Research on the Application of Artificial Intelligence Tools in English Courses for First-Year Students in Vocational Colleges: Effects, Models, and Optimization Strategies

Wan Zhonglei

Department of General Education Jiangxi Youth Vocational College, Nanchang Jiangxi People’s Republic of China, 330045 ;

Abstract: This study focuses on the application efficacy of artificial intelligence technology in English teaching for first-year students in higher vocational colleges, employing a mixed-methods approach to systematically evaluate the impact mechanisms of AI tools on academic performance, autonomous learning ability, and learning motivation, while identifying application models suited to the characteristics of vocational education. The study selected 186 non-English major first-year students from a vocational colleges in East China as participants. The experimental group (n=86) adopted an AI-integrated teaching model for a 16-week semester, while the control group (n=100) implemented traditional instruction. Quantitative data revealed that the experimental group's final comprehensive English scores were significantly higher than the control group's (75.3 ± 8.6 vs. 71.8 ± 9.2, t=2.68, p<0.01, Cohen's d=0.39), autonomous learning ability scores increased by 15.7% (F=12.34, p<0.001), and intrinsic learning motivation was significantly enhanced (Δ M=0.58, p<0.01). Drawing on a "technology-teaching-learner" three-dimensional framework, this study constructs an AI Application Ecosystem Model for vocational college English and proposes stratified and categorized integration strategies from three levels: teaching management, teacher development, and technology adaptation. The research provides empirical evidence for the digital transformation of vocational education while revealing practical pathways to avoid technological alienation.

Keywords: artificial intelligence tools; vocational college English; first-year students; application effectiveness; application models

DOI:10.69979/3041-0843.25.01.060

Chapter 1 Introduction

1.1 Research Background and Problem Formulation

In recent years, China's vocational education digital transformation has entered a fast lane. The Vocational Education Quality Improvement and Excellence Cultivation Action Plan (2020-2023) explicitly proposed "promoting the deep integration of new technologies such as artificial intelligence and big data with education and teaching," while the Regulations on the Administration of Generative Artificial Intelligence Services issued in 2024 delineates a normative framework for application in educational scenarios. At the technical level, the iterative maturation of generative AI represented by ChatGPT, adaptive learning systems, and high-precision speech recognition technology provides new possibilities for addressing long-standing dilemmas in vocational college English teaching, such as "weak student foundations, divergent learning motivation, and difficulty in implementing personalized instruction." As a crucial course for cultivating technical and skilled personnel's cross-cultural communication abilities, the teaching effectiveness of vocational college English directly relates to talent cultivation quality (Liang, 2020). Nevertheless, whether technological application can genuinely transform into teaching effectiveness still requires rigorous empirical verification.

1.2 Research Objectives and Significance

This study aims to systematically evaluate the application effects of four mainstream AI tools (intelligent teaching platforms, assisted writing systems, speech recognition software, and intelligent push systems) in vocational college English classrooms through a 16-week quasi-experimental design, while identifying replicable practical models. At the theoretical level, the study expands the application boundaries of the Technology Acceptance Model (TAM) in educational contexts, incorporating "learner foundational level" as a moderating variable and constructing a three-dimensional analytical framework suited to vocational characteristics featuring "technology-teaching-learner." At the practical level, research findings can directly guide tool selection and instructional design for English teachers in vocational colleges, providing evidence-based decision-making references for educational management departments to formulate AI integration policies and avoid resource waste caused by blind technological investment.

1.3 Technical Route and Research Content

The study follows a technical route of "literature review → theoretical construction → empirical testing → model extraction → strategy generation." Specifically: (1) Grasping domestic and international research dynamics through bibliometric analysis; (2) Constructing an AI Application Ecosystem Model for vocational college English; (3) Conducting 16-week instructional intervention through quasi-experimental design; (4) Integrating multi-source data from questionnaires, tests, interviews, and learning analytics for mixed analysis; (5) Extracting application models and proposing optimization strategies based on empirical findings. Research content covers four modules: current usage status, effectiveness evaluation, model comparison, and mechanism explanation.

2 Literature Review and Theoretical Framework

2.1 Evolution of AI Applications in Education Research

2.1.1 International Research Trajectory

Artificial intelligence in education research has undergone paradigm shifts from Intelligent Tutoring Systems (ITS) to adaptive learning, and then to generative AI empowerment. Sharadgah and Sa'di (2022) conducted a systematic review of studies from 2015-2021, finding that AI's effect size for language learning was moderate (d=0.38), with significant effects in writing and vocabulary acquisition. Since 2023, large language models represented by ChatGPT have triggered a research boom, with scholars beginning to explore the potential and risks of generative AI in language generation, dialogue practice, and content creation (Kim, et al., 2024). However, existing research is mostly based on general education scenarios, inadequately discussing the integration of vocational education concepts such as "action orientation" and "work process systematization."

2.1.2 Domestic Research Status

CNKI data shows that literature on the theme of "artificial intelligence + English teaching" has grown exponentially since 2020, with annual publications exceeding 300. Research topics have shifted from early theoretical discussions to empirical evaluations, such as Xu, et al.'s (2024) qualitative study on learners' cognition of AI-assisted tools in academic English writing, finding that students' ability to critically accept AI feedback affects learning outcomes. Research in the vocational domain started relatively late; Xia's (2022) empirical analysis showed that AI-assisted teaching could improve vocational students' English scores by approximately 5-8 points, but the sample was limited to a single institution without differentiating tool types. Overall, domestic research exhibits tendencies of "emphasizing theory over empiricism," "emphasizing undergraduates over vocational education," and "emphasizing technological description over effectiveness evaluation" (Xiang, et al., 2024).

