**International Journal of Education and Humanities**
ISSN: 2770-6702 | Vol. 24, No. 1, 2026

# Reconstruction and Empirical Research on Full-Process Maritime English Teaching in Higher Vocational Colleges Supported by Dual Collaborative Generative AI Models

**Xiang Huang**

Jiangsu Maritime Institute, Nanjing, Jiangsu 211170, China

**Abstract:** Maritime English teaching in higher vocational colleges is plagued by practical dilemmas including insufficient provision of occupation-based scenarios, limited oral training channels, unimplemented student stratification, and rigid summative assessment systems. To advance the digital transformation of maritime specialized English education, this study adopts situated learning, task-based language teaching, and multi-dimensional process evaluation theories to construct a three-stage full-process teaching framework featuring the collaborative operation of DeepSeek and Doubao. The framework forms a closed-loop teaching system of "pre-class differentiated guided learning – in-class situated interactive training – post-class data-driven assessment". This research aims to develop a low-cost, replicable digital teaching implementation scheme, providing a localized practical paradigm for the intelligent teaching reform of English for Specific Purposes (ESP) in higher vocational education.

**Keywords:** Generative Artificial Intelligence, Maritime English in Higher Vocational Colleges, Dual-Model Collaboration, Full-Process Teaching Empirical Research.

## 1. Introduction

Against the backdrop of digital transformation in vocational education, the deep integration of artificial intelligence with specialized courses has become the core direction of industry-specific English teaching reform[1]. The Revised Vocational Education Law of the People's Republic of China (2022) clearly proposes building personalized and scenario-based education models via digital technologies. The Jiangsu Vocational Education Quality Improvement Action Plan (2025–2027) further emphasizes aligning general basic courses with industrial post standards and empowering classroom quality improvement through information technology. As a core ESP course for shipping majors, Maritime English complies with the Standards of Training, Certification and Watchkeeping for Seafarers (STCW) and seafarers' competency assessment requirements, cultivating students' professional linguistic competence in ship watchkeeping communication, maritime emergency disposal, and port foreign exchange.

Current Maritime English instruction in domestic higher vocational colleges suffers from multiple structural contradictions. Classroom teaching mainly focuses on vocabulary and grammar without authentic maritime scenarios such as VTS traffic scheduling, collision avoidance at sea, and ship customs clearance. Restricted by faculty scale, training venues, and class hours, students lack regular oral dialogue practice opportunities[2]. Significant gaps in English proficiency exist among students recruited from general high schools and secondary vocational schools, making individualized instruction unfeasible under unified teaching schedules. Course evaluation has long relied on final written examinations, failing to capture full-cycle learning behaviors including preview, in-class interaction, and after-class exercises, thus incapable of objectively measuring seafarers' practical post-related English competence[3].

Generative large language models support automatic scenario material generation, real-time voice interaction, learning data statistics, and differentiated task customization, offering technical solutions to address the above teaching bottlenecks. DeepSeek excels at proofreading maritime professional texts and developing standardized shipping scenario scripts, while Doubao has advantages in oral dialogue simulation, differentiated learning analysis, and lightweight teaching material generation. The differentiated collaboration of the two models balances professional rigor and classroom interactivity. A review of existing literature reveals that overseas maritime colleges have deployed intelligent English training platforms earlier, yet significant discrepancies exist between domestic and foreign maritime regulations and seafarer assessment criteria, making foreign models unsuitable for direct localization. Domestic studies on AI-enabled Maritime English mostly remain fragmented applications such as pronunciation correction and exercise generation, lacking systematic designs covering pre-class, in-class, and post-class teaching links. Empirical research on dual-model collaboration and full-cycle data-based assessment remains scarce[4][5]. Accordingly, this study reconstructs the full-process Maritime English teaching system relying on dual generative AI models, tests the model's effectiveness through teaching experiments, and forms a promotable standardized teaching reform pathway.

