21. Langacker, R. W. (2008). *Cognitive grammar: A basic introduction*. Oxford University Press.
22. Mishra, P., & Koehler, M. J. (2006). Technological pedagogical content knowledge: A framework for teacher knowledge. *Teachers College Record*, 108(6), 1017–1054. <https://doi.org/10.1111/j.1467-9620.2006.00684.x>
23. Mishra, P., Warr, M., & Islam, R. (2023). TPACK in the age of ChatGPT and generative AI. *Journal of Digital Learning in Teacher Education*, 39(4), 235–251. <https://doi.org/10.1080/21532974.2023.2247480>
24. President of the Republic of Uzbekistan. (2021). *On measures to bring the activities for popularising the study of foreign languages in the Republic of Uzbekistan to a qualitatively new level* (Resolution No. PQ-5117, 19 May 2021). <https://lex.uz>
25. Pulatova, D. R. (2026). *Korpus ta'lim texnologiyalari asosida nofilologik ta'lim yo'nalishi talabalarining kasbiy kommunikativ kompetensiyasini rivojlantirish metodikasini takomillashtirish* [Improving the methodology for developing the professional communicative competence of non-philology students on the basis of corpus technologies] [Doctoral dissertation abstract]. Tashkent.
26. Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Houghton Mifflin.

27. Tai, H.-C., Pan, M.-Y., & Lee, B.-O. (2015). Applying technological pedagogical and content knowledge (TPACK) model to develop an online English writing course for nursing students. *Nurse Education Today*, 35(6), 782–788. <https://doi.org/10.1016/j.nedt.2015.02.016>
28. Tashmatova, G. R. (2025). *Nofilologik yo 'nalish talabalarining kasbiy kommunikativ kompetensiyasini CLIL texnologiyasi asosida rivojlantirish metodikasi* [Methodology for developing the professional communicative competence of non-philology students on the basis of CLIL] [PhD dissertation abstract]. Fergana.
29. UNESCO. (2023). *Guidance for generative AI in education and research*. UNESCO. <https://doi.org/10.54675/EWZM9535>
30. Voogt, J., Fisser, P., Pareja Roblin, N., Tondeur, J., & van Braak, J. (2013). Technological pedagogical content knowledge – a review of the literature. *Journal of Computer Assisted Learning*, 29(2), 109–121. <https://doi.org/10.1111/j.1365-2729.2012.00487.x>
31. Vygotsky, L. S. (1986). *Thought and language* (A. Kozulin, Trans.). MIT Press.
32. 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, 39. <https://doi.org/10.1186/s41239-019-0171-0>
33. Zimmerman, B. J. (2002). Becoming a self-regulated learner: An overview. *Theory Into Practice*, 41(2), 64–70. https://doi.org/10.1207/s15430421tip4102_2