World Journal of Innovation and Modern Technology, Vol. 9, Issue 10 (Oct) 2026
ISSN 2682-5910
DOI: 10.66069/ojspub.26821009

From Research to Classroom: Transforming a Teacher–Student–AI Agent Collaborative Model into ESP Teaching Materials and Cases

Xiaoli Hu

Jiangsu Maritime Institute, Nanjing 211170, Jiangsu, China

Abstract: ***The persistent research–practice gap in vocational education—whereby research outcomes rarely materialize as usable classroom resources—remains a critical bottleneck in English for Specific Purposes (ESP) instruction. Although prior work has established the theoretical viability and empirical effectiveness of a Teacher–Student–AI Agent (TSA) collaborative model for vocational ESP, the translational mechanism through which such research is systematically converted into textbooks and teaching cases has not been examined. This paper reports a research-to-teaching project that operationalized the TSA model on the Chaoxing AI Agent platform and translated its core components into a modular loose-leaf textbook, Cruise Service Post English, together with three scenario-based teaching cases. A four-stage transformation pathway is proposed—outcome distillation, pedagogical reengineering, material and case development, and iterative classroom validation—and each stage is documented with its design decisions, outputs, and revision triggers. The textbook's workplace-task architecture and the three cases (front-desk check-in, cabin service complaint resolution, and medical emergency response) are presented in detail, with an explicit mapping from research components to material forms. Formative validation was conducted in two intact classes (N = 94) over a 16-week semester using teacher reflection logs, student feedback forms, focus group interviews, and platform interaction logs. Findings indicate high perceived relevance, reduced speaking anxiety, and improved instructional precision, while also revealing limitations in automatic speech recognition and the need for stronger teacher scaffolding for lower-proficiency learners. Because individual-level paired observations were not retained, only aggregate platform statistics are reported and no causal claim is advanced regarding the association between practice frequency and rubric scores. The study contributes to materials development research by showing that a platform-integrated AI agent can serve simultaneously as an instructional collaborator and as a source of evidence for curriculum revision, offering a replicable and traceable paradigm for vocational ESP instructors.***

Keywords: Research-to-teaching, Teacher–Student–AI Agent, Vocational ESP, Loose-leaf textbook, Teaching case development, Materials development, AI literacy.

1. INTRODUCTION

Vocational English for Specific Purposes (ESP) instruction in Chinese higher vocational colleges faces a persistent structural tension: faculty members are increasingly expected to undertake funded research projects, yet the outputs of these projects—models, frameworks, and empirical findings—rarely travel the “last mile” into the textbooks, lesson plans, and assessment tasks that shape students' daily learning experiences. This disjuncture, commonly described as the “research–practice gap” or the “two skins” problem, is particularly acute in applied language teaching, where findings on collaborative learning, AI integration, and scenario-based pedagogy frequently remain at the level of academic articles rather than becoming the materials that teachers actually use in class (Tomlinson, 2013).

The present study addresses this gap by reporting a deliberate research-to-teaching transformation project situated within the Cruise Service English course at Jiangsu Maritime Institute. The project takes as its source a completed research program that developed and tested a Teacher–Student–AI Agent (TSA) collaborative model for vocational ESP through a quasi-experimental design (Hu, 2026). The question animating the present work is not whether the TSA model works—that has been addressed previously—but how its constituent elements can be systematically reengineered into teaching materials and cases usable by instructors who were not part of the original research team.

Three specific objectives guide the study. First, to articulate a transformation pathway that moves from research outcomes to pedagogical resources in a principled, traceable manner. Second, to document the development of *Cruise Service Post English*, a loose-leaf textbook organized around the TSA model's scenario bank and assessment logic. Third, to present three fully deconstructed teaching cases—AI-simulated front desk check-in, cabin service complaint resolution, and medical emergency response—that illustrate how the model's four stages

(pre-class diagnosis, in-class interaction, post-class evaluation, and continuous optimization) are enacted in concrete instructional sequences.