## 2. Domestic and International Research Status

### 2.1. Foreign Research

Overseas maritime colleges have pioneered intelligent seafarer English training. Warsash Maritime Academy in the UK and Singapore Polytechnic have built standardized virtual

ship dialogue platforms to boost oral training frequency via benchmark management, with studies confirming AI's capacity to reduce learners' anxiety in oral expression[6]. Distributed computing frameworks have also been applied to project-based Maritime English design, yet such outcomes target short-term overseas seafarer training and feature low compatibility with domestic higher vocational maritime regulations and assessment systems, without differentiated stratification mechanisms tailored to mixed-ability domestic student cohorts.

Two core research strands dominate general second language education. First, Godwin-Jones (2024) put forward the concept of distributed agency in generative AI-assisted second language learning, arguing that human-machine collaboration expands language practice channels but carries drawbacks including insufficient cultural authenticity and model hallucinations[7]. Second, Kukulska-Hulme (2024) elaborated personalized assessment pathways from the perspective of mobile AI learning[8]. Meanwhile, Mohamed (2025) and Hockley (2024) separately warned of long-term risks including distorted AI-generated maritime English materials and digital divides, providing international theoretical references for constructing the risk balancing system of this study[9].

## 2.2. Domestic Research

Domestic research falls into three strands. First, partial AIGC teaching applications such as the 321 writing model for marine engineering English and integrating AI into maritime vocational safety culture, which fail to cover complete teaching cycles[4][5]. Second, traditional blended Maritime English reforms adopting OBE-based role-play classrooms with insufficient digital integration and no dual-model differentiated design[10][11]. Third, research on risk control in AI teaching, where scholars proposed a three-stage technical access mechanism that reduces the proportion of AI-generated text by 61%, delivering viable ideas for restraining students' cognitive laziness[12].

A comprehensive review of CNKI literature identifies three core research gaps. First, fragmented technical applications without full-process instructional designs dividing labor between DeepSeek and Doubao. Second, overreliance on classical pedagogical theories lacking original middle-range theories adapted to maritime human-AI interaction scenarios. Third, insufficient empirical rigor: most teaching reforms omit pre-test balance tests and effect size reporting, undisclosed stratification rules and prompts, resulting in poor research reproducibility. This study conducts systematic investigations, model construction, and controlled empirical research to fill these gaps.

# 3. Current Status of Generative AI Empowering Full-Process Maritime English Teaching in Higher Vocational Colleges

To accurately identify practical frontline teaching pain points, this study conducted field investigations at three maritime higher vocational colleges in Jiangsu, distributing 216 questionnaires to teachers and students, interviewing 8 shipping captains, and observing 12 class hours of instruction. Quantitative and qualitative data jointly summarize four implementation dilemmas, distinct from the macro theoretical pain points outlined in the introduction. This section focuses on concrete obstacles teachers encounter in AI operation.

## 3.1. Monolithic AI Tool Usage Without Dual-Model Collaborative Division of Labor

87.3% of surveyed teachers only utilize a single general large language model for lesson preparation and pronunciation correction, failing to distinguish functional differences between text generation and oral interaction models. General large models tend to produce erroneous descriptions of ship communication workflows and STCW clauses when generating maritime scripts. Without AI assistants specializing in professional text proofreading, teachers spend excessive time revising content. Furthermore, a single model cannot simultaneously support mass learning performance calculation and interactive voice dialogue, creating technical deficiencies in pre-class differentiation and in-class training links.

## 3.2. Fragmented Teaching Segments Lacking a Pre-Class–In-Class–Post-Class Closed Loop

Existing AI applications are mostly confined to isolated links such as classroom exercises and oral dialogues. There is an absence of AI-driven pre-class learning diagnosis and differentiated preview push before class, as well as full-cycle data aggregation and dynamic differentiation adjustment after class. Uniform preview materials and assignments fail to match students' varied English foundations. Differentiated teaching relies on manual classification, which is labor-intensive and subjectively biased. Monthly learning data cannot feed back to optimize instructional design, leading to fragmented implementation.

## 3.3. Formalistic Process Assessment With Missing Quantitative Standards

91.7% of teachers support the implementation of process evaluation yet lack quantifiable composite AI-human assessment schemes. Traditional process records depend on manual teacher logging, unable to retain complete data of oral interaction, preview, and error exercises. No three-dimensional weighted assessment criteria integrating AI voice quantification, teacher qualitative evaluation, and industrial grading exist for implicit post competencies such as on-site maritime emergency communication and cross-cultural adaptability. Process score assignment is highly subjective, and assessment scales lack reliability and validity testing, undermining the persuasiveness of evaluation results.

