AI-Supported EFL Speaking Practice in Chinese Vocational Education: Practice Behaviors, Skill-Specific Gains, and Moderation by Baseline Proficiency

Hanxu Tu[1], Wan Zhou[2]

[1,2] *Wuhan Vocational College of Software and Engineering (Wuhan Open University), Wuhan, China*
[2]*Corresponding author.*

ABSTRACT

China's English Curriculum Standards for Higher Vocational Education (2021) lists workplace-oriented communication among the core competencies of vocational English education. Large classes and limited contact time, however, make sustained individual speaking practice difficult. This study examined how students' practice behaviors were associated with speaking gains and whether gains differed between pronunciation and communicative effectiveness. It also tested whether baseline speaking proficiency changed the association between total attempts and gain. A semester-long single-group pretest–posttest classroom study involved 99 second-year students in a 16-week General Workplace English course. Students used an AI-supported speaking practice platform, while parallel teacher-rated tasks measured overall speaking, pronunciation, and communicative effectiveness. Platform logs provided total attempts and repeat frequency. Overall speaking scores increased from 62.30 to 70.40, a mean gain of 8.10 points. Both practice indicators were positively associated with overall speaking gain. The association between attempts and gain strengthened as baseline proficiency increased. The simple slope was weak and nonsignificant at one standard deviation below the baseline mean, but positive at the mean and one standard deviation above it. Pronunciation and communicative effectiveness both improved, with pronunciation gaining 1.92 points more on matched 20-point scales. Teachers can read practice logs alongside baseline proficiency to identify participation patterns and plan support. In vocational English courses, platform work can provide frequent rehearsal and pronunciation practice, while classroom interaction can develop extended responses, role-play, and meaning negotiation.

Keywords: *AI-supported speaking, Vocational EFL, Practice behaviors, Speaking gains.*

1. INTRODUCTION

In 2021, the General Office of China's Ministry of Education issued the English Curriculum Standards for Higher Vocational Education (2021). The standards place workplace-oriented communication among the core competencies of vocational English education and require students to use English effectively in everyday and workplace settings [1]. For vocational students, this includes explaining procedures, responding to clients, presenting products, and handling routine workplace exchanges. These abilities need sustained oral practice, yet large classes and limited contact hours leave little time for each student to speak repeatedly and receive individual feedback.

AI-supported speaking tools offer one practical way to extend oral work beyond class time. Students can record, listen, and try again without waiting for a turn in class. Liu et al. [2] found that EFL research on ASR has concentrated on pronunciation accuracy, especially segmental production. This evidence is strongest for focused pronunciation practice. Vocational English, however, also asks learners to sustain longer responses and complete communicative tasks with other speakers.

Tool design shapes the kind of oral work students can do. Structured ASR tasks usually elicit short utterances, whereas chatbots and other interactive systems can support longer exchanges. Ge and Zheng [3] show that AI speaking studies differ in instructional function, classroom integration, and reported outcomes. Studies therefore need to be read in relation to task design, teacher involvement, and assessment choice. These details determine what kind of practice a tool provides and which speaking outcomes it is likely to support.

Another gap concerns what students actually do with the practice they receive. Group comparisons and pretest–posttest gains show outcome differences, but reveal little about variation in students' practice histories. Platform logs can add this process view when their indicators are linked to speaking assessed outside the platform. Starting level may also shape how students use repeated practice, and different speaking skills may change at different rates. For classroom teaching, the useful question is how practice behavior and starting level relate to gains in the skills a course is designed to develop.

The present study addressed this question in a 16-week General Workplace English course at a vocational college in Hubei, China. Ninety-nine second-year students completed parallel teacher-rated speaking tasks before and after the semester, and platform logs provided total attempts and repeat frequency. The analyses examined semester change in overall speaking, the associations between the two practice behaviors and speaking gain, moderation by baseline speaking proficiency, and the difference between pronunciation and communicative-effectiveness gains. Three research questions guided the study:

RQ1. How did students' overall speaking performance change over one semester of AI-supported speaking practice?