The significance of this work lies in its attempt to render the research-to-teaching process itself an object of systematic inquiry. Scholarship on vocational ESP has produced valuable models and empirical findings, yet it has paid comparatively little attention to the translational infrastructure—the design decisions, material formats, and validation mechanisms—that determines whether those findings ever reach students. Three research questions are therefore posed: (a) which components of a completed research model are transferable into teaching materials, (b) through what design decisions is that transfer accomplished, and (c) what forms of classroom evidence can be used to validate the resulting materials formatively? By documenting this infrastructure in detail, the study offers a template that can be adapted across ESP domains, from maritime English to medical and business English.

2. LITERATURE REVIEW

2.1 The Research-to-Teaching Gap in Vocational Education

The expectation that research should inform teaching is widely endorsed in policy discourse, yet empirical accounts of how this occurs in vocational contexts remain sparse (Macalister & Nation, 2020). In Chinese vocational colleges, research outputs are frequently assessed through publication counts and project completions rather than through evidence of curricular impact. This incentive structure contributes to a "research–practice divide" in which faculty develop sophisticated models that remain disconnected from the textbooks and lesson materials their students encounter.

Recent scholarship has begun to address this gap by examining the conditions under which research outputs can be translated into pedagogical resources. Studies of curriculum development in vocational ESP emphasize that translation is not a straightforward application of findings but a design-intensive process requiring attention to learner proficiency, instructional time constraints, and the affordances of available platforms. The loose-leaf, or modular, textbook format has emerged as a particularly promising vehicle for such translation, precisely because its structure allows research-derived scenarios and tasks to be updated without wholesale revision of a bound textbook. What remains undocumented, however, is the sequence of design decisions that connects a specific research model to a specific set of classroom materials.

2.2 AI Agents and Scenario-Based ESP Instruction

The integration of AI agents into language instruction has moved beyond the question of whether such tools can support learning to the question of how they should be configured within coherent pedagogical designs. In vocational ESP, AI conversational agents and scenario-dialogue functions offer a means of simulating workplace communication that traditional classrooms struggle to replicate (Godwin-Jones, 2022). The TSA model, grounded in symbiotic theory, positions the AI agent as a "digital collaborator" with diagnostic, interactive, and evaluative functions rather than as a passive tool or an autonomous tutor (Wang & Lee, 2026).

Crucially for the present study, the AI agent's role in scenario-based practice is not limited to student-facing interaction. The agent also generates data—diagnostic profiles, interaction logs, and performance scores—that can inform both instructional adjustment and material revision. This dual function (instructional and informational) makes the agent a potentially powerful instrument not only for teaching but for the iterative development of teaching materials themselves (Pokrivcakova, 2019). The present study exploits this duality by treating interaction logs and scoring outputs as inputs into the revision of the textbook and cases.

Empirical work in closely related contexts is emerging but remains oriented toward intervention effects rather than translation processes. In a quasi-experimental study of maritime ESP at a Chinese vocational college, Wang (2025) combined a structured BOPPPS lesson design with LLM-generated texts and reported gains in language proficiency together with positive learner attitudes. Such studies establish that AI-supported designs can be effective; they do not, however, report how an evaluated model was subsequently converted into reusable materials for instructors outside the research team. The present study targets that missing translational step.

2.3 Loose-Leaf Textbooks as Translational Vehicles

The loose-leaf textbook format has been promoted in vocational education as a format suited to the pace of industry change and the diversity of learner needs. Its defining features—modular units, task-based organization, space for formative recording, and capacity for local adaptation—align well with the requirements of research-to-teaching translation (Tomlinson & Masuhara, 2018). A research-derived scenario can be inserted or revised as a discrete module; an assessment rubric can be appended as a standalone component; an AI interaction guide can be updated as platform functions evolve.

