International Journal of Social Science, Education and Humanities
ISSN Print: 3104-4239
ISSN Online: 3104-4247
CD ISSN: 3104-4255
https://doi.org/10.63313/IJSSEH.9002

AI-Empowered Pedagogy in ESP Curriculum: A Constructivist Approach to Active Learning

Xiaoxuan Yang
School of English Studies, Zhejiang Yuexiu University, Shaoxing 312030, China
Email: 20251024@zyufl.edu.cn

How to cite this paper: Yang, X. X. (2025). AI-Empowered Pedagogy in ESP Curriculum: A Constructivist Approach to Active Learning. International Journal of Social Science, Education and Humanities, 1(1), 7-23. ISSN Print: 3104-4239; ISSN Online: 3104-4247.
https://doi.org/10.63313/IJSSEH.9002
Published: 2025-07-07

Abstract

This study empirically investigates the integration of domestic AI tools (iFlytek Spark, iWrite, Pigai.org) within a constructivist framework for English for Spe-cific Purposes (ESP) technical writing instruction for Chinese engineering un-dergraduates. Employing a sequential explanatory mixed-methods design (qua-si-experimental, N=120), it examines AI's impact on writing proficiency, cogni-tive efficiency, metacognition, and collaboration. Quantitative results show the AI-integrated group achieved significantly greater gains in overall writing profi-ciency ($F(1,117)=28.37$, *p*<.001, $\eta^2=0.195$), particularly in audience adap-ta-tion (*d*=1.42) and genre conventions (*d*=1.37), alongside 25.4% faster task completion (*p*<.001), 41.7% fewer revisions (*p*<.001), enhanced meta-cog-nitive awareness (*d*=1.15), and 68% more substantive peer collaboration (*p*<.001). Qualitative analysis reveals AI functioned as a "discourse catalyst," deepening collaborative knowledge construction through analytical prompts and fostering critical engagement with feedback. Domain-specific limitations highlight the necessity for human-AI orchestration. The study proposes a peda-gogical model positioning AI for cognitive load management, learners for au-thentic problem-solving, instructors as cognitive coaches, and structured col-laborative knowledge building. Implications emphasize customizable platforms, faculty TPACK development, and cross-domain research.

Keywords

Artificial Intelligence (AI); English for Specific Purposes (ESP); Constructivist pedagogy; Collaborative learning; Technical writing; Mixed-methods research

1. Introduction

The accelerating convergence of sophisticated artificial intelligence (AI) technologies with established pedagogical methodologies has generated transformative potential for specialized language instruction, particularly within the demanding field of English for Specific Purposes (ESP). ESP, as a distinct pedagogical discipline, is fundamentally concerned with equipping learners with context-dependent communication competencies tailored to specific professional, academic, or vocational domains. This specificity necessitates instruction that transcends generic language skills, focusing instead on the mastery of discipline-specific genres, terminologies, discourse conventions, and communicative strategies relevant to the target context. Within this framework, AI technologies offer unprecedented opportunities for simulating authentic professional environments and providing adaptive, immediate, and highly contextualized learning experiences that were previously unattainable through conventional means. The imperatives of Industry 4.0, characterized by digitalization, automation, and globalized knowledge exchange, have dramatically accelerated the need for graduates possessing not just general English proficiency, but demonstrable fluency in the specialized English registers of their chosen fields. This pressing demand highlights a critical shortcoming: traditional ESP instruction often remains constrained by static materials, standardized assessment frameworks, and limited capacity for providing individualized, iterative practice within realistic scenarios (Hyland 133). This pedagogical limitation becomes particularly acute and evident in technical writing courses, where nuanced genre conventions, sophisticated audience awareness, rhetorical precision, and the iterative nature of drafting and revision require situated, experiential practice that conventional teaching methods, often burdened by large class sizes and resource limitations, struggle to provide efficiently and at scale.

Constructivist learning theory provides a robust and highly relevant theoretical foundation for reimagining ESP instruction in the AI era. Rooted in the work of Piaget and Vygotsky, constructivism fundamentally posits that knowledge is not passively received but actively constructed by learners through meaningful social interaction, authentic problem-solving, and reflection upon experience (Vygotsky 86). Within ESP, this translates to pedagogical approaches where learners engage in tasks mirroring the communicative demands of their target professional environments, constructing their understanding of discipline-specific discourse through guided participation and collaborative knowledge building. When strategically aligned with AI's inherent capacity for providing immediate, granular feedback, generating complex scenario simulations, and facilitating structured collaboration, core constructivist principles—such as scaffolding, situated learning, and social negotiation (Luckin et al. 32). Current scholarship exploring this intersection, however, reveals significant and persistent gaps in our understanding. Specifically, there is limited empirical investigation into how AI tools functionally mediate the complex social learning processes central to constructivism within ESP contexts, or how they contribute to the development of crucial metacognitive awareness—learners' ability to plan, monitor, evaluate, and regulate their own learning and writing processes. Furthermore, the theoretical implications of positioning AI explicitly within sociocultural frameworks like Vygotsky's, particularly concerning the nature of mediation and the Zone of Proximal Development (ZPD), remain underexplored. This study directly addresses these critical lacunae through a comprehensive empirical investigation of an AI-integrated technical writing curriculum for engineering students. It examines not only the quantifiable learning outcomes but, crucially, delves into the pedagogical mechanisms and theoretical dynamics through which these outcomes are achieved, providing a richer understanding of the human-AI pedagogical partnership.