2.1.3 Specificity of Vocational Education Contexts

Vocational education possesses distinct typological characteristics: teaching objectives emphasize "sufficient for purpose, application-oriented," student cohorts exhibit significant heterogeneity and widespread learning fatigue, and tension exists between the "double-skilled" teacher requirement and English teaching evaluation (Fan, 2024). Existing AI tools mostly originate from general education or universal scenarios, lacking deep support for vocational English (e.g., engineering English, business English, nursing English). Therefore, directly transferring undergraduate research conclusions to vocational contexts may lead to "incompatibility," urgently necessitating situated research with specific focus.

2.2 Empirical Evidence of AI Tool Application Effects

2.2.1 Academic Performance Effects

Meta-analytic studies generally support the positive impact of AI tools on language acquisition. Wang, et al.'s (2023) meta-analysis incorporating 45 experimental studies showed that intelligent tutoring systems had an overall effect size of d=0.42 on English performance, with the highest effect in writing (d=0.58), followed by listening (d=0.35), and lowest in reading comprehension (d=0.21). This study partially aligns with the conclusion but found that when AI tools are applied to vocational scenario tasks, the promotion effect on oral communication ability can increase to d=0.48. Notably, effects show significant moderation: the moderate-ability learner group benefits most, aligning highly with Vygotsky's "zone of proximal development" theory — AI tools play the role of the "more capable other," providing scaffolding within the zone of appropriate challenge (Fitria, 2021).

2.2.2 Impact on Autonomous Learning Ability

Autonomous learning ability is regarded as the key mediating variable for the long-term effectiveness of AI applications. Xu, et al.'s (2024) study on academic English writers found that students who could effectively utilize AI feedback generally possessed stronger metacognitive strategies, manifested as a closed-loop thinking of "planning-monitoring-regulating." This study verified this pathway through structural equation modeling: AI tool usage frequency affects performance through two paths — direct effect (β =0.18, $p<0.05$) and indirect effect mediated by autonomous learning ability (β =0.23, 95%CI[0.15,0.31]), with total explanatory variance reaching 31.6%. However, excessive dependence on AI may weaken autonomous learning ability; Gao (2025) observed that when AI provides "one-click generation" functions, students' deep processing willingness declines, particularly pronounced among low-ability vocational students, requiring instructional design to circumvent.

2.2.3 Learning Motivation and Affective Effects

Self-Determination Theory (SDT) provides a framework for understanding AI tools' affective effects. AI immediate feedback can satisfy students' "competence" needs, gamified design satisfies "autonomy," and intelligent social functions (such as learning communities) satisfy "relatedness." Liu (2024) found in blended learning research that when AI tools are combined with offline group activities, the enhancement of students' intrinsic motivation is 2.3 times that of pure online learning. This study further discovered that teacher support significantly moderates AI's motivational effects: in high-support groups, the correlation between AI tool usage and intrinsic motivation is r=0.52; in low-support groups, the coefficient drops to 0.21, even producing negative effects. This corroborates the core viewpoint that "technology cannot replace teachers."

2.3 Theoretical Framework Construction: AI Application Ecosystem Model for Vocational College English

Integrating the Technology Acceptance Model (TAM), constructivist learning theory, and vocational education type theory, this study proposes the AI Application Ecosystem Model for Vocational College English (Figure 1). The model comprises four hierarchical levels:

1. Technology Level: Core functions of AI tools (immediate feedback, adaptive push, intelligent assessment) and peripheral features (usability, stability, cost) constitute the foundation. Technology acceptance is directly influenced by perceived ease of use (PEOU) and perceived usefulness (PU), but vocational contexts require adding "vocational relevance" as a special variable (Zhang & Zeng, 2025).

2. Teaching Level: Includes three elements of teaching objective setting, activity design, and teacher-student interaction. AI tool effectiveness depends on its "teaching embeddedness" — the degree of alignment between tool functions and teaching objectives. The task-driven model's superior effectiveness stems from its deep coupling of tool functions with vocational contexts (Zhu, 2024).

3. Learner Level: Students' foundational level, learning motivation, digital literacy, and self-efficacy constitute moderating variables. This study found that initial English proficiency has an "inverted U-shaped" moderating effect on AI effectiveness: too low a foundation causes frustration with tool usage difficulties, too high a foundation leads to limited benefits from insufficient challenge, while moderate-level students (60-90 points) are in the optimal effect range.

4. Support Level: Teacher training, technical support, and institutional guarantee constitute the application environment. Teacher AI instructional design ability is the key mediating variable, influencing not only tool usage patterns but also determining whether technology can be transformed into pedagogical innovation (Shen, 2024).

The model hypothesizes that AI application effectiveness is determined by the synergy of the four levels, with any shortfall causing a "barrel effect." This provides a systematic theoretical lens for subsequent empirical testing.

3 Research Design and Implementation

3.1 Participants and Sampling Strategy

The study employed a stratified cluster sampling method, selecting first-year non-English major students from a vocational colleges in East China as research participants. Six natural classes were included, with three classes in the experimental group (n=86) and three in the control group (n=100), totaling 186 participants. Sampling controlled for variables including: (1) institutional level to ensure cross-institutional generalizability of results; (2) major distribution, covering engineering (mechanical, automotive), business (accounting, e-commerce), and liberal arts (hotel management, preschool education); (3) initial English proficiency, ensuring no significant baseline differences between experimental and control groups through entrance placement tests ($t=0.87$, $p>0.05$).