However, the loose-leaf format also poses design challenges. Without a strong organizing logic, modularity can devolve into fragmentation. The present study addresses this challenge by adopting a workplace-task architecture: each unit is organized around a recognizable cruise service task (e.g., boarding welcome, cabin inquiry, complaint handling), and within each unit the TSA stages structure the learning sequence. This architecture preserves the research model's logic even as individual components are revised.

3. THE TRANSFORMATION PATHWAY

The transformation from research to teaching materials proceeded through four iterative stages, each with distinct design tasks and outputs. This pathway was not a linear sequence but a cycle in which classroom feedback from later stages prompted revisions to earlier design decisions. Table 1 summarizes the pathway and its feedback structure; Sections 3.1–3.4 describe the stages and Section 3.5 specifies the validation procedures.

Table 1: The Four-Stage Research-to-Teaching Transformation Pathway

Stage	**Key Design Task and Output**
Stage 1: Outcome Distillation ↓	Identify conceptual, technical, and assessment components: three-subject role structure, four-stage cycle, scenario bank, prompt templates, and analytic rubrics.
Stage 2: Pedagogical Reengineering ↓	Map each component onto a material form: role cards, unit template, task briefs, AI interaction guides, and reflection/assessment sections (Table 1).
Stage 3: Material & Case Development ↓	Produce the eight-unit loose-leaf textbook Cruise Service Post English and three elaborated teaching cases (Tables 2–3).
Stage 4: Iterative Classroom Validation ↓	Collect teacher logs, student feedback, focus group interviews, and platform logs; revise prompts, add troubleshooting support, and rebalance in-class reflection.
↺ Feedback loop	Validation findings return to Stages 1–3, so the pathway is cyclical rather than linear.

Note. Downward arrows indicate the sequence of design work; the return arrow denotes the feedback loop through which classroom evidence revises earlier design decisions.

3.1 Stage One: Outcome Distillation

The first stage involved distilling the research program into its transferable components. Three categories of outcome were identified. Conceptual components included the three-subject role structure (teacher as designer and interpreter, student as active practitioner and reflector, AI agent as diagnostic and evaluative collaborator) and the four-stage instructional cycle. Technical components included the scenario bank comprising eight cruise service tasks, the AI agent configuration scripts, and the prompt templates used to structure scenario dialogues. Assessment components included the analytic rubrics for oral performance (fluency, lexical resource, grammatical accuracy, pronunciation) and the platform-generated learning analytics indicators.

Distillation required judgment about which components were essential to the model's integrity and which could be adapted or omitted. The three-subject role structure and the four-stage cycle were treated as non-negotiable design principles; specific scenario contents and prompt phrasings were treated as adaptable. This distinction is recorded explicitly because it determines which parts of the materials may be localized without compromising the model's theoretical coherence.

3.2 Stage Two: Pedagogical Reengineering

The second stage translated research components into instructional materials. Each conceptual component was mapped onto a corresponding material form. The role structure became a "role card" system specifying what the teacher, student, and AI agent are expected to do at each stage. The four-stage cycle became the unit template: every unit opens with a pre-class diagnostic task, proceeds through in-class scenario dialogue and teacher-guided reflection, and closes with post-class AI evaluation and a student self-assessment record. The scenario bank became eight textbook units, each anchored to a workplace task and populated with vocabulary support, AI interaction scripts, and a professional communication rubric. Table 1 maps each research component onto its reengineering logic and resulting material form.

A key design decision at this stage concerned the level of AI specificity. Early drafts embedded detailed platform-specific click-by-click instructions, which risked rapid obsolescence. The final design instead articulated interaction principles—how to formulate a scenario prompt, how to structure a multi-turn dialogue, and how to evaluate the agent's response for professional appropriateness—that remain valid across platform updates.