2. Theoretical Framework and Literature Review

2.1. Constructivist Foundations of ESP Pedagogy

The epistemological core of constructivism fundamentally challenges transmission models of learning. It posits that learners actively construct knowledge schemata through direct experience, interaction with the environment, and crucially, through social negotiation and discourse, rather than passively receiving pre-formed information from an instructor (Piaget 15). Lev Vygotsky's sociocultural extension of constructivism placed paramount emphasis on the critical role of social interaction, cultural tools (including language), and more knowledgeable others in cognitive development (Vygotsky 86). His concept of the Zone of Proximal Development (ZPD) represents the dynamic space between what a learner can achieve independently and what they can achieve with guidance and collaboration from a more capable peer or instructor. Effective pedagogy, according to this view, involves providing appropriate "scaffolding" within the ZPD—temporary support structures that enable learners to tackle challenging tasks—which are gradually withdrawn as competence increases.

In ESP contexts, these constructivist principles translate directly into pedagogical imperatives. Learners must engage not merely in abstract language exercises, but in authentic disciplinary tasks that reflect the communicative challenges they will face in their professional lives. This involves grappling with genuine textual genres (e.g., research proposals, technical reports, patient case notes, legal briefs), understanding specific audience expectations, and mastering the discourse conventions unique to their field. Dudley-Evans and St John contend that effective ESP instruction must meticulously mirror the communicative contexts, rhetorical purposes, and textual genres of the target professional environments (102). This enables learners to internalize discipline-specific discourse conventions not through rote memorization, but through guided participation and situated practice, essentially "apprenticing" into the discourse community. The role of the ESP instructor thus shifts from being the sole source of knowledge to becoming a facilitator and co-constructor of knowledge, designing authentic tasks and providing the necessary scaffolding to support learners' entry into their target discourse communities. Assessment within this framework must move beyond decontextualized grammar tests towards evaluating performance on authentic communicative tasks within simulated or real-world contexts

2.2. AI as a Meditational Tool in Language Learning

Artificial intelligence, particularly Natural Language Processing (NLP), is rapidly transforming its role within language education, transcending its conventional perception as merely an automated tutor or drill provider. Contemporary NLP-powered platforms function increasingly as sophisticated cognitive partners or mediational tools that scaffold learning in ways deeply resonant with constructivist and sociocultural principles (Ware and Warschauer 112). These tools scaffold learning through several primary mechanisms, each aligning with facets of constructivist pedagogy:

- Adaptive Feedback Systems: Advanced AI systems can diagnose errors not just at the surface level of grammar and spelling, but also at deeper syntactic, lexical (e.g., domain-specific terminology usage), pragmatic, and rhetorical levels (e.g., thesis clarity, argument structure, audience appropriateness). This feedback can be immediate and highly contextualized, allowing for rapid iteration and refinement. For instance, an AI system designed for engineering ESP can flag inappropriate verb tense use in a methods section or suggest more precise technical terminology.
- Collaborative Interfaces: AI platforms can facilitate distributed knowledge construction by structuring peer review processes, managing group workflows, analyzing contributions for equity, and prompting deeper levels of commentary and discussion. They can mediate collaboration by providing frameworks for feedback, highlighting areas of agreement or disagreement across multiple reviewers, and synthesizing comments. This transforms peer interaction from potentially superficial exchanges to focused, evidence-based discussions about writing choices.
- Simulation Environments: Perhaps most powerfully for ESP, AI can generate dynamic, realistic simulations of workplace communication scenarios. Learners can practice writing emails to simulated clients, drafting reports based on simulated datasets, or responding to simulated technical inquiries, all within a safe environment that provides immediate feedback on their performance. This capability provides the crucial contextual authenticity that constructivist approaches require but is often difficult to replicate consistently in a classroom.

These capabilities strongly align with David Jonassen's influential concept of "mindtools" – technologies that do not simply deliver information but engage learners in critical thinking, knowledge representation, and meaningful problem-solving. AI, when designed and implemented effectively, functions as such a mindtool in ESP. It allows learners to externalize their thinking (e.g., through drafting and revision cycles supported by feedback), model complex processes (e.g., structuring a research article), and engage in authentic knowledge construction within their specific domain. For ESP instruction, AI's unique ability to generate and respond to domain-specific scenarios (e.g., simulating the drafting of an engineering feasibility report, a medical discharge summary, or a business proposal) provides the contextual authenticity and situated practice opportunities that are central to constructivist approaches but often logistically challenging to provide.