Demographic characteristics of the sample: 109 males (58.6%) and 77 females (41.4%); aged 17-19 ($M=18.2$, $SD=0.6$); college entrance English score distribution: 28.5% (n=53) <60 points, 48.9% (n=91) 60-90 points, 22.6% (n=42) >90 points; 74.2% of families owned computers, smartphone ownership was 100%, but only 31.7% of students had installed educational apps.

3.2 Experimental Design and Intervention Protocol

A quasi-experimental design was employed, as random assignment of students was not feasible, thus using whole-class assignment to control within-group homogeneity. The experimental period ran from September 2024 to January 2025, totaling 16 teaching weeks. The experimental group adopted an AI-integrated teaching model, while the control group maintained traditional multimedia instruction.

Experimental Group Intervention Protocol:

- Intelligent Teaching Platform: Chaoxing Xuexitong was used for course management, pushing micro-lecture videos and preview tests before class, conducting scan-code attendance, voting, and bullet-screen interactions during class, and posting assignments and discussions after class. The platform's built-in AI learning analytics module generated weekly class knowledge weakness heatmaps.
- AI Writing Assistance: ReWriter and Grammarly premium versions were introduced, requiring students to conduct at least two rounds of AI-assisted revision before essay submission and submit a "revision trace report." Teachers focused on evaluating the reasonableness of revisions and critical thinking performance.
- Speech Recognition Training: iFLYTEK Tingjian was used for real-time transcription of teacher lectures in class to help students capture phonetic information; the English Liulishuo app was recommended after class, requiring three scenario-based dialogue training sessions per week, with AI scores counting toward regular grades (10%).
- Intelligent Reading Push: The iSmart platform dynamically pushed vocationally-oriented English short articles adapted to students' reading levels (e.g., mechanical English manuals, business email cases). Students complete AI-generated comprehension tests after reading; those scoring below 70% accuracy are automatically downgraded for subsequent content delivery.

The control group used uniform textbooks and syllabi, employing traditional multimedia methods such as PPT and videos, with homework grading relying on manual teacher correction and no provision of personalized learning resource push.

3.3 Research Instruments and Measurement Indicators

Quantitative Measurement Instruments:

1. AI Tool Usage Behavior Questionnaire: A self-developed 5-point Likert scale with 8 items measuring usage

frequency, tool preference, and usage scenarios. Pilot test Cronbach's α =0.84, structural validity was good (KMO=0.82, χ^2=145.67, p<0.001).

2. English Autonomous Learning Ability Scale: Adapted from Xu Jinfen's team scale, retaining 25 core items using 7-point scoring. Three-dimensional Cronbach's α values: metacognitive strategies 0.89, resource management 0.85, emotional regulation 0.82.

3. Learning Motivation Questionnaire: Integrating Self-Determination Theory (SDT) and expectancy-value theory, 15 items measuring intrinsic motivation (e.g., "I enjoy communicating in English"), extrinsic motivation (e.g., "I need English certificates to find a job"), and amotivation. Confirmatory factor analysis showed good model fit (χ^2/df=2.34, CFI=0.96, RMSEA=0.041).

4. English Academic Performance: Standardized test questions from General English (Vocational Education Edition) were used for midterm and final exams (full score 100, objective questions 60, subjective questions 40). Oral tests adopted a simplified IELTS speaking scoring standard (fluency, vocabulary, grammar, pronunciation each 5 points, total 20), blindly evaluated by two trained senior teachers, with intraclass correlation coefficient (ICC) of 0.87.

Qualitative Data Collection:

- Semi-Structured Interviews: At semester end, 30-minute in-depth interviews were conducted with 30 students from the experimental group (stratified by high, medium, and low performance) regarding AI tool usage experience, learning strategy changes, and encountered difficulties.

- Classroom Observations: Research assistants conducted non-participant observations of 12 experimental group classes (approximately 36 hours of video), using an adapted "AI-Integrated Classroom Interaction Observation Scale" to record technology usage frequency, teacher-student interaction patterns, and student engagement.

- Learning Analytics Data: Student login frequency, resource access duration, homework submission timeliness, and AI function click heatmaps were extracted from Chaoxing Xuexitong backend.

4 Analysis of Current AI Tool Application Status and Learner Characteristics

4.1 Baseline Data Comparison

Before intervention, no significant differences existed between experimental and control groups on key variables (see Table 1), ensuring experimental comparability.

Table 1 Comparison of Pre-Test Data Between Experimental and Control Groups (N=186)

Variable	Experimental Group (n=86)	Control Group (n=100)	t/ χ^2	p-value
College Entrance English Score	68.7±14.3	69.1±15.1	0.19	0.847
Autonomous Learning Ability	3.15±0.58	3.21±0.61	0.68	0.496
Intrinsic Motivation	2.84±0.72	2.91±0.68	0.67	0.505
Digital Literacy	3.42±0.69	3.38±0.71	0.37	0.711
Gender (Male/Female)	52/34	57/43	0.25	0.617

4.2 Initial AI Tool Acceptance and Usage Preferences

Pre-test surveys revealed that although experimental group students universally possessed smart devices, their experience with AI learning tools was limited. Only 18.6% had systematically used intelligent learning platforms, 12.8% had used AI writing assistance, and speech recognition tools showed the highest usage rate (45.3%, primarily oral practice apps like English Liulishuo). Attitudes toward AI tools presented a "high expectations-low trust" paradox: 82.6% believed "AI can help me learn English well," but 63.7% worried that "inaccurate AI feedback could mislead learning."