Table 2: Mapping of Research Components onto Pedagogical Materials

Research Component (Source)	**Pedagogical Reengineering (Process)**	**Material Form (Output)**
Conceptual: Three-subject role structure (Teacher, Student, Agent)	Defined specific actions and expectations for each actor at every stage of the lesson.	Role Cards embedded at the beginning of each textbook unit.
Conceptual: Four-stage instructional cycle	Transformed into a standardized template for organizing every lesson.	Unit Template (Pre-class diagnosis → In-class interaction → Post-class evaluation → Optimization).
Technical: Scenario bank (8 cruise tasks)	Converted into specific workplace communicative tasks with contextual details.	Task Briefs and Professional Language Banks (8 textbook units).
Technical: AI prompt templates and interaction scripts	Simplified into general interaction principles to ensure durability across platform updates.	AI Interaction Guides providing step-by-step prompt strategies and troubleshooting.
Assessment: Analytic rubrics and learning analytics	Adapted for student self-assessment and peer review, moving beyond exclusive teacher use.	Reflection and Assessment sections (Rubrics + structured reflection prompts).

Note. The table illustrates the systematic translation of abstract research outputs into concrete, usable instructional materials for vocational ESP classrooms. It corresponds to Sections 3.1 and 3.2.

3.3 Stage Three: Material and Case Development

The third stage produced the two principal deliverables: the loose-leaf textbook and three teaching cases. The textbook, *Cruise Service Post English*, comprises eight units organized around the cruise service cycle. Each unit follows a four-part structure: (1) a Task Brief describing the workplace scenario and communicative goal; (2) an AI Interaction Guide providing prompt templates and dialogue strategies; (3) a Professional Language Bank with service expressions and register notes; and (4) a Reflection and Assessment section containing the rubric and self-assessment prompts.

The three teaching cases were developed as elaborated exemplars of the textbook's unit logic, intended for use in teacher training and peer observation. Each case documents the full instructional sequence—pre-class, in-class, and post-class—with specific attention to teacher moves, expected AI agent responses, and contingencies for when the agent's output is inadequate or inaccurate.

3.4 Stage Four: Iterative Classroom Validation

The fourth stage implemented the materials in two intact Cruise Service English classes ($N = 94$) over a 16-week semester. Validation focused not on learning outcomes (which had been examined in the prior quasi-experimental study reported by Hu, 2026) but on material usability and instructional fit. Data sources included weekly teacher reflection logs, student end-of-unit feedback forms, platform interaction logs, and end-of-semester focus group interviews with 12 students and both teachers. Revisions triggered by this feedback included simplifying the

prompt templates for lower-proficiency students, adding a troubleshooting guide, and lengthening the in-class teacher-guided reflection segment relative to independent AI practice.

3.5 Validation Design and Analytic Procedures

To render the validation auditable, its design is specified here. Participants were the 94 students enrolled in the two intact classes and the two teachers who delivered the course; no comparison group was constituted, because the validation targeted usability and instructional fit rather than learning gains. Four data sources were collected across the semester: (a) weekly teacher reflection logs, (b) end-of-unit student feedback forms, (c) end-of-semester focus group interviews with 12 students and both teachers, and (d) platform interaction logs exported from the Chaoxing AI Agent environment. Interview and open-response data were examined following the six-phase thematic analysis procedure (Braun & Clarke, 2006): familiarization, initial coding, searching for themes, reviewing themes, defining and naming themes, and reporting. Coding was conducted by the author; because the analysis was formative and no second coder was involved, no inter-rater reliability coefficient is reported (Koo & Li, 2016). Platform logs were summarized descriptively using means, standard deviations, and a Pearson correlation. Quoted comments are reproduced as recorded in the project documentation.

4. TEXTBOOK AND CASE DESIGN

4.1 The Loose-Leaf Textbook Structure

The textbook's eight units correspond to the scenario bank's workplace tasks: boarding welcome and document verification, cabin facility introduction, dining service and dietary inquiry, entertainment and activity explanation, safety briefing and emergency instruction, complaint handling and service recovery, disembarkation assistance, and cross-departmental coordination. Each unit is designed as a self-contained loose-leaf module, permitting instructors to select and sequence units according to course hours and learner needs.