2.3. Current Gaps in AI-ESP Integration Research

Despite the promising theoretical alignment and burgeoning technological capabilities, empirical studies specifically investigating AI integration within ESP contexts remain relatively nascent and often narrowly focused. A significant portion of existing research concentrates on measuring isolated linguistic outcomes, such as vocabulary acquisition, grammar accuracy, or surface-level error reduction, rather than examining holistic pedagogical integration and its impact on higher-order communicative competence. For example, Chen and Wang demonstrated the effectiveness of NLP tools in reducing surface-level errors in technical writing while also noting some positive impact on rhetorical awareness (1790). Warschauer documented significant efficiency gains in drafting processes through technology-mediated collaboration (153). While valuable, such studies often stop short of exploring the underlying pedagogical mechanisms or the broader theoretical implications.

Critical gaps persist in the current literature:

- Metacognitive Awareness: There is limited research examining AI's role in developing learners' metacognitive awareness within ESP. How does interacting with AI feedback influence learners' ability to plan, monitor, evaluate, and regulate their own writing processes and learning strategies? Does it foster critical engagement with feedback or passive acceptance?
- Social Learning Processes: While collaboration is sometimes studied, research rarely delves deeply into how AI tools functionally mediate social learning processes and collaborative knowledge construction. How does AI alter the dynamics of peer review? Does it facilitate more meaningful social negotiation of meaning within disciplinary contexts?
- Theoretical Framing: Existing studies often lack robust engagement with sociocultural learning theories. There is insufficient exploration of how AI functions as a mediational tool within Vygotskian frameworks, how it might redefine the ZPD, or how the concept of scaffolding applies when the "more knowledgeable other" is an algorithm. The theoretical implications of human-AI collaboration in knowledge construction are under-theorized.
- Contextual Specificity and Limitations: Research often underreports the limitations of AI tools, particularly concerning their performance across diverse ESP sub-genres and highly specialized domains. How do domain-specific nuances impact the accuracy and usefulness of AI feedback? What are the boundary conditions for effective AI integration in ESP?

This research directly addresses these significant gaps through its explicit grounding in sociocultural constructivist theory and its employment of a multidimensional methodology designed to capture not only learning outcomes but also the processes of metacognitive development, collaborative interaction, and the lived experience of learners and instructors within an AI-integrated ESP technical writing curriculum.

3. Methodology

3.1. Research Design

A sequential explanatory mixed-methods design was implemented across a 14-week technical writing course at a comprehensive Chinese university. The quantitative phase employed a quasi-experimental pretest-posttest control group design to measure learning outcomes, followed by qualitative investigation of learner experiences and pedagogical processes. Methodological triangulation strengthened validity while capturing both outcome metrics and the phenomenological dimensions of AI integration (Creswell and Plano Clark 73). The study proceeded in two distinct, interconnected phases:

- Quantitative Phase: Employed a quasi-experimental pretest-posttest control group design. This phase aimed to objectively measure differences in learning outcomes (writing proficiency, task efficiency, metacognitive awareness) between the experimental group (using AI tools) and the control group (conventional instruction).
- Qualitative Phase: Conducted following the quantitative phase, this phase involved in-depth qualitative investigation of learner experiences, pedagogical processes, instructor perspectives, and the dynamics of human-AI interaction within the experimental group. Its purpose was to provide rich explanations and contextual understanding for the quantitative findings and to explore the "how" and "why" questions related to the observed outcomes.

Methodological triangulation—using multiple data sources (tests, surveys, interviews, focus groups, artifacts, behavioral logs) and multiple methods (quantitative and qualitative)—was a core principle to enhance the validity, reliability, and depth of the findings. This approach allowed for capturing both measurable outcome metrics and the complex phenomenological dimensions of AI integration within the authentic classroom context.

3.2. Participants

Participants comprised 120 second-year undergraduate students enrolled in the "Technical Writing for Engineers" course. All participants were engineering majors (Mechanical, Civil, Electrical, Chemical Engineering). The gender distribution was 72 male (M) and 48 female (F), with a mean age of 20.3 years (SD = 0.87). All participants possessed intermediate English proficiency, verified by institutional placement tests and standardized as equivalent to IELTS band scores of 5.5 to 6.0, indicating they could cope with academic language but likely required support with complex

ESP tasks.

Participants were randomly assigned to either the experimental group (*n* = 60) or the control group (*n* = 60). This random assignment aimed to minimize selection bias and ensure group equivalence at the outset concerning language ability, major, and other potential confounding variables. Both groups received the same core curriculum content and contact hours (3 hours per week for 14 weeks), taught by instructors with similar qualifications and experience, who received standardized training on the core pedagogical approach to minimize instructor effect differences. The key distinction lay in the pedagogical tools and feedback mechanisms:

1) Control Group: Received conventional instruction

This involved textbook-based exercises, in-class lectures and workshops on technical writing principles, manual drafting and revision cycles, instructor-provided written feedback on assignments, and traditional face-to-face peer review sessions organized by the instructor.