RQ2. How were total attempts and repeat frequency associated with overall speaking gains, and did baseline speaking proficiency moderate the association between total attempts and speaking gains?

RQ3. Were pronunciation gains greater than gains in communicative effectiveness?

2. LITERATURE REVIEW

2.1 AI-supported Practice in Speaking Development

Research on AI-supported speaking has developed along two related lines. The larger body of work uses ASR for short, repeatable pronunciation practice. Liu et al. [2] reviewed 24 EFL studies and found that most assessed pronunciation accuracy, particularly segmental production. A newer line of research uses chatbots, mobile applications, and generative systems for longer or more varied speaking tasks. Ge and Zheng [3] found considerable variation in the teaching functions, classroom integration, and outcomes reported for speaking chatbots. The same broad label can therefore refer to quite different forms of oral work.

Ngo et al.'s meta-analysis [4] provides the clearest overall estimate for ASR-based pronunciation practice. Across 15 studies and 38 effect sizes, the mean effect was $g = .69$. Results varied with feedback type, pronunciation target, training length, activity organization, and learner level. These conditions matter because a semester of short, repeatable sound-level work creates a different learning opportunity from open interaction, even when both are described as AI-supported speaking.

Broader gains have appeared when digital practice is combined with interaction or other instructional support. In a study of 61 Chinese EFL learners, Sun [5] found improvements in accentedness, comprehensibility, and global speaking after ASR practice with peer correction. Fathi et al. [6] reported gains across several IELTS speaking dimensions and willingness to communicate in AI-mediated interaction. In both studies, technology was part of a wider teaching arrangement.

Other studies show that gains do not spread evenly across speaking dimensions. Ma et al. [7] reported clearer improvement in pronunciation and fluency than in vocabulary and grammar after a 10-week mobile-app intervention with Chinese university students. During generative-AI speaking activities, Wu et al. [8] found gains in lexical and syntactic complexity and in fluency, with less clear change in accuracy. Task design and measurement shaped what counted as improvement. A single total-speaking score would hide much of that variation.

Research now extends from tightly structured pronunciation drills to a wider range of speaking activities. The evidence remains uneven. Tool functions, task organization, teacher support, and outcome measures vary substantially, and group-level gains say little about how individual students used the practice. Studies of local accuracy offer a clear account of what repeated production can improve. Work on broader speaking outcomes shows the value of interaction and instructional support, but actual practice behavior is often absorbed into the treatment label. Few studies show how individual practice histories within the same class relate to independently assessed gains. For vocational English, the useful question is which practice patterns are associated with which speaking outcomes under ordinary course conditions. This line of inquiry keeps the technology in view while placing the analysis closer to course teaching and student performance.

2.2 Practice Behaviors Recorded in Platform Logs

Outcome scores leave much of the practice process unseen. Platform logs can show how much work accumulated and how often learners returned to it, but a count needs to be interpreted in relation to the activity that produced it. Inceoglu et al. [9] observed several responses to ASR recognition failure. Some learners repaired their speech, some recorded again without a clear repair, and others ignored the result and moved on. The same repetition value can represent different learning actions.

Nickolai [10] reached a related conclusion from eight years of data on one ASR platform. Repeated submissions were uneven across learners and activities, and platform scores sometimes changed across attempts. Such records are useful for tracing work within the platform. Evidence of broader performance requires a new speaking task assessed independently of that platform.

Platform logs become useful process evidence when the recorded behavior is clearly defined. Total activity and repeated activity describe different participation patterns. Their instructional value becomes clearer when they are linked to teacher-rated performance on new prompts. This connection moves the analysis from activity within one scoring system to performance on course-relevant speaking tasks. It also lets teachers examine participation without assuming that every attempt represents the same kind of work. A log can identify students whose patterns deserve attention, while selected recordings and classroom responses help explain those patterns. Few classroom studies have combined semester-level log indicators with independently rated speaking gains. Fewer still have examined whether the association changes with students' starting proficiency. The same amount of recorded practice may have a different value for learners who enter the course with different speaking resources.