The four-part unit structure translates the TSA model's logic into a usable form. The Task Brief establishes the professional context and communicative objective, ensuring that language practice is anchored to a recognizable service encounter. The AI Interaction Guide provides scaffolding for the student–agent dialogue, including opening prompts, follow-up strategies for sustaining the interaction, and criteria for judging whether the agent's response is professionally appropriate. The Language Bank supplies the lexical and phraseological resources that the scenario demands. The Reflection and Assessment section presents the analytic rubric and prompts students to identify specific moments in the AI dialogue where their communicative choices could be improved. Table 3 summarizes this structure, and Table 4 provides a comparative overview of the three cases detailed in Sections 4.2–4.4.

Table 3: Structure of the Loose-Leaf Textbook Units

Unit Section	Core Function	Integration with the TSA Model
1. Task Brief	Establishes the professional context, roles, and communicative goals.	Anchors language practice to a recognizable service encounter, framing the scenario for the AI agent.
2. AI Interaction Guide	Provides opening prompts, follow-up strategies, and troubleshooting tips.	Scaffolds the Student–Agent interaction, enabling students to sustain a multi-turn dialogue.
3. Professional Language Bank	Supplies lexical and phraseological resources organized by communicative functions.	Supports linguistic accuracy and register appropriateness, informed by the research corpus.
4. Reflection & Assessment	Presents the analytic rubric and structured reflection prompts.	Facilitates the Post-class Evaluation and Continuous Optimization stages of the TSA cycle.

Note. Each unit is a self-contained module, allowing instructors to sequence units based on course hours and learner proficiency.

Table 4: Deconstruction of the Three Representative Teaching Cases

Case Focus	Pre-Class (Agent & Student)	In-Class (Teacher, Student, Agent)	Post-Class (Agent & Student)
Case 1: Front Desk Check-In	Agent acts as a guest with an expired visa page; the student diagnoses baseline ability to verify documents and offer solutions.	Teacher models one polite exchange; students practice repeatedly with the agent, who responds with confusion if the student is pragmatically abrupt.	Agent generates a performance summary; the student writes a reflection on one "communication repair" to make next time.
Case 2: Cabin Service Complaint	Agent simulates a guest complaining about cabin noise; the student is tested on baseline empathy and apology moves.	Teacher shows a video of a service encounter to set standards; students practice while the agent escalates dissatisfaction if the guest's perspective is not acknowledged.	Agent provides rubric scores; the student writes a "service recovery reflection" connecting practice to industry standards.
Case 3: Medical Emergency	Agent is configured with a time-sensitive dialogue structure; the student practices gathering information under simulated pressure.	Teacher emphasizes the difference between conversational fluency and emergency clarity; students practice issuing clear, sequenced instructions.	Agent analyzes dialogue logs; teacher and students debrief on information accuracy and instruction sequencing.

Note. The cases were developed as elaborated exemplars of the textbook's unit logic for use in teacher training and peer observation; see Sections 4.2–4.4.

4.2 Case One: AI-Simulated Front Desk Check-In

The first case simulates a foreign guest's check-in at the cruise terminal. In the pre-class stage, students complete a diagnostic dialogue with the AI agent configured as a guest with a documentation discrepancy (e.g., an expired visa page). The agent's diagnostic output identifies whether the student's baseline performance includes the target language functions: verifying identity, explaining the discrepancy politely, and proposing a resolution pathway. The teacher reviews these profiles before class and selects the focus for in-class work.

In class, the teacher introduces the scenario and models one exchange before students pair with the agent for repeated practice. The agent responds contingently: if the student produces a grammatically accurate but pragmatically abrupt request, the agent may respond with hesitation or confusion, prompting the student to reformulate. The teacher circulates, notes recurring difficulties, and pauses the class for brief whole-group clarification. Post-class, the agent generates a performance summary across the four rubric dimensions, and students complete a structured reflection identifying one "communication repair" they would make in a subsequent attempt.