## 3.4. Absent Risk Balancing Mechanisms for AI Teaching

Field investigations reveal two prevalent risks in frontline teaching. First, AI hallucinations distort professional terminology and maritime workflows without a dual calibration workflow of "AI generation – preliminary teacher review – final captain approval"[12]. Second, students over-rely on AI to complete dialogues and written compositions, weakening independent language output capabilities. Meanwhile, uneven digital operational proficiency among students creates a new digital stratification divide[13]. Current teaching schemes lack standardized risk restraint norms regulating AI usage boundaries and differentiation gradients.

# 4. Architecture of the Dual Generative AI Collaborative Full-Process Maritime English Teaching Model for Higher Vocational Colleges

## 4.1. Model Design Philosophy

Integrating situated cognition theory and the concept of distributed agency, this study positions AI as a collaborator rather than a substitute in teaching activities, constructing a dual-AI collaborative framework centered on "human-machine labor division and complementary advantages". The framework encompasses four core elements: differentiated learners, standardized maritime scenario resources, human-machine interaction rules, and three-dimensional assessment. DeepSeek builds rigorous professional maritime text libraries to mitigate hallucination risks, while Doubao constructs low-anxiety interactive oral scenarios to ease students' reluctance to speak. Teachers and shipping mentors act as calibration subjects to counteract systematic disturbances including AI content bias and student cognitive laziness. The framework incorporates operating mechanisms of differentiated adaptation, monthly dynamic iteration, and triple risk balancing.

## 4.2. Functional Division of Dual Models

DeepSeek Functions: Maritime professional terminology verification; standardized VTS/emergency scenario script generation; batch learning data calculation; differentiation threshold judgment; intelligent grading of written assignments; maritime error question aggregation.

Doubao Functions: Lightweight pre-class listening and speaking assessments; interactive role-play dialogue simulation; real-time oral pronunciation correction; lightweight push of differentiated preview materials; real-time storage of in-class interaction data.

Collaborative Functions of Teachers & Shipping Mentors: Overall teaching design; intensive explanation of key and difficult classroom points; dual review of AI-generated content; subjective scoring in three-dimensional assessment; dynamic adjustment of monthly differentiation gradients; guidance on students' digital literacy.

## 4.3. Reproducible Three-Stage Teaching Implementation Framework

The complete operating logic is divided into three phases: pre-class differentiated guided learning, in-class maritime situated training, and post-class multi-dimensional iterative assessment.

Pre-class Differentiated Guided Learning: Differentiation thresholds are set based on pre-test score distributions and seafarer competency exam passing standards: total score below 60 = Basic Level; 60–79 = Intermediate Level; ≥80 = Advanced Level. Core reproducible differentiation prompt: Based on students' total listening and speaking scores in Maritime English pre-assessment, classify learners into three tiers with thresholds of 60 and 80. Generate gradient preview materials compliant with STCW VII/2 Convention covering three scenario categories: daily ship operations, port services, and maritime emergencies, avoiding errors in maritime terminology and communication procedures, while outputting a diagnostic report on common class weaknesses. After Doubao completes pre-assessment, data is imported into DeepSeek to draft differentiated materials, which are revised by teachers to eliminate maritime deviations before distributing preview tasks.

In-Class Maritime Situated Training: All scenario scripts generated by DeepSeek must receive final approval from shipping captains prior to classroom application. Doubao simulates VTS operators and foreign seafarers for real-time voice dialogue while synchronously recording oral error data. Teachers reduce repetitive vocabulary and grammar instruction, focusing on interpreting post competencies and cultivating professional literacy to retain classroom dominance.

Post-Class Multi-Dimensional Assessment & Monthly Iteration: AI automatically aggregates error questions and pushes differentiated extended tasks. Weighted process assessment accounts for 50% of the total grade: 40% AI voice quantification, 35% on-site teacher evaluation, 25% shipping captain post-standard scoring; final written examination accounts for the remaining 50%. Full-cycle learning data is compiled monthly to dynamically adjust differentiation thresholds and task difficulty based on changes in student ability, enabling iteration of the virtual linguistic ecosystem.