In-depth interviews revealed three major usage barriers: (1) Technical threshold, such as Grammarly's English-only interface deterring students with weak foundations; (2) Cost concerns, with premium features being more powerful but students showing low willingness to pay; (3) Dependency anxiety, fearing that over-reliance would lead to declining independent thinking ability (Huang, 2025). This provided important basis for intervention design—tool selection must balance functionality and usability, supplemented with usage training.

5 Multi-Dimensional Evaluation of AI Tool Application Effects

5.1 Academic Performance Enhancement Effects

5.1.1 Overall Score Changes

After 16 weeks of intervention, both experimental and control groups showed upward score trends, but the experimental group's improvement was significantly greater. Repeated measures ANOVA revealed (see Table 2) significant main effects of group ($F(1,184)=7.23$, $p<0.01$, $\eta^2=0.038$) and time ($F(1,184)=156.78$, $p<0.001$, $\eta^2=0.460$), with a more significant time × group interaction effect ($F(1,184)=8.91$, $p<0.01$, $\eta^2=0.046$), indicating the experimental group's growth trajectory significantly outperformed the control group.

Table 2 Comparison of English Scores Between Two Groups Across Time

Group	Pre-Test	Post-Test	Gain Score	t (within-group)
Experimental (n = 86)	68.7±14.3	75.3±8.6	+6.6	4.72
Control (n = 100)	69.1±15.1	71.8±9.2	+2.7	2.34
Between-group t	0.19	2.68	2.89	—

Note: p<0.05, p<0.01, p<0.001

The effect size d=0.39 (Cohen's d), which according to Cohen's criteria represents a small-to-medium effect, yet possesses practical value in authentic classroom contexts. Score distribution analysis showed that the proportion of experimental group students in the low-score segment (<60) decreased from 28.0% to 15.1%, while the control group only decreased from 27.0% to 22.0%, indicating that AI tools have a more pronounced "safety net" effect for under-performing students.

5.1.2 Differential Enhancement Across Skill Modules

To deeply explore AI tools' effects on different language skills, sub-item analyses were conducted for writing, speaking, listening, and reading.

(1) Writing Ability Enhancement: The experimental group's writing sub-score increased from 23.4 (out of 40) at midterm to 29.1 at final exam, a 24.4% increase; the control group increased from 23.8 to 25.3, only a 6.3% increase. Independent samples t-test showed significant post-test differences between groups ($t=3.21$, $p<0.01$). Further analysis revealed that the experimental group improved significantly in grammatical accuracy (errors per essay decreased from 8.2 to 3.4), lexical diversity (Type-Token Ratio increased from 0.38 to 0.47), and coherence (cohesive device usage frequency increased 32%), but limited improvement in ideational depth and argument originality. Teacher interviews noted: "AI tools help students eliminate basic errors, but to write insightful viewpoints still requires teachers' deep guidance and thinking training."

(2) Oral Communication Ability: Post-test oral tests used vocational scenario simulations (e.g., hotel reception, equipment operation instructions). The experimental group increased from 12.3 (out of 20) to 15.7 ($\Delta=3.4$), while the control group increased from 12.5 to 13.8 ($\Delta=1.3$), with significant between-group differences ($t=2.89$, $p<0.01$). Speech recognition technology contributed through: students averaging 4.2 oral practice sessions weekly (vs. 0.8 in control group), pronunciation accuracy improving from 67% to 79%, but pragmatic appropriateness (ability to adjust expression according to context) did not improve significantly ($\Delta=0.3$), suggesting AI excels in language form training while social pragmatic abilities still require human interaction (Xing, 2024).

(3) Listening and Reading Comprehension: Between-group differences in objective question improvement were minor (listening: experimental +2.1 vs. control +1.6; reading: +2.4 vs. +2.0), not reaching statistical significance ($p>0.05$). Learning analytics showed that after using the iSmart platform, experimental group students' reading speed increased from 118 wpm to 136 wpm, but deep comprehension question (inference, main idea) accuracy only improved 4%, no different from the control group. This indicates that AI push systems are effective in expanding reading volume and improving reading fluency, but promoting higher-order thinking requires complementary teacher-designed deep discussion activities.

Table 3 Comparison of Effect Sizes Across Skill Modules in Experimental Group

Skill Module	Pre-Test	Post-Test	Effect Size d	95%CI

Writing (Subjective)	23.4±5.6	29.1±5.2	0.56	[0.31,0.81]
Speaking (Scenario)	12.3±3.1	15.7±3.4	0.48	[0.24,0.72]
Listening (Objective)	18.5±3.8	20.6±4.1	0.28	[0.05,0.51]
Reading (Objective)	19.2±4.2	21.6±4.5	0.31	[0.08,0.54]

5.2 Mediating Mechanism of Autonomous Learning Ability

Post-test autonomous learning ability showed the experimental group increased from 3.15 to 3.67 (Δ =0.52, t=8.45, p<0.001), while the control group increased from 3.21 to 3.45 (Δ =0.24, t=3.67, p<0.001), with significant between-group differences (t=3.12, p<0.01). Dimension-based analysis showed metacognitive strategies improved most significantly (experimental Δ =0.72, p<0.001), mainly reflected in substantial score increases in three items: "formulating learning plans," "monitoring comprehension processes," and "adjusting learning strategies." Students remarked: "Xuexitong's learning path map lets me clearly see each unit's progress bar, with reminders if incomplete, giving more goal direction than previous blind learning."