4.3 Case Two: Cabin Service Complaint Resolution

The second case involves a guest complaint about cabin noise. The AI agent is configured to express escalating dissatisfaction if the student's responses are dismissive or fail to acknowledge the guest's perspective. This design deliberately creates pragmatic pressure: the student must deploy apology, empathy, and solution-offering moves in an appropriate sequence. The pre-class diagnostic assesses the student's baseline use of these moves. In class, the teacher uses a brief video of a service encounter to establish the professional standard, and students then practice with the agent while the teacher monitors and intervenes where a student becomes stuck. Post-class evaluation includes both the agent's rubric scores and a student-written "service recovery reflection" that connects the AI practice to professional service standards.

4.4 Case Three: Medical Emergency Response

The third case simulates a medical emergency requiring the student to gather information, issue clear instructions, and coordinate with shore-based support—all in English under time pressure. The AI agent is configured with a time-sensitive dialogue structure: delayed or ambiguous responses from the student trigger increasingly urgent prompts from the agent. This case is the most demanding of the three and is positioned late in the textbook sequence. In class, the teacher emphasizes the difference between conversational fluency and emergency communication clarity and uses the agent's dialogue log to conduct a post-practice debrief focused on information accuracy and instruction sequencing.

5. IMPLEMENTATION AND EVALUATION

5.1 Teacher Perceptions

End-of-semester interviews with the two participating teachers revealed three themes, identified through the thematic analysis procedure described in Section 3.5 (Braun & Clarke, 2006). First, the loose-leaf format reduced the perceived burden of adopting a research-based model: “I didn't have to redesign my whole course. I could start with one unit and see how it worked.” Second, the AI Interaction Guide was identified as the most valuable component, particularly its troubleshooting section: “When the AI didn't understand a student, the guide told them what to try next, so I wasn't answering the same question twenty times.” Third, teachers reported that the in-class reflection segment was essential: without it, students tended to treat the AI dialogue as a game rather than as professional practice.

5.2 Student Perceptions

Student feedback forms and focus group interviews indicated high perceived relevance of the workplace tasks. A representative comment was: “The scenarios are what we will actually face—checking documents, handling complaints. It doesn't feel like textbook English.” Students also reported reduced speaking anxiety relative to whole-class oral performance, attributing this to the private, repeatable nature of AI dialogue. However, students at lower proficiency levels reported frustration when the agent's speech recognition failed to capture their pronunciation: “Sometimes the AI doesn't understand me, and I don't know if it's my English or the system.” This feedback prompted the addition of a troubleshooting module and a recommendation that lower-proficiency students begin with text-based rather than voice-based interaction.

5.3 Platform Interaction Patterns

Platform logs for the 94 students showed a mean of 18.3 AI practice sessions per student (SD = 6.7) across the semester, with a mean total practice duration of 241 minutes (SD = 72). Scenario dialogue completion rates averaged 76.8% (SD = 14.2 percentage points), somewhat lower than the 82.4% reported in the prior quasi-experimental study (Hu, 2026); this difference is plausibly attributable to the broader range of student proficiency in the present implementation classes, although the two cohorts were not directly compared. Practice frequency was moderately associated with end-of-unit rubric scores, r = .43, p < .001. Because the analysis is correlational and no individual-level paired observations were retained, this association is not interpreted as a dose–response or causal effect; it indicates only that uptake of the materials' practice opportunities was uneven. Aggregate statistics are reported in Table 5 in place of the originally proposed scatter plot, which cannot be produced from the available records.

Table 5: Platform Interaction Statistics (N = 94)

Indicator	*M*	*SD*
AI practice sessions per student	18.3	6.7
Total practice duration (minutes)	241	72
Scenario dialogue completion rate (%)	76.8	14.2

Note. N = 94. Practice frequency was moderately associated with end-of-unit rubric scores, r = .43, p < .001. Completion rates are reported in percentage points. Individual-level paired observations were not retained in the project records; accordingly, no scatter plot is presented and the association is not interpreted as causal. The completion rate reported in the prior quasi-experimental study was 82.4% (Hu, 2026).