## 4.4. Three-Dimensional Full-Process Assessment System

A self-developed Three-Dimensional Process Assessment Scale for Higher Vocational Maritime English covers three dimensions: written maritime competence, oral emergency communication, and cross-cultural & seafarer professional literacy. The overall scale Cronbach's $\alpha = 0.872$, with each sub-dimension $\alpha > 0.8$. Exploratory factor analysis yields a cumulative variance interpretation rate of 76.35%, indicating qualified reliability and validity capable of quantifying implicit post competencies such as on-site seafarer emergency response.

# 5. Teaching Empirical Research and Results

## 5.1. Empirical Research Design

Two parallel classes from a maritime higher vocational college in Jiangsu were selected: 43 students in the experimental class received dual generative AI collaborative instruction, while 42 students in the control class adopted traditional lecture-based teaching for one semester. The two groups shared identical textbooks, class hours, instructors, and student demographic structures.

A mixed research methodology was adopted, including literature review, frontline field investigation, four cycles of four-week action research, and parallel controlled experiments. Quantitative analytical tools included independent samples t-test, ANCOVA, Cohen's d effect size, and reliability & validity testing. Control variables: a unified pre-test was administered before instruction, with pre-test scores used as covariates to calibrate post-test outcomes. Equal total class duration, assignment volume, and training frequency were maintained across groups to eliminate interference from unequal instructional input.

## 5.2. Quantitative Empirical Results

Baseline Equilibrium Test of Comprehensive Maritime English Scores

Pre-test: Experimental class $M = 56.32$, $SD = 7.14$; Control class $M = 55.97$, $SD = 6.89$. Independent samples t-test $p = 0.712 > 0.05$, indicating no significant initial performance

difference and balanced baselines. After ANCOVA calibration, the experimental class post-test M = 72.15 versus the control class M = 67.54, with significant inter-group differences ($p < 0.05$). Cohen's d = 0.51, falling at the lower bound of a medium effect size. Effect gains contain short-term Hawthorne novelty effects, requiring cross-academic-year follow-up verification for long-term stable improvements.

Maritime Oral Situational Test

The oral proficiency pass rate reached 86.05% in the experimental class versus 74.70% in the control class, with Cohen's d = 0.48 approaching a medium effect size. Low-pressure AI interaction effectively alleviated oral speaking anxiety.