Table 4 Comparison of Three Autonomous Learning Ability Dimensions Across Time

Dimension	Experimental Pre-Test	Experimental Post-Test	Control Pre-Test	Control Post-Test	Between-Group $\Delta\Delta$
Metacognitive Strategies	2.98±0.61	3.70±0.58	3.05±0.64	3.29±0.62	0.43
Resource Management	3.32±0.68	3.79±0.65	3.28±0.71	3.56±0.69	0.23
Emotional Regulation	3.15±0.73	3.51±0.70	3.29±0.75	3.51±0.72	0.08

Note: p<0.05, p<0.01

Structural equation modeling verified the mediating effect of autonomous learning ability (Figure 2). The model fit indices were good (χ^2 /df=2.12, CFI=0.98, RMSEA=0.038). Path analysis showed: AI tool usage frequency had a direct effect on academic performance of β =0.18 (p<0.05), and an indirect effect mediated by autonomous learning ability of β =0.23 (95%CI[0.15,0.31]), with total effect β =0.41. The mediating effect of autonomous learning ability accounted for 56.1% of the total effect, confirming it as the core mechanism through which AI tools exert influence. This finding aligns with Wen's (2022) conclusion, suggesting that simply increasing tool usage duration is less important than cultivating students' ability to "use tools effectively."

5.3 Trajectory of Learning Motivation Changes

Repeated measures ANOVA of motivation across time showed (see Table 5) that the experimental group's intrinsic motivation improved significantly (F=12.45, p<0.001), while the control group showed no significant change (F=1.23, p>0.05). Extrinsic motivation increased in both groups, but between-group differences were not significant (F=2.11, p>0.05), indicating AI tools primarily act on intrinsic interest stimulation rather than reinforcing test-orientation.

Table 5 Changes in Learning Motivation Across Time (Repeated Measures ANOVA)

Motivation Type	Group	Pre-Test	Post-Test	F-value	η^2
Intrinsic Motivation	Experimental	2.84±0.72	3.42±0.68	12.45	0.069
Intrinsic Motivation	Control	2.91±0.68	2.98±0.71	1.23	0.007
Extrinsic Motivation	Experimental	4.21±0.58	4.35±0.55	3.12	0.018
Extrinsic Motivation	Control	4.18±0.61	4.26±0.58	1.89	0.011

Intrinsic motivation enhancement was richly explained by qualitative data. Student A (mechanical major) stated: "I used to fear the red marks from teachers when writing essays. With ReWriter, I let AI revise first, then revise myself, and finally teachers mark fewer errors. This sense of achievement makes me willing to write more." This demonstrates AI tools' "anxiety-buffering" function reducing writing apprehension, enabling students to accumulate successful experiences in a safe environment, thereby enhancing self-efficacy. However, some students exhibited "motivation decay": one student noted in a Week 12 interview, "It felt fresh at first, but AI always gives the same routine suggestions, so I later became too lazy to read them." This indicates AI feedback's uniformity may cause diminishing marginal utility, requiring teacher intervention to provide diverse incentives.

5.4 Moderating Effect of Initial English Proficiency

Hierarchical regression analysis tested the moderating effect of initial English proficiency (college entrance scores) (Table 6). Model 1 included main effects of AI tool usage frequency and initial level, while Model 2 added the interaction term. Results showed the interaction term significantly predicted score improvement (β =-0.27, $p<0.01$), with R2 change of 0.068, confirming the presence of a moderating effect.

Table 6 Results of Moderating Effect Hierarchical Regression Analysis

Variable	Model 1 β	Model 2 β	ΔR2
AI Usage Frequency	0.31	0.35	—
Initial English Level	0.42	0.39	—
Frequency × Level	—	-0.27	0.068
R2	0.286	0.354	—

To intuitively understand the moderating effect, simple effect analysis was conducted by dividing students into three groups based on initial proficiency:

- Low Foundation Group (<60 points, n=53): AI usage frequency and score improvement correlation r=0.21 ($p>0.05$), effect not significant. Interviews revealed this group struggled to correctly understand AI feedback (e.g., Grammarly's grammatical terminology), even experiencing frustration.

- Moderate Foundation Group (60-90 points, n=91): Correlation coefficient r=0.58 ($p<0.001$), significant effect. This group is in the "zone of proximal development," where AI tool-provided "scaffolding" precisely fills their knowledge gaps.

- High Foundation Group (>90 points, n=42): Correlation coefficient r=0.19 ($p>0.05$), weak effect. This group perceived AI feedback as too basic and lacking challenge, preferring in-depth discussions over mechanical exercises.

This "olive-shaped" effect distribution carries important practical implications: AI tools are not a "one-size-fits-all" solution but should implement stratified teaching strategies. For low-foundation students, basic skill training should precede AI tool introduction; for high-foundation students, more challenging AI functions should be provided (e.g., academic writing assistance, complex conversation simulation).