2) Experimental Group

Utilized an integrated AI platform suite alongside core instruction. This suite featured:

- iFlytek Spark API: Integrated into the writing environment to provide real-time, contextual drafting assistance. This included domain-aware prompts, sentence completion suggestions tailored to engineering genres, terminology suggestions, and initial grammar/style checks during the drafting phase.
- iWrite English Writing System: Used for primary drafting, revision, and feedback. This system incorporates specific engineering-writing style rules (e.g., passive voice conventions in methods sections, precision in results description, conciseness in abstracts) and provides discipline-specific feedback on drafts, highlighting issues related to genre conventions, technical accuracy, clarity, and conciseness beyond basic grammar.
- Pigai.org Automated Peer-Review Analytics System: Structured the peer review process. Students submitted drafts to Pigai.org, which distributed them anonymously among peers. Crucially, Pigai.org analyzed the peer feedback itself, providing metrics on comment depth, specificity, focus (e.g., surface vs. content vs. organization), and prompting reviewers to elaborate on vague comments. It also generated comparative analytics on drafts (e.g., thesis statement clarity scores across versions).
- WELearn Digital Platform: Served as the central Learning Management System (LMS), hosting materials, assignments, and crucially, capturing detailed behavioral logs. These logs tracked time spent on tasks, frequency of accessing different AI tools, revision history, response time to feedback, and interactions within the peer review system, providing objective data on engagement and cognitive load proxies

3.3. Data Collection Instruments

1) Quantitative Measures:

- Technical Writing Proficiency Assessment: A custom-designed, analytically scored rubric was the primary outcome measure. This 25-point rubric assessed five critical dimensions of ESP technical writing (each dimension scored 0-5):
  - Audience Awareness: Appropriateness of tone, level of detail, and explanation for the target reader.
  - Genre Conventions: Adherence to structural, stylistic, and linguistic norms of the specific technical document type (e.g., lab report, proposal).
  - Technical Accuracy: Correct and precise use of domain-specific terminology, data presentation, and factual content.
  - Coherence: Logical flow of ideas, effective paragraphing, and use of cohesive devices.
  - Concision: Efficiency and clarity of expression, avoiding redundancy.

Two experienced ESP raters, blinded to group assignment, scored all pre- and post-test essays. Inter-rater reliability was high (Cohen's $\kappa = 0.87$); discrepancies were resolved through discussion. Pre-test and post-test tasks were parallel forms involving comparable technical writing prompts (e.g., summarizing a technical process, describing experimental apparatus).

- Task Efficiency Metrics: Derived from WELearn platform logs and draft version histories:
  - Time-on-task: Total time spent drafting and revising each major assignment.
  - Revision Cycles: Number of substantive revisions made per document before final submission.
  - Error Identification/Correction Rates: Measured on specific error correction tasks embedded in the curriculum (accuracy in identifying and correcting errors in sample texts).
- Metacognitive Awareness Inventory (MAI): A 28-item, 5-point Likert scale survey adapted from the well-validated instrument developed by Schraw and Dennison (1994, p. 460). This instrument measures key facets of self-regulated learning: Knowledge of Cognition (awareness of one's own thinking) and Regulation of Cognition (planning, monitoring, debugging, evaluating). Its reliability in this context was confirmed (Cronbach's $\alpha = 0.83$). Administered pre- and post-intervention.

2) Qualitative Measures:

- Semi-structured Interviews: Conducted post-intervention with a purposive sample of 24 experimental-group participants (representing different engineering majors and initial proficiency levels based on pre-test scores). Interviews (avg. 45 mins) explored experiences with the AI tools, perceptions of feedback usefulness, changes in writing strategies, perceptions of collaboration, and development of critical awareness regarding AI suggestions. All interviews were audio-recorded and transcribed verbatim.

- Focus Groups: Two focus groups were conducted with the 6 instructors involved in teaching the experimental sections (3 per group). These sessions (approx. 90 mins each) explored instructors' observations of student engagement, perceived strengths/weaknesses of the AI tools, changes in their teaching role, challenges encountered, and overall assessment of the integration. Sessions were audio-recorded and transcribed.
- Artifact Analysis: A detailed analysis of 180 AI-annotated drafts (3 drafts per experimental participant across different assignment types) was performed. This involved examining the nature of AI feedback (type, frequency, specificity), learner responses to feedback (acceptance, modification, rejection), revision patterns between drafts, and the evolution of writing quality and engagement with feedback over time.

Table 1. Research Design Overview.