6. DISCUSSION

6.1 The Transformation Pathway as a Replicable Design

The four-stage pathway proposed here—distillation, reengineering, development, and iterative validation—offers a structured alternative to the implicit assumption that research findings will simply "trickle down" into teaching. Its value lies in its explicitness: each stage specifies design decisions and their rationales, making the translation process visible and therefore revisable (Macalister & Nation, 2020). For vocational ESP instructors who wish to draw on research without becoming researchers themselves, the pathway provides a manageable sequence: identify the transferable components, map them onto material forms, build the materials, and refine them through classroom use. The emphasis on making evidence visible also accords with syntheses identifying feedback as a high-impact influence on learning (Hattie, 2009).

A distinctive feature of this pathway is its treatment of the AI agent as both an instructional collaborator and a material-development instrument. The agent's interaction logs and diagnostic outputs informed revisions to the prompt templates and troubleshooting guides, creating a feedback loop in which classroom use directly shaped the evolution of the materials. This dual role—agent as teacher and agent as informant—departs from conventional materials development, which typically relies on teacher judgment and student feedback alone.

6.2 Design Tensions and Resolutions

Three tensions surfaced during development. First, specificity versus durability: platform-specific instructions were useful in the short term but risked rapid obsolescence. The resolution—articulating interaction principles rather than click-by-click procedures—sacrificed some immediate usability for longer-term resilience. Second, standardization versus flexibility: a uniform unit template facilitated teacher adoption but constrained adaptation to local conditions. The loose-leaf format mitigated this tension by permitting unit-level substitution while preserving the overall architecture. Third, AI autonomy versus teacher authority: the model positions the teacher as the ultimate interpreter of AI outputs, but in practice students sometimes treated AI scores as authoritative. The in-class reflection segment was strengthened to reassert the teacher's interpretive role.

6.3 Limitations

Several limitations constrain the claims that can be made. The validation reported here is formative rather than summative: it documents usability and instructional fit but does not assess learning outcomes against a comparison condition, and the prior quasi-experimental study (Hu, 2026) remains the source of efficacy evidence. The single-institution, single-course context limits generalizability, and thematic coding was conducted by a single researcher without a reliability check (Koo & Li, 2016). Agreement between AI-generated rubric scores and teacher-assigned scores was likewise not evaluated, so the reliability of the automated scoring remains unverified. Only aggregate platform statistics were retained; individual-level paired data were unavailable, precluding visual or multivariate analysis of the practice–performance association. The materials were developed for cruise service English, and although the transformation pathway is designed to be transferable, its application to other ESP domains would require substantial re-specification of scenarios, language banks, and AI configurations. Finally, students' difficulties with speech recognition point to a technical constraint that lies beyond the materials themselves.

6.4 Implications for Practice

For vocational ESP instructors, the study suggests that research-to-teaching translation is feasible without extensive additional resources if the translation task is broken into manageable stages and if the resulting materials are designed for modular adoption. For administrators, the findings support investment in loose-leaf materials development and in teacher training that focuses on the interpretive and reflective dimensions of AI-supported instruction. For platform developers, student feedback on speech recognition underscores the continued importance of robust voice interaction, particularly for learners whose pronunciation is still developing.

6.5 Future Research

Three lines of inquiry follow from these limitations. First, a controlled comparison is needed to establish whether

use of the translated materials produces learning gains beyond those documented for the original TSA model. Second, inter-rater agreement between AI-generated and teacher-assigned rubric scores should be estimated so that automated assessment can be interpreted with known reliability (Koo & Li, 2016). Third, replication across institutions and ESP domains would test whether the four-stage pathway itself—rather than the particular cruise service content—accounts for successful translation. Retaining individual-level platform data in future implementations would permit the visual and multivariate analyses that the present study could not conduct.