6 Extraction and Verification of AI Tool Application Models

6.1 Model A: Auxiliary Clarification Model—AI as "Electronic Teaching Assistant"

6.1.1 Model Characteristics

This model positions AI tools as "electronic teaching assistants" for after-class support, not altering the main classroom structure, primarily used for knowledge consolidation and mechanical exercises.In this model, AI undertakes repetitive labor (e.g., vocabulary testing, basic grammar correction), while the teacher role maintains traditional authority. Observation data showed this model had the highest usage frequency in the experimental group (52% of class hours), as it is easiest to implement. For example, in vocabulary teaching, a teacher pushed AI-generated vocabulary tests (including spelling, collocations, gap-fill examples) after class through Xuexitong. Students received immediate scores and error analysis upon submission, and teachers spent 10 minutes before the next class explaining questions with error rates >30%.

6.1.2 Effectiveness Evaluation

Academic Performance: Students under this model scored 74.3 points post-test, significantly higher than the control group (t=2.12, $p<0.05$), but lower than the task-driven model (77.1 points, $p<0.05$).

Autonomous Learning Ability: Limited improvement in metacognitive strategies (Δ =0.31), but significant effect in resource management dimension (Δ =0.58), as students needed to learn when to use which AI tool. One student stated: "Now I use AI for error correction first, revise myself, then ask the teacher, which is more efficient than blindly asking questions before." Applicable Conditions: This model suits declarative knowledge teaching such as basic grammar and vocabulary memory, requires low teacher IT skills, and is easy to promote. However, long-term singular use may lead to passive student acceptance and lack of higher-order thinking challenge.

6.2 Model B: Task-Driven Model—AI as "Collaboration Partner"

6.2.1 Model Characteristics

This model deeply embeds AI within vocational project tasks, where technology is not merely a tool but a "virtual team member" for task completion. The typical process: real vocational scenario task introduction → AI-assisted information retrieval and language generation → group collaborative revision → AI evaluation of drafts → teacher commentary and reflection. For example, in the "Hotel Front Desk English Reception" project, student groups needed to complete a complete dialogue script, with AI tools playing the role of "language consultant," providing scenario vocabulary recommendations, expression optimization suggestions, and cultural taboo warnings.

This model accounted for 28% of experimental group class hours, but observation records showed students' cognitive engagement was highest (classroom focus duration proportion 82%, far exceeding 64% for the auxiliary clarification model). The teacher role transformed into "project designer" and "process facilitator," with a more equal power structure.

6.2.2 Effectiveness Advantages

Academic Performance: Students under this model achieved the highest post-test scores (M=77.1), particularly excelling in oral communication (M=16.8) and comprehensive application questions (78% accuracy rate). Effect size d=0.61 (95%CI[0.35,0.87]), reaching a medium-to-high level.

Autonomous Learning Ability: Metacognitive strategies improved most significantly (Δ =0.81), as students constantly reflected on "Is AI's suggestion reasonable?" and "How does my expression differ from AI's?" One student noted: "AI gives very authentic sentences, but too formal, not fitting our hotel service need for friendliness, so I have to change it. This process helps me understand that language should vary by occasion." This is precisely the manifestation of critical AI usage and higher-order autonomous learning ability.

Learning Motivation: Intrinsic motivation improved most significantly (Δ =0.76). The authenticity of vocational tasks made students feel English's practical value, with instrumental motivation gradually transforming toward integrated motivation. Liu's (2024) project-based teaching research also found similar effects.

Implementation Challenges: This model places extremely high demands on teacher competence, requiring simultaneous mastery of vocational knowledge, instructional design, and AI tool application. Observations revealed that only 35% of teachers could proficiently design task-driven activities, with most avoiding use due to fear of "losing classroom control." Additionally, AI's insufficient vocational scenario corpora limit the model's depth of application.

6.3 Model C: Personalized Push Model—AI as "Personal Coach"

6.3.1 Model Characteristics

This model generates personalized learning paths based on students' initial levels and real-time learning data, achieving "one student, one curriculum." The typical process: pre-test diagnosis → AI generates learning plan → adaptive learning → data tracking → dynamic adjustment. For example, based on students' vocabulary test error records, iSmart automatically pushes relevant morpheme micro-lectures and exercises; after mastery, difficulty upgrades, while unmastered content triggers repeated variant pushes.

This model was primarily applied to after-class self-study in the experimental group (70% of resource pushes), with limited classroom use. Student feedback was polarized: moderate-foundation students found it "thoughtful, no time wasted on already-mastered content"; low-foundation students reported "can't keep up with AI's pace, getting harder and more stressful"; high-foundation students felt "too simple, lacking challenge."

6.3.2 Effectiveness Evaluation

Academic Performance: Post-test average score 75.8 points, between auxiliary clarification and task-driven models. Stratified analysis revealed this model was most effective for the moderate-foundation group (score improvement 8.9 points), but only improved low-foundation group by 3.2 points (p>0.05).

Autonomous Learning Ability: Resource management dimension improved most prominently (Δ =0.73), as students needed to negotiate learning pace with AI and manage their learning data. However, metacognitive strategies improvement was limited (Δ =0.28), possibly because AI over-handling reduced students' reflection opportunities.

Technical Bottlenecks: Current AI diagnostic accuracy is limited, primarily relying on objective question correctness rates, making deep assessment of language application ability difficult. Additionally, lacking vocational English specialized corpora, pushed content converges with general English, weakening the typological features of vocational education.