Dimension	Experimental Group	Control Group	Data Collection
Core Pedagogy	Constructivist task-based + AI tools	Constructivist task-based	Instructor logs, lesson plans
Writing Tasks	8 AI-scaffolded technical documents	8 conventional assignments	Pre/post writing assessments (Rubric scored)
Drafting Tools	iWrite + iFlytek Spark integration	Word Processor	WELearn behavioral logs, Draft version histories
Feedback Source	Triangulated: AI (iWrite) + Peer (Pigai.org) + Instructor	Instructor-only written feedback	Feedback annotations, Revision histories, Interview/Focus group data
Collaboration	Pigai.org-mediated peer review with analytics	Manual peer review (instructor assigned/managed)	Pigai.org interaction logs, Peer feedback artifacts, Interview data
Duration	14 weeks (3 contact hrs/week)	14 weeks (3 contact hrs/week)	All process metrics (Time-on-task, Revision cycles)

3.4. Analytical Procedures

1) Quantitative Data Analysis

Pretest scores on the writing rubric and MAI were analyzed using independent samples t-tests to confirm initial group equivalence. Primary analysis of post-test outcomes employed Analysis of Covariance (ANCOVA) with the corresponding pretest scores as covariates. This statistically controls for initial differences and provides a more precise estimate of the intervention's effect on the post-test scores. Effect sizes were calculated using Cohen's *d* to quantify the magnitude of differences between groups. Task efficiency metrics were analyzed using independent samples t-tests comparing experimental and control group means. All quantitative analyses were performed using SPSS software (v28), with significance level set at $\alpha = 0.05$.

2) Qualitative Data Analysis

Interview and focus group transcripts, along with open-ended survey responses, underwent rigorous thematic analysis following the framework outlined by Braun and Clarke (2006). This involved familiarization with the data, generating initial codes, searching for themes, reviewing themes, defining and naming themes, and producing the analysis. NVivo software (v12) aided in data management and coding. Themes were identified both inductively (emerging from the data) and deductively (informed by the theoretical framework, e.g., scaffolding, metacognition, collaboration). Constant comparative analysis ensured consistency. Artifact analysis employed genre analysis techniques (Swales, 1990) to trace developmental trajectories in writing, focusing on changes in rhetorical moves, linguistic features, and responsiveness to different feedback sources over the sequence of drafts. Analyst triangulation was employed, with two researchers independently coding a subset of the qualitative data and comparing interpretations to enhance reliability.

3) Integration:

The sequential explanatory design prioritized using qualitative findings to explain and elaborate on the quantitative results. For example, quantitative findings showing significant gains in audience awareness were explored qualitatively through interview data on how AI tools helped students visualize and adapt to specific audiences. Similarly, patterns observed in artifact analysis regarding feedback engagement were triangulated with MAI scores and interview reflections on metacognitive development.

4. Results

4.1. Quantitative Findings

1) Writing Proficiency Development

ANCOVA results, controlling for pretest scores, revealed that the experimental group demonstrated statistically significant and substantially greater improvement in overall technical writing proficiency compared to the control group ($F(1,117) = 28.37$, $p < .001$, partial $\eta^2 = 0.195$). This large effect size ($\eta^2 > .14$) indicates that the AI-integrated approach explained nearly 20% of the variance in post-test scores beyond the variance explained by initial ability.

As detailed in Table 2, the gains were not uniform across all dimensions. Substantial and highly significant gains occurred in areas requiring deep contextual understanding and rhetorical skill:

- Audience Awareness: The experimental group significantly outperformed the control group (Adj. M = 4.82 vs. 3.91, F=18.29, *p*<.001, *d*=1.42). This very large effect size suggests AI tools, particularly contextual prompts and feedback on tone/clarity, effectively supported students in tailoring their writing to specific readers.
- Genre Conventions: Similarly strong gains were observed (Adj. M = 4.75 vs. 3.87, F=15.76, *p*<.001, *d*=1.37). The AI's genre-specific templates, style rules, and feedback helped students internalize structural and stylistic norms.
- Coherence: Significant improvement was also evident (Adj. M = 4.57 vs. 4.05, F=8.93, *p*<.01, *d*=0.81), indicating better organization and flow of ideas, likely aided by feedback on paragraph structure and linking devices.

- Technical Accuracy: While improvement occurred (Adj. M = 4.63 vs. 4.21, F=6.42, *p*<.05, *d*=0.73), the effect size was moderate. This suggests AI tools were helpful for terminology and factual precision, but perhaps less impactful than instructor feedback on highly specialized content nuances.
- Concision: Gains were present but not statistically significant at the p<.05 level in this analysis (data not shown in table for brevity, Adj. M Exp=4.64 vs Ctrl=4.33, F=3.21, p=.076).

Table 2. Writing Proficiency Gains by Dimension (Post-test Adjusted Means)

Assessment Dimension	Experimental Group M (SD)	Control Group M (SD)	F-value	Effect Size (Cohen's *d*)
Audience Awareness	4.82 (0.51)	3.91 (0.63)	18.29***	1.42 (very large)
Genre Conventions	4.75 (0.48)	3.87 (0.71)	15.76***	1.37 (very large)
Technical Accuracy	4.63 (0.56)	4.21 (0.59)	6.42*	0.73 (medium)
Coherence	4.57 (0.62)	4.05 (0.67)	8.93**	0.81 (large)
Composite Score	23.41 (1.82)	18.97 (2.15)	28.37*	1.92 (very large)