Classroom Learning Engagement Scale

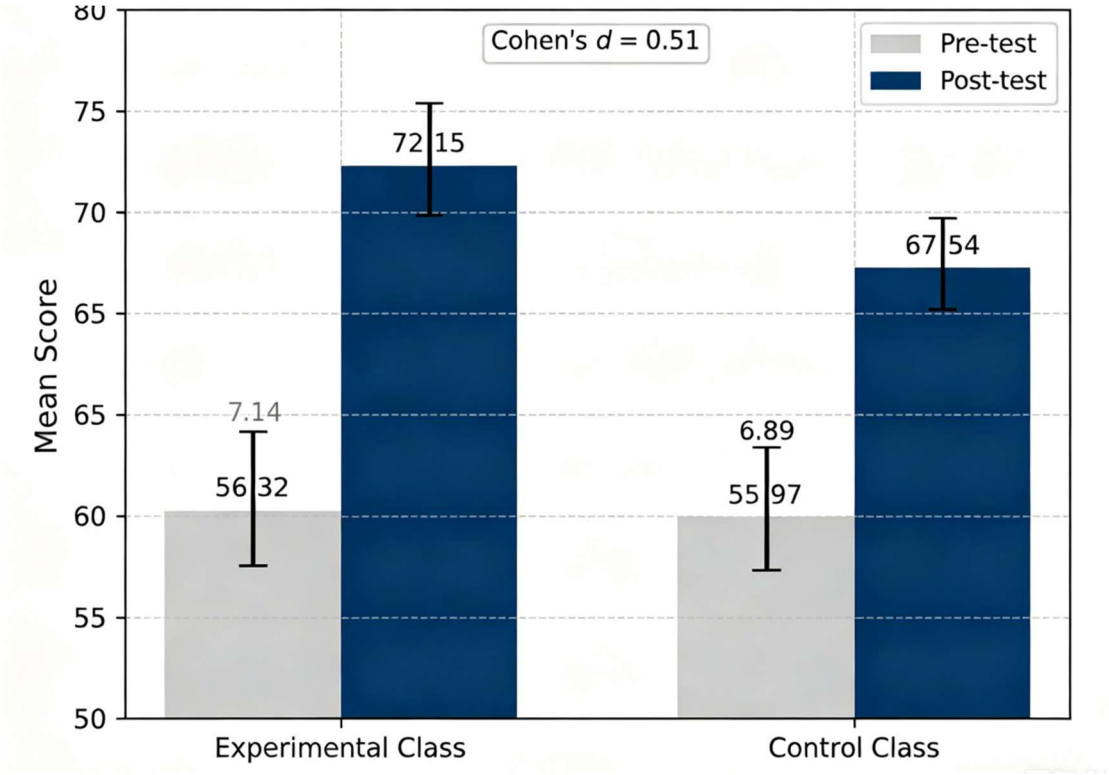



**Figure 1.** Pre-test and Post-test Score Comparison (Grouped Bar Chart)

The experimental class scored significantly higher than the control group ($p < 0.05$, Cohen's d = 0.54), with prominent improvements in active in-class dialogue and task completion participation.

Reliability and Validity of the Assessment Scale

The three-dimensional process assessment scale achieved an overall Cronbach's $\alpha = 0.872$ with a stable three-factor structure, capable of objectively measuring students' full-cycle comprehensive maritime competence.

## 5.3. Qualitative Practical Conclusions and Risk Balancing Strategies

Combining 12 hours of classroom observation, 26 teacher-student interviews, and professional evaluations from 8 shipping captains, three practical values of the model are summarized. First, distributed dual-model labor division fills gaps in maritime scenario and differentiated training resources. Second, AI undertakes mechanical work including material development and assignment grading, freeing teachers' time for personalized tutoring and literacy cultivation. Third, the dual review mechanism of "teachers + captains" drastically reduces the probability of AI maritime hallucinations.

Standardized balancing schemes are matched to three implementation constraints identified in the investigation:

High calibration costs for AI professional content: Build a dedicated maritime prompt library to unify scenario generation standards and cut iterative revision workloads.

Weakened independent expression from AI over-reliance: Clarify classroom AI usage boundaries and raise assessment weights for on-site impromptu dialogue and original maritime writing.

Digital literacy stratification divide: Deploy 5–10 minute pre-class micro-lessons on digital operations to narrow gaps in tool proficiency.

## 5.4. In-Depth Interpretation of Empirical Limitations

First, the experimental cycle spans only one semester, with intervention effects subject to obvious Hawthorne novelty effects driven by student curiosity toward AI tools. Long-term educational stability across academic years cannot be confirmed, requiring two-year follow-up experiments to isolate short-term gains.

Second, the sample is limited to a single maritime higher vocational college in Jiangsu, lacking coverage of regional institutions and diverse student cohorts, leaving the theory and model's generalizability to be verified via cross-institutional horizontal comparison experiments.

Third, differentiation thresholds of 60/80 are manually set based on existing exam standards without adaptive differentiation algorithms built on long-term student learning big data, leaving room to improve intelligent matching precision.

Fourth, this study did not collect data on teachers' TPACK time allocation, lacking quantitative comparisons of lesson preparation and tutoring durations under the dual-model framework. Empirical evidence regarding human-machine

collaboration's impacts on teachers' instructional behaviors remains absent.

# 6. Conclusion

## 6.1. Core Research Conclusion

Targeting four prominent pain points of Maritime English in higher vocational colleges - insufficient scenario provision, unenforceable differentiated teaching, scarce oral training, and single summative assessment - this study constructs an original middle-range theory of distributed maritime virtual linguistic ecosystems supported by situated cognition, distributed agency, and human-machine collaborative resilience theories. A full-process teaching framework with dual collaborative generative AI (DeepSeek & Doubao) is established to form an integrated closed-loop system of "pre-class differentiated guided learning – in-class embodied maritime situated training – post-class three-dimensional multi-dimensional assessment".