2.3 Baseline Proficiency and Skill-Specific Development

Speaking improvement is shaped by what an assessment counts as successful performance. Pronunciation concerns the production of speech sounds, stress, rhythm, and related features. Communicative effectiveness concerns whether a learner conveys relevant meaning and completes the purpose of a task. The two outcomes are related, but classroom practice may provide many opportunities for one and relatively few for the other.

Derwing and Munro [11] distinguish accent from intelligibility and wider communicative success. Their work makes listener understanding the central goal of pronunciation teaching. Levis [12] similarly places pronunciation within the broader goal of oral communication. These distinctions matter in vocational tasks. A student may pronounce individual words clearly and still need to reorganize an explanation, clarify a request, or respond to an unexpected question.

Learners also begin platform practice with different speaking resources. Vocabulary, pronunciation control, task familiarity, and the ability to interpret feedback may affect how quickly an attempt becomes useful rehearsal. Ngo et al. [4] found that learner level contributed to variation in ASR pronunciation effects, although the evidence came from comparisons across studies or groups. That finding leaves a student-level question unresolved: within one classroom, does baseline speaking proficiency change the association between recorded practice and gain?

Vocational English requires intelligible production and successful task completion. Existing studies give reason to expect uneven gains, yet they offer little direct evidence on how starting proficiency changes the value associated with practice behavior. They also rarely compare pronunciation and communicative-effectiveness

gains within the same students on matched scales. If repeated practice concentrates on short individual production, its association with gain may depend on the resources students bring to the task, and the strongest change may appear in the skill most directly rehearsed. A classroom study can examine this pattern through speaking outcomes assessed independently of the platform. It also preserves a direct link between practice behavior, learner starting points, and the speaking demands of vocational coursework. The present study therefore links semester platform logs with teacher-rated parallel speaking tasks and compares two course-relevant skill outcomes within the same students.

3. METHODS

3.1 Research Design and Context

The study used a semester-long single-group pretest–posttest classroom design. It took place from September 1, 2025, to January 9, 2026, at a vocational college in Hubei, China. The course was General Workplace English, a required 64-hour course delivered across 16 weeks, including the examination period. Three intact second-year classes in an intelligent optoelectronics program took part. Two teachers delivered the course, and all three classes used an AI-supported speaking practice platform as part of regular instruction.

Students were the unit of analysis, and the three classes contributed to one combined dataset. The analyses addressed semester change and associations among student-level measures.

3.2 Participants

The sample contained 99 students, all of whom completed the pretest and posttest and had records for the main practice variables. Their ages ranged from 18 to 20 years (M = 19.02, SD = 0.67). Fifty-seven students were male and 42 were female.

The 99 enrolled students agreed to the research use of their de-identified course scores, platform records, and recordings. The school completed its ethics review and teaching-research approval procedures. The present analysis used the teacher-rated speaking measures and platform behavior indicators specified below.

3.3 AI-supported Speaking Practice

Students used an AI-supported speaking practice platform throughout the semester. The assigned work focused on course-linked structured speaking tasks based on unit topics and weekly speaking requirements. After each submission, the platform used automated speech evaluation to provide an overall score and immediate feedback on pronunciation accuracy, fluency, and completeness. Students could review the feedback, listen to their recordings, and attempt the same task again. The platform stored the attempts and exported the two aggregated behavior indicators. Classroom work included teacher-led instruction, presentations, and role-play, so platform practice formed one part of the course.

3.4 Speaking Assessment

Speaking performance was assessed before and after the semester. Each assessment followed the same structure: a short presentation and a situational role-play under comparable time limits, task requirements, and difficulty. The prompts differed between the pretest and posttest, so students responded to new content at each time point.