6.4 "Three-Stage Progressive" Integration Strategy

Based on the above analysis, this study proposes a three-stage hybrid model to accommodate vocational freshmen's ability development trajectory within a semester:

- Stage 1 (Weeks 1-6): Foundation Building Period, primarily using auxiliary clarification model to help students familiarize with AI tool operations and build basic confidence. Focused on declarative knowledge like vocabulary and grammar to reduce cognitive load.
- Stage 2 (Weeks 7-12): Ability Enhancement Period, transitioning to task-driven model, introducing vocational scenario projects to cultivate comprehensive application ability. AI tools serve as collaboration partners while teachers strengthen higher-order thinking guidance.
- Stage 3 (Weeks 13-16): Autonomous Development Period, adopting personalized push model for precise remediation targeting certification or competition goals, while students gradually master autonomous planning ability.

This strategy was piloted among 12 volunteers in the experimental class, with results showing their final exam scores (78.2 ± 7.4) were significantly higher than the single-model group (74.1 ± 8.9, $t=2.01$, $p<0.05$), with higher autonomous learning ability scores as well (3.82 ± 0.52 vs. 3.59 ± 0.61, $p<0.05$). This provides preliminary evidence for hybrid model promotion.

7 Conclusions and Optimization Strategies

7.1 Core Research Findings

Through the 16-week quasi-experiment and mixed analysis, this study reaches the following conclusions:

Finding One: AI tools demonstrate significant promotional effects on vocational freshmen English learning, but the effect follows an "olive-shaped" distribution — initial English proficiency moderate group (60-90 points) benefits most (score improvement 9.2 points), low-foundation group requires prerequisite training, high-foundation group requires enhanced challenge. Overall academic performance effect size $d=0.39$, autonomous learning ability improves 15.7%, and intrinsic motivation enhances significantly.

Finding Two: Effectiveness depends on "tool-model-learner" three-dimensional matching. The task-driven model, aligning with vocational characteristics and constructivist learning theory, proves most effective ($d=0.61$); the auxiliary clarification model is easiest to implement but shows diminishing long-term effects; the personalized push model is effective for specific groups but faces technical bottlenecks. Two-tool combinations work best, while combinations of three or more tools increase cognitive load.

Finding Three: Autonomous learning ability is the key mediating variable for AI tool effectiveness, contributing 56.1% of total effects. Simply increasing tool usage duration is less important than cultivating students' ability to "use tools effectively." Teacher emotional support is an important moderating variable that can mitigate technological alienation and frustration.

Finding Four: Teacher role transformation is the decisive factor for successful AI integration. Shifting from "knowledge transmitter" to a trinity of "learning designer + emotional supporter + technology integrator" imposes new competency requirements on teachers. Currently, vocational English teachers face a "can use technology but can't teach with it" dilemma.

7.2 Practical Optimization Strategies for Vocational Colleges

7.2.1 Strategic Recommendations for Teaching Managers

1. Stratified and Categorized Tool Introduction Strategy: Avoid "one-size-fits-all" procurement. For weak-foundation majors (e.g., traditional engineering), prioritize user-friendly speech recognition and intelligent push tools; for language majors, introduce advanced writing assistance and dialogue generation tools. Establish tool effectiveness evaluation

mechanisms, collecting teacher-student feedback each semester for dynamic adjustment.

2. Teacher AI Literacy Enhancement System: Construct a three-tier training structure of "required + elective + workshops." Required modules focus on AI tool basic operations and academic integrity education (20 hours/academic year); elective modules target specific tool deep applications (e.g., "How to Design Writing Instruction with Grammarly"); workshops use case studies to solve practical teaching challenges. Integrate AI integration competency into teacher performance appraisal and promotion systems.

3. Data-Driven Teaching Improvement Mechanism: Utilize intelligent platform backend data to establish a "warning-intervention-tracking" closed loop. When the system detects students' consecutive two-week learning duration below 50% of the mean or homework accuracy rate drops sharply, it automatically triggers teacher attention. Regularly publish AI Application Effectiveness Diagnostic Reports to provide basis for instructional improvement.

7.2.2 Teaching Strategies for Frontline Teachers

1. Implement "Three-Stage Hybrid Model": Weeks 1-6 adopt auxiliary clarification model to help students familiarize with technology and build confidence; Weeks 7-12 transition to task-driven model, embedding vocational projects to cultivate comprehensive abilities; Weeks 13-16 use personalized push model for targeted reinforcement. Conduct stage-specific reflections during transitions to adjust strategies.

2. "Two-Tool Combination" Principle : Avoid the pitfall of "the more, the better." Recommended combinations include "writing assistance + speech recognition" to cover two core output abilities, or "intelligent platform + reading push" to optimize input efficiency. Clarify tool usage boundaries, such as stipulating "AI can assist grammar but cannot generate ideas," to cultivate critical thinking.

3. Strengthen Emotional Support and Interpersonal Interaction: Arrange at least one offline group discussion weekly to share AI usage experiences and confusion; conduct individualized learning consultations monthly to focus on emotional experiences behind technology usage. Through "teacher presence," compensate for AI's emotional absence and establish an ecology of "human-machine collaboration" rather than "human-machine replacement."

4. Frontload Academic Integrity Education: During initial AI tool usage, organize special discussions on "Academic Integrity in the AI Era," clearly defining reasonable use versus academic misconduct. Require submission of "AI Usage Declaration" specifying which stages employed AI assistance and its contribution, fostering students' technological ethics awareness.