*Note: Maximum score per dimension = 5; *p*<.05, **p*<.01, ***p*<.001*

2) Cognitive Efficiency Metrics

The AI-integrated group demonstrated significantly enhanced cognitive efficiency, suggesting reduced extraneous cognitive load:

- Task Completion Time: Decreased by an average of 25.4% compared to the control group (*t*(118) = 4.18, *p* < .001). Students spent less time wrestling with lower-order concerns like phrasing and basic editing.
- Revision Cycles: The average number of substantive revisions per document before final submission dropped by 41.7% in the experimental group (*t*(118) = 5.73, *p* < .001). This indicates more effective initial drafting and targeted revisions.
- Error Identification Accuracy: On dedicated error correction tasks, the experimental group showed a 32.8% improvement in accuracy compared to controls (*t*(118) = 3.94, *p* < .001), suggesting better developed proofreading and linguistic awareness skills.

3) Metacognitive and Collaborative Development

- Metacognitive Awareness: Self-reported metacognitive capabilities, as measured by the MAI composite score, showed a marked and significant improvement in the experimental group (Adj. M = 4.31/5 vs. 3.42/5 for controls; *t*(118) = 6.18, *p* < .001, *d*=1.15). Gains were particularly strong in the Regulation of Cognition subscale (planning, monitoring, evaluating).
- Collaborative Engagement: Analysis of peer feedback generated through Pigai.org revealed that experimental-group participants contributed, on average, 68% more substantive comments per review (defined as comments addressing content, organization, argumentation, or specific suggestions for improvement beyond surface corrections) compared to the comments generated in the control group's manual peer reviews (*p* < .001). The analytics and prompts within Pi-

gai.org appeared to foster deeper engagement.

4.2. Qualitative Findings

1) Collaborative Knowledge Construction
The qualitative data provided rich insights into how the AI tools, particularly Pigai.org, transformed collaboration. An instructor observed: "The biggest shift wasn't just more comments, but the kind of comments. Before, peer review often meant 'Good job' or 'Fix this grammar.' Pigai.org changed that. For example, the system automatically calculated and displayed thesis statement clarity metrics across different drafts side-by-side. Suddenly, students weren't just saying 'unclear thesis,' they were pointing to the metrics and saying, 'Look, your thesis score dropped in draft 2 because you added too much detail here. Let's compare how I phrased mine...' They started debating what 'clarity' actually meant in an engineering context, referencing examples from the style guide within the tool. It became a genuine discussion about disciplinary writing conventions, almost organically." This evolution aligns closely with Vygotsky's concept of collective scaffolding, where peers mediate each other's learning within the ZPD (Vygotsky 89), facilitated here by the AI's analytical prompts.
2) Evolution of Metacognitive Awareness and Critical Engagement
Interview data revealed a distinct evolution in how students engaged with AI feedback, indicative of developing metacognitive regulation. A Civil Engineering student articulated this progression clearly: "In the first few weeks, honestly, I just accepted almost every edit suggestion iFlytek gave me. It was faster and seemed smart. But around Week 5 or 6, working on a structural analysis report, it suggested simplifying a description of a load-bearing calculation. I realized that simplification would actually lose important technical nuance for my professor audience. I overrode the suggestion and kept my original phrasing. That felt like a turning point. Later, our teacher ran workshops on how to critically evaluate AI feedback – asking 'Does this change the technical meaning?', 'Is this appropriate for my specific reader?', 'Is this just stylistic preference or a real convention?'. By Week 8, I was actively questioning suggestions, especially those related to technical content or passive/active voice choices. I learned to use the AI as a starting point, not the final word." This journey from uncritical acceptance to selective, reasoned adoption signifies the development of crucial metacognitive awareness and strategic knowledge application (Flavell 907).
3) Contextual Limitations and the Need for Human Oversight
Both students and faculty consistently noted domain-specific limitations inherent in the current AI tools, reinforcing that they are not infallible replacements for human expertise, especially in highly specialized ESP contexts.

- Genre Variability: Feedback accuracy varied noticeably across sub-genres. Technical proposals and equipment descriptions generally received more accurate and useful feedback than experimental reports requiring nuanced interpretation of results or complex methodological descriptions. A Chemical Engineering student noted: "The feedback on my lab safety report was spot on. But for my chromatography results analysis, some suggestions about how to phrase the discussion were just... off. It didn't grasp the significance of the peak resolution data."

- Disciplinary Nuances: Specific linguistic features posed challenges. A Mechanical Engineering professor highlighted a recurring issue: "The system consistently flagged passive voice constructions in the Methods section as 'potentially weak' and suggested active voice alternatives. But in engineering writing, passive voice is often preferred in Methods to emphasize the process over the actor. We had to explicitly teach students to ignore that specific suggestion type in that specific section." This aligns with known challenges in NLP handling domain-specific pragmatics and register (Yang et al. 155).
- Depth of Understanding: While good at surface and structural issues, AI feedback rarely grasped deeper conceptual flaws or the true technical implications of phrasing changes. Instructors remained essential for addressing these higher-order concerns. As one instructor summarized: "The AI catches the misplaced modifiers and vague pronouns brilliantly, freeing me up. But it won't spot a fundamental misunderstanding of tensile strength concepts embedded in a sentence that's grammatically perfect."