Each class's pretest and posttest performances were scored by the same course teacher using a shared scoring standard. Overall speaking captured performance across the two tasks and was scored on a 0–100 scale. The pronunciation score represented pronunciation performance across the two tasks. Communicative effectiveness reflected how well the response conveyed meaning and fulfilled the task purpose. Both skill outcomes were scored on a 0–20 scale.

Baseline speaking proficiency was defined as the overall pretest score. Overall speaking gain was the overall posttest score minus the pretest score. Pronunciation gain and communicative-effectiveness gain were calculated in the same way. The skill-gain contrast for RQ3 was pronunciation gain minus communicative-effectiveness gain.

3.5 Platform-recorded Practice Behaviors

Total attempts was the number of speaking attempts recorded for each student across the semester. Values ranged from 19 to 74 (M = 44.98, SD = 11.96). Repeat frequency was the semester-average repetition indicator exported by the platform. It ranged from 1.27 to 3.98 (M = 2.40, SD = 0.60).

Both indicators came from the same semester record, but represented different features of

participation. Keeping them separate allowed the model to examine whether each had a distinct association with speaking gain.

3.6 Procedure

Students completed the parallel speaking pretest at the start of the course. AI-supported speaking practice then continued as part of regular instruction during the semester. The posttest was administered after the instructional period. After the course, platform records were matched to the de-identified assessment data by student identifier.

The final dataset contained the pretest and posttest scores, total attempts, repeat frequency, demographic variables, and calculated gains for all 99 students. No values were missing for the main variables used in RQ1–RQ3.

3.7 Data Analysis

Analyses followed the three research questions. For RQ1, a paired-samples t test compared overall speaking pretest and posttest scores. Reporting included the two means and standard deviations, mean gain, a 95% confidence interval for gain, $t(98)$, the p value, and the paired-design effect size d_z. The effect size was calculated as t divided by the square root of n.

For RQ2, Pearson correlations described the zero-order relations among baseline speaking proficiency, total attempts, repeat frequency, and overall speaking gain. An ordinary least squares model then predicted overall gain from centered total attempts, repeat frequency, centered baseline speaking proficiency, and the product of the two centered variables. The model was:

$$\text{overall speaking gain} = b_0 + b_1(\text{total attempts}_c) + b_2(\text{repeat frequency}) + b_3(\text{baseline}_c) + b_4(\text{total attempts}_c \times \text{baseline}_c) + \text{error}.$$

Unstandardized coefficients, standard errors, standardized coefficients, t values, p values, and 95% confidence intervals were reported. The attempts slope was estimated at one standard deviation below the baseline mean, at the mean, and at one standard deviation above the mean. Figure 1 fixed repeat frequency at its sample mean of 2.3999 and used the observed attempts range of 19–74.

For RQ3, paired-samples tests described pretest–posttest change in pronunciation and communicative effectiveness. The main comparison used each student's two gain scores and tested the mean of pronunciation gain minus communicative-effectiveness gain against zero. Because both outcomes used the same 0–20 scale within one scoring standard, their gain scores were compared directly. The analysis reported the mean and standard deviation of the difference, its 95% confidence interval, $t(98)$, p, and d_z.

4. RESULTS

"Table 1" presents descriptive statistics for the two practice behaviors and the three speaking outcomes. The results then follow RQ1–RQ3.