One semester of balanced baseline controlled experiments and four rounds of iterative action research verify that fixed differentiation thresholds and standardized maritime prompts enable highly reproducible differentiated teaching, generating large volumes of standardized post interaction scenarios compliant with STCW conventions. ANCOVA and Cohen's d results demonstrate medium positive intervention gains in the experimental class's written performance, oral proficiency, and classroom learning engagement. The self-developed three-dimensional assessment scale exhibits sound reliability and validity to objectively quantify implicit on-site seafarer emergency competencies. This study establishes a dual content review workflow of "AI draft – preliminary teacher audit – final shipping captain audit", systematically counteracting three teaching risks: AI hallucinations, digital literacy divides, and student cognitive laziness. A lightweight, transferable complete set of standardized teaching reform implementation specifications is formed, delivering a localized paradigm integrating theoretical construction and empirical evidence for digital differentiated ESP education in domestic maritime higher vocational colleges.

## 6.2. Prospects for Further Research

First, extend the experimental cycle to conduct cross-academic-year follow-up teaching to isolate Hawthorne novelty effects and test the model's long-term educational resilience. Collaborate with maritime higher vocational colleges across multiple regions to carry out horizontal comparison experiments, refine the adaptive boundary of the distributed maritime virtual linguistic ecosystem theory, and improve the generalizability of research conclusions.

Second, develop adaptive differentiation algorithms based on long-term student learning big data to realize fully automatic dynamic adjustment of differentiation thresholds, reduce teachers' manual differentiation workload, and further enhance personalized matching precision.

Third, integrate VR maritime virtual simulation scenarios with dual generative AI to build an integrated virtual-physical embodied training classroom, strengthen multi-sensory post linguistic experience, and deepen the application of situated cognition theory in Maritime English instruction.

Fourth, collect supplementary data on teachers' TPACK classroom time allocation to quantitatively compare shifts in teachers' work structures between traditional classrooms and the dual-AI collaborative model, supplementing empirical evidence on human-machine collaboration's impacts on teaching practitioners. Simultaneously develop pre-course digital literacy micro-lessons to eliminate digital stratification divides.

# References


[1] Wang, H. X. (2024). Digital transformation path of ESP courses in higher vocational education from the perspective of smart education. *Chinese Vocational and Technical Education*, (28), 62–68.

[2] Dong, J. W. (2025). Practical dilemmas and countermeasures of maritime English teaching in higher vocational colleges. *Times Characters*, (30), 175–176.

[3] Wang, S. R. (2025). Restructuring vocational-oriented English teaching in higher vocational colleges in the new era. *Foreign Language World*, (01), 41–47.

[4] Wang, J. (2025). An analysis of integrating maritime vocational safety culture into higher vocational English teaching from the perspective of AIGC. *Pearl River Water Transport*, (10), 106–108.

[5] Zhang, Y. X. (2025). The 321 teaching model of generative AI in marine engineering English writing. *Overseas English*, (22), 112–115.

[6] Mohamed, A. (2025). AI-generated maritime English training materials for seafarers. *Maritime Education*, 19(2), 76–92.

[7] Godwin-Jones, R. (2024). Distributed agency in second language learning and teaching through generative AI. *Language Learning & Technology*, 28(2), 5–30.

[8] Kukulska-Hulme, A. (2024). Mobile language learning: Reflecting on the past and envisioning the future. In (Book Title Editor) *Theory and practice in vocabulary research in digital environments*. Routledge.

[9] Hockley, T. (2024). Generative AI and digital divide in EFL teaching. *ELT Journal*, 78(4), 478–485.

[10] Li, Y., & Zhao, C. Y. (2022). Blended teaching practice of maritime English based on OBE concept. *Maritime Education Research*, (04), 67–71.

[11] Zhuo, Y. Q. (2025). Design of outcome-oriented role-play teaching method for maritime English. *Maritime Education Research*, (04), 43–47.

[12] Li, M. (2025). Risks and balancing paths of generative AI hallucinations in foreign language teaching. *Computer-Assisted Foreign Language Education*, (02), 18–25.

[13] Chen, X. (2024). LLM-assisted differentiated adaptation strategies for higher vocational English from the perspective of digital literacy divide. *Modern Foreign Languages*, (06), 821–830.