5. Discussion

The empirical evidence gathered in this study substantiates the significant potential of strategically integrating sophisticated domestic AI tools (iFlytek Spark, iWrite, Pigai.org) within a constructivist ESP pedagogical framework. The quantitative findings demonstrate clear and substantial advantages for the AI-integrated group, particularly in context-dependent dimensions like audience adaptation and genre convention mastery. These gains resonate strongly with sociocultural theory's emphasis on authentic contextualization and the mediation of learning through cultural tools. The AI platforms functioned effectively as localized mediational tools, providing the situated practice and scaffolded support crucial for internalizing the complex discourse conventions of engineering English (Zhang et al. 97). The very large effect sizes (*d* > 1.3) observed in these areas underscore the practical significance of this approach for ESP curriculum development.

The cognitive efficiency metrics provide compelling evidence for AI's critical function in managing extraneous cognitive load within the demanding context of ESP writing (Sweller 78). By automating the detection and correction of surface-level errors (grammar, spelling, basic punctuation) and providing immediate feedback on lower-order concerns through tools like iWrite, learners were demonstrably freed to allocate greater cognitive resources to higher-order rhetorical and conceptual challenges—structuring arguments, tailoring content for specific audiences, refining technical explanations. The 41.7% reduction in revision cycles is a particularly telling indicator; it suggests that the AI scaffolding enabled learners to produce more polished initial drafts and focus subsequent revisions on substantive improvements rather than basic error correction, leading to more efficient writing processes overall.

The qualitative findings offer invaluable insights into the pedagogical mechanics underpinning these outcomes. The documented evolution in learners' engagement with AI feedback—from initial uncritical acceptance towards selective, reasoned adoption based on audience, purpose, and disciplinary norms—represents a core manifestation of developing metacognitive regulation (Flavell 909). Learners increasingly developed the ability to distinguish between mechanical corrections (which could often be safely automated) and rhetorical or conceptual decisions requiring human judgment. This critical discernment is a vital skill in an era of pervasive AI writing assistance. The instructor's role in explicitly teaching students how to evaluate AI suggestions emerged as a crucial factor in fostering this metacognitive growth, highlighting that tool integration must be coupled with pedagogical guidance.

The collaborative dimension revealed AI's powerful role as a "discourse catalyst." The Pigai.org platform, by quantifying feedback characteristics and providing comparative analytics on drafts, fundamentally altered the nature of peer interaction. Peer review transcended superficial commentary ("good," "awkward") and evolved into evidence-based "writing conferences", where students used concrete data (e.g., clarity scores, comparative thesis statements) to ground their discussions about disciplinary writing choices. This transformation strongly supports Gerry Stahl's framework of collaborative knowledge building, where technology mediates interactions to deepen collective understanding and co-construction of meaning. The significant increase in substantive peer comments quantifies this qualitative shift.

However, the study also surfaced critical tensions and limitations, primarily concerning contextual boundaries. AI's inconsistent performance across technical sub-genres and its struggles with highly specialized linguistic nuances (like the passive voice convention in engineering Methods sections) reinforce that it is not a panacea. These limitations underscore the continued indispensability of human expertise within ESP. The findings strongly advocate for "human-AI orchestration" models, where teachers strategically deploy technology to handle specific tasks (e.g., basic error checking, initial genre structure feedback, facilitating peer review logistics) while focusing their expertise on higher-order conceptual guidance, nuanced feedback on technical accuracy and depth, fostering critical thinking, and addressing the specific limitations of the AI tools within their domain (Chen & Li 1465). The AI acts as a powerful augmentative tool within a pedagogy still fundamentally guided by the instructor.

6. Conclusion

This study provides robust empirical validation for the efficacy of integrating constructivist pedagogical principles with advanced AI tools within Chinese ESP technical writing instruction for engineering students. It demonstrates that a thoughtfully designed AI-integrated curriculum, utilizing platforms like iFlytek Spark, iWrite, and Pigai.org, leads to statistically significant and educationally meaningful improvements in writing proficiency—especially in crucial areas like audience awareness and genre mastery—alongside enhanced cognitive efficiency and the development of metacognitive awareness and collaborative skills. Crucially, the findings illustrate that these tools, when implemented effectively, do not replace instruction or human interaction; instead, they enhance social learning processes by providing scalable, individualized scaffolding, facilitating deeper collaboration, and allowing instructors to focus on higher-order cognitive and disciplinary coaching.