Table 1. Practice behaviors and pre–post speaking outcomes

Panel A. Platform-recorded practice behaviors

Measure	M	SD	Range
Total attempts	44.98	11.96	19–74
Repeat frequency	2.40	0.60	1.27–3.98

Panel B. Teacher-rated speaking outcomes

Outcome	Pre M (SD)	Post M (SD)	Gain M (SD)	95% CI	t(98)	dz
Overall speaking	62.30 (9.50)	70.40 (9.96)	8.10 (4.49)	[7.20, 9.00]	17.95	1.80
Pronunciation	12.30 (2.50)	14.92 (2.58)	2.62 (1.21)	[2.38, 2.86]	21.51	2.16
Communicative effectiveness	13.10 (2.80)	13.79 (3.27)	0.69 (1.50)	[0.40, 0.99]	4.62	0.46

[a] Note. N = 99. Gain = posttest − pretest. CI = confidence interval. dz = paired-design effect size. All paired tests had p < .001.

4.1 Overall Speaking Change

Overall speaking scores increased from 62.30 (SD = 9.50) at pretest to 70.40 (SD = 9.96) at posttest. The mean gain was 8.10 points (SD = 4.49), 95% CI [7.20, 9.00]. The paired difference was statistically significant, $t(98) = 17.95$, $p < .001$, $d_z = 1.80$.

4.2 Practice Behaviors and Moderation by Baseline Proficiency

"Table 2" gives the means, standard deviations, and zero-order correlations for baseline speaking proficiency, total attempts, repeat frequency, and overall speaking gain.

Table 2. Correlations among baseline proficiency, practice behaviors, and overall speaking gain

Variable	*M*	*SD*	1	2	3	4
1. Baseline speaking proficiency	62.30	9.50	1.000			
2. Total attempts	44.98	11.96	−.123	1.000		
3. Repeat frequency	2.40	0.60	.050	.310**	1.000	
4. Overall speaking gain	8.10	4.49	−.132	.411***	.337***	1.000

a Note. $N = 99$. Gain = posttest − pretest. ** $p < .01$. *** $p < .001$.

Total attempts was positively correlated with overall speaking gain, $r = .411$, $p < .001$. Repeat frequency was also positively correlated with gain, $r = .337$, $p < .001$. The two practice behaviors were moderately related, $r = .310$, $p = .002$. Baseline speaking proficiency was not significantly correlated with attempts, repeat frequency, or overall gain.

The moderation model was significant, $R^2 = .273$, adjusted $R^2 = .242$, $F(4, 94) = 8.84$, $p < .001$. As shown in Table 3, centered total attempts had a positive association with overall gain, $B = .116$, $SE = .035$, $\beta = .310$, $t = 3.31$, $p = .001$, 95% CI [.046, .186]. Repeat frequency also had a positive association, $B = 1.834$, $SE = .696$, $\beta = .245$, $t = 2.64$, $p = .010$, 95% CI [.453, 3.215]. ("Table 3")

Table 3. Moderation model predicting overall speaking gain

Panel A. Model coefficients

Predictor	B	SE	β	t	p	95% CI
Constant	3.814	1.715	n/a	2.224	.029	[0.409, 7.219]
Total attempts (centered)	.116	.035	.310	3.306	.001	[.046, .186]
Repeat frequency	1.834	.696	.245	2.636	.010	[.453, 3.215]
Baseline proficiency (centered)	−.077	.044	−.162	−1.763	.081	[−.163, .010]
Attempts × baseline	.0083	.0034	.221	2.429	.017	[.0015, .0151]
Model summary: $R^2 = .273$; adjusted $R^2 = .242$; $F(4, 94) = 8.84$, $p < .001$.						

Panel B. Conditional effect of total attempts

Baseline level	Score	B	SE	t	p	95% CI
−1 SD	52.8	.037	.049	.753	.453	[−.061, .135]
Mean	62.3	.116	.035	3.306	.001	[.046, .186]
+1 SD	71.8	.195	.047	4.201	< .001	[.103, .288]

a Note. Repeat frequency was included in the model. Conditional effects are reported at the sample mean and ±1 SD of baseline proficiency.

The baseline main effect was negative and did not reach statistical significance, $B = −.077$, $SE = .044$, $\beta = −.162$, $t = −1.76$, $p = .081$, 95% CI [−.163, .010]. The total-attempts × baseline interaction was positive, $B = .00833$, $SE = .00343$, $\beta = .221$, $t = 2.43$, $p = .017$, 95% CI [.00152, .01513].