7. CONCLUSION

This paper has documented a research-to-teaching transformation project that converted the conceptual and empirical outputs of a TSA collaborative model study into a loose-leaf textbook and three teaching cases for vocational cruise service English. The four-stage pathway—distillation, pedagogical reengineering, development, and iterative validation—provides a structured account of how research findings can become classroom-ready materials without losing their theoretical coherence. The textbook's task-based architecture and the cases' detailed instructional sequences demonstrate that the TSA model's logic can be preserved in usable form, while classroom validation revealed both affordances (professional relevance, reduced speaking anxiety) and constraints (speech recognition limitations, the need for stronger teacher scaffolding at lower proficiency levels). The evidence reported here is formative and correlational; claims about learning effects rest on the prior quasi-experimental study (Hu, 2026). As vocational institutions continue to invest in both research and AI-supported teaching, the translational infrastructure described here offers a practical, auditable template for ensuring that the former genuinely informs the latter.

ACKNOWLEDGMENTS

This research is supported by the Jiangsu Maritime Institute Research-to-Teaching Special Project “Research on Teaching Transformation and Case Development of Scientific Research Achievements in the ‘Teacher–Student–AI Agent’ Collaborative Teaching Model for Vocational Industry English” (Project Number: 2026kyfb11). It is a phased achievement of the 2026 National Joint Conference of Presidents of Higher Vocational Colleges project “Research on the Teacher–Student–AI Agent Collaborative Teaching Model for Higher Vocational Industry English Based on AI Agents” (Project No. XLH202632). We extend our gratitude to the faculty and students of the Cruise Service Management program for their participation and feedback during classroom implementation of the teaching cases.

REFERENCES

[1] Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. Qualitative Research in Psychology, 3(2), 77–101. https://doi.org/10.1191/1478088706qp063oa
[2] Godwin-Jones, R. (2022). Partnering with AI: Intelligent writing assistance and instructed language learning. Language Learning & Technology, 26(2), 5–24. https://doi.org/10.64152/10125/73474
[3] Hattie, J. (2009). Visible learning: A synthesis of over 800 meta-analyses relating to achievement. Routledge.
[4] Hu, X. (2026). Effects of a Teacher–Student–AI Agent collaborative model on oral English proficiency and learning motivation in vocational cruise English: A quasi-experimental study with triangulated evidence. Journal of Research in Vocational Education, 8(9), 20–27. https://bryanhousepub.com/index.php/jrve/article/view/4019
[5] Koo, T. K., & Li, M. Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. Journal of Chiropractic Medicine, 15(2), 155–163. https://doi.org/10.1016/j.jcm.2016.02.012
[6] Macalister, J., & Nation, I. S. P. (2020). Language curriculum design (2nd ed.). Routledge.
[7] Pokrivcakova, S. (2019). Preparing teachers for the application of AI-powered technologies in foreign language education. Journal of Language and Cultural Education, 7(3), 135–153. https://doi.org/10.2478/jolace-2019-0025
[8] Tomlinson, B. (Ed.). (2013). Developing materials for language teaching (2nd ed.). Bloomsbury Academic.
[9] Tomlinson, B., & Masuhara, H. (2018). The complete guide to the theory and practice of materials development for language learning. Wiley-Blackwell.
[10] Wang, F.-Y., & Lee, S.-Y. (2026). Symbiotic learning: Developing multiliteracies in human–AI collaboration. Pedagogies: An International Journal, 21(2), 329–352. https://doi.org/10.1080/1554480X.2025.2606234

[11] Wang, J. (2025). Large-language-model-supported BOPPPS teaching for maritime ESP: A quasi-experimental study in a Chinese vocational college. In Proceedings of the 2025 3rd International Conference on Information Education and Artificial Intelligence (pp. 1084–1090). Association for Computing Machinery. https://doi.org/10.1145/3799457.3799635