The findings support a refined pedagogical model for AI-integrated ESP:

- AI as Cognitive Load Manager: AI tools excel at automating surface-level editing and providing immediate feedback on lower-order concerns, freeing cognitive resources.
- Learners as Active Constructors: Students engage in authentic, discipline-specific problem-solving tasks within simulated or realistic contexts, constructing knowledge through interaction with content, peers, instructors, and AI tools.
- Instructors as Cognitive Coaches: Teachers transition towards roles focused on designing authentic tasks, facilitating critical discussions, providing nuanced feedback on higher-order concerns, guiding metacognitive strategy development, and orchestrating the effective use of AI tools.
- Collaboration as Structured Knowledge Building: Peer interaction is enhanced through AI-mediated platforms that provide structure, analytics, and prompts for substantive, evidence-based feedback, fostering collective knowledge construction.

For Chinese ESP practitioners and curriculum designers seeking to implement similar approaches, the study offers concrete recommendations:

- Pedagogical Focus: Prioritize AI tools that explicitly enhance social learning processes (collaboration, discussion, peer review) and metacognitive development (critical feedback evaluation, self-regulation), not just isolated language practice.
- Tool Selection: Invest in platforms offering robust discipline-specific customization, such as iWrite's engineering-writing modules, to maximize contextual relevance and feedback accuracy. Prioritize tools with strong collaborative and analytical features like Pigai.org.
- Faculty Development: Institutions must invest significantly in building instructors' Technological Pedagogical Content Knowledge (TPACK) (Mishra & Koehler, 2006). Training should move beyond tool operation to focus on:
  - Integrating AI seamlessly within constructivist lesson design.
  - Developing strategies for teaching students critical AI literacy and feedback

evaluation.

- ■ Understanding AI limitations within specific ESP domains.
- ■ Orchestrating effective human-AI partnerships in feedback and instruction.

- Curriculum Design: Embed AI tools as integral components within authentic task sequences, ensuring alignment between tool capabilities, learning objectives, and disciplinary communication needs. Design activities that explicitly leverage AI for collaboration and metacognitive reflection.

7. Limitations and Future Research

This study, while rigorous, has limitations. The single-institution sample within the Chinese engineering ESP context limits generalizability. Future research should examine longitudinal impacts to assess the durability of gains and skill transfer; replicate the study across diverse ESP domains (e.g., Business English, Medical English, Legal English) to explore domain-specific variations in effectiveness and limitations, investigate diverse cultural and institutional contexts beyond China; conduct micro-genetic studies to trace in real-time the cognitive and collaborative processes mediated by different AI tools; research optimal feedback calibration strategies in localized AI systems to improve accuracy in specialized sub-genres and address identified limitations (e.g., passive voice conventions); and explore the impact of specific faculty development models on successful AI integration outcomes.

As domestic AI capabilities continue to evolve at a rapid pace, continuous pedagogical re-evaluation and adaptation remain essential. This study provides a foundational framework and empirical evidence for harnessing AI's potential to create more effective, efficient, and engaging constructivist learning environments within ESP, while acknowledging the irreplaceable role of the skilled instructor in guiding the process. The future of ESP lies not in AI replacing teachers, but in teachers strategically leveraging AI to empower learners within their specific discourse communities.

References

[1] Belcher, Diane. English for Specific Purposes in Theory and Practice. University of Michigan Press, 2009.

[2] Chen, Xiaoling, and Bo Li. "AI-Powered Peer Review in Chinese Engineering ESP Courses: A Case Study of Pigai.org." Computer Assisted Language Learning, vol. 36, no. 8, 2023, pp. 1452–70.

[3] Creswell, John W., and Vicki L. Plano Clark. Designing and Conducting Mixed Methods Rsearch. 3rd ed., Sage Publications, 2017.

[4] Dudley-Evans, Tony, and Maggie Jo St John. Developments in English for Specific Purposes: A Multi-disciplinary Approach. Cambridge University Press, 1998.

[5] Flavell, John H. "Metacognition and Cognitive Monitoring." American Psychologist, vol. 34, no. 10, 1979, pp. 906–11.

[6] Hyland, Ken. "Disciplinary Interactions: Metadiscourse in L2 Postgraduate Writing." Jounal of Second Language Writing, vol. 13, no. 2, 2004, pp. 133–51.

[7] Jonassen, David H. Computers as Mindtools for Schools. 2nd ed., Prentice Hall, 2006.

[8] Ministry of Education of China. Education Informatization 2.0 Action Plan. 2018.

[9] Schraw, Gregory, and Rayne Sperling Dennison. "Assessing Metacognitive Awareness." Contemporary Educational Psychology, vol. 19, no. 4, 1994, pp. 460–75.

[10] Stahl, Gerry. Group Cognition: Computer Support for Building Collaborative Knowledge. MIT Press, 2006.

[11] Sweller, John. Cognitive Load Theory. Springer, 2011.

[12] Vygotsky, Lev S. Mind in Society: The Development of Higher Psychological Processes. Harvard University Press, 1978.

[13] Yang, Yu-Fen, et al. "Blending AI and Human Feedback in EFL Writing." Educational Technology & Society, vol. 25, no. 2, 2022, pp. 148–62.

[14] Zhang, Hui, et al. "Localizing AI Writing Supporters for Chinese STEM Students: The iWrite Experience." English for Specific Purposes, vol. 68, 2022, pp. 89–103.