The interaction indicates that the positive association between attempts and gain increased with baseline proficiency.

Simple slopes clarified this pattern. At one standard deviation below the baseline mean (52.8), the attempts slope was $B = .037$, $SE = .049$, $t = .75$, $p = .453$, 95% CI [−.061, .135]. At the mean baseline score (62.3), the slope was $B = .116$, $SE = .035$, $t = 3.31$, $p = .001$, 95% CI [.046, .186]. At one standard deviation above the mean (71.8), it was $B = .195$, $SE = .047$, $t = 4.20$, $p < .001$, 95% CI [.103, .288]. "Figure 1" displays the model-predicted lines across the observed range of total attempts.

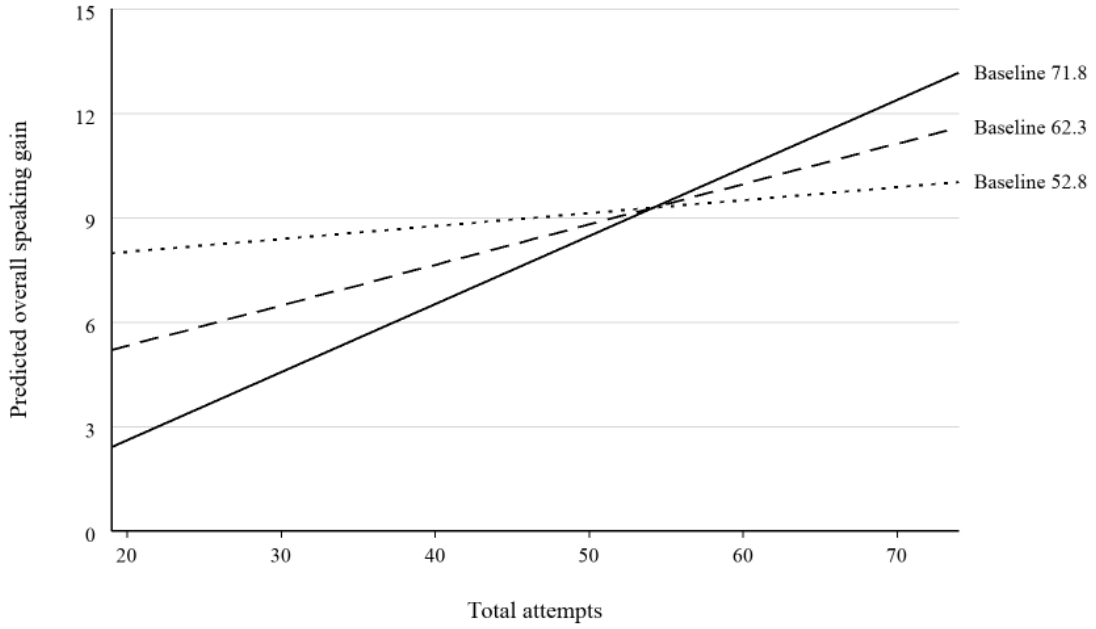


Figure 1 Model-predicted association between total attempts and overall speaking gain at three levels of baseline proficiency.

a Note. Total attempts ranged from 19 to 74; repeat frequency was fixed at 2.3999.

4.3 Skill-specific Gains

Pronunciation increased from 12.30 ($SD = 2.50$) to 14.92 ($SD = 2.58$). The mean gain was 2.62 points ($SD = 1.21$), 95% CI [2.38, 2.86], $t(98) = 21.51$, $p < .001$, $d_z = 2.16$. Communicative effectiveness increased from 13.10 ($SD = 2.80$) to 13.79 ($SD = 3.27$). Its mean gain was .69 points ($SD = 1.50$), 95% CI [.