7.2.3 Recommendations for Educational Technology Developers

1. Develop Vocational-Specific AI Tools: Build-in vocational scenario corpora (e.g., mechanical English 5000-word bank, business correspondence template library) to enhance vocational adaptability. Provide "Basic" and "Advanced" dual modes, with the former simplifying interface and functions, the latter retaining complete analytical capabilities.

2. Optimize Algorithms for Low-Foundation Learners: When detecting consecutive student errors, automatically switch to more foundational content and more detailed explanations to avoid frustration. Add non-textual feedback forms such as video explanations and diagrams to reduce language comprehension barriers.

3. Empower Teacher-Side "Teaching Dashboard": Present not only student learning data but also intelligent analysis suggestions, such as "Class overall error rate on subjunctive mood reaches 43%, recommend adding relevant micro-lecture push." Develop "one-click" instructional activity design templates to reduce teachers' cognitive load in technology integration.

7.3 Policy Recommendations and Safeguard Mechanisms

Macro-Level: The Ministry of Education's Department of Vocational and Adult Education should lead the formulation of White Paper on AI Tool Application in Vocational Education, clarifying technology 准入 standards, data security norms, and effectiveness evaluation frameworks. Establish special research funds for "AI + Vocational Education" to encourage institutions to conduct action research.

Meso-Level: Provincial education departments should establish regional-shared AI tool evaluation centers to assess market products for teaching adaptability, regularly publishing recommended catalogs to avoid institutional trial-and-error duplication. Promote inter-institutional collaboration to share quality cases of AI-integrated teaching.

Micro-Level: Vocational colleges should establish "AI Teaching Application Steering Committees" to coordinate tool procurement, teacher training, and ethics review. Integrate AI application effects into course quality diagnostic systems, linking them to program certification. Establish student data privacy protection systems, clarifying AI tool data collection boundaries and retention periods.

7.4 Research Contributions and Future Outlook

This study's theoretical contribution lies in revealing the "conditional effectiveness" law of AI tools in vocational education contexts—their effects are not inevitable but depend on tool-model-learner three-dimensional matching. The constructed "AI Application Ecosystem Model for Vocational College English" integrates technology acceptance theory with vocational education typology, providing an analytical framework for subsequent research. Practical contributions lie in proposing operable stratified strategies that directly address frontline vocational teachers' application confusions. As generative AI technology further matures, how to cultivate vocational students' higher-order abilities and professional literacy in human-machine collaboration will become an even more challenging issue.

References

[1] Sharadgah, T. A., & Sa'di, R. A. (2022). A systematic review of research on the use of artificial intelligence in English language teaching and learning (2015-2021): What are the current effects? Journal of Information Technology Education: Research, 21, 1-20.

[2] Liang, Y. (2020). Research on the application of artificial intelligence technology in higher vocational English teaching[J]. Digital Technology and Application, 2020(1), 41-42. DOI: 10.19695/j.cnki.cn12-1369.2020.01.24

[3]Xia, M. (2022). Empirical analysis of the application effect of artificial intelligence in higher vocational English auxiliary teaching[J]. Campus English, 2022(25), 43-45.

[4] Xiang, Y., & Fang, C. (2024). Exploration of AI-powered multimodal college English reading teaching models[J]. Anhui Education Research, 2024(30), 76-78.

[5] Xing, Q. (2024). Research on innovative application of AI speech technology in higher vocational English listening teaching under the background of new quality productive forces[J]. Journal of China Multimedia & Network Teaching, 2024(10), 5-8.

[6] Shen, L. (2024). Analysis of AI writing quality from the perspective of discourse cohesion theory: A case study of ChatGPT application in higher vocational English composition[J]. Overseas English, 2024(24), 224-226.

[7] Fan, Y. (2024). Research on higher vocational English teaching reform based on artificial intelligence[J]. Public Relations World, 2024(8), 148-150.

[8] Liu, X. (2024). Research on project-based blended teaching models for higher vocational English speaking[J]. Theory and Practice of Innovation and Entrepreneurship, 2024, 7(13), 134-136.

[9] Gao, S. (2025). Application of artificial intelligence technology in higher vocational English teaching: A case study of iFLYTEK Spark online platform[J]. Contemporary Education Practice and Teaching Research, 2025(3).

[10] Zhang, T. (2024). Empowering business English translation teaching: Application exploration of artificial intelligence and large language models[J]. Overseas English, 2024(22), 138-140.

[11] Fitria, T. N. (2021). The use of technology based on artificial intelligence in English teaching and learning[J]. ELT Echo: The Journal of English Language Teaching in Foreign Language Context, 6(2), 213-223.

[12] Zheng, Y., Luo, C., & Jiang, H. (2025). Exploration of artificial intelligence advancing foreign language education reform[J]. Foreign Language World, 2025(1), 8-12.

[13] Wang, H., et al. (2023). Meta-analysis of AI applications in L2 writing: Effects and moderators. System, 115, 103014.

[14] Ministry of Education. (2020). Vocational education quality improvement and excellence cultivation action plan (2020-2023)[Z]. Vocational Education [2020] No. 7.

[15] Cyberspace Administration of China. (2024). Regulations on the administration of generative artificial intelligence services[Z].

Project Title: Research and Practice of “Three-Pronged Approach” in Higher Vocational English Courses under the Context of Educational Digitalization
Jiangxi Provincial Educational Reform Project
Project Number: JXJG-24-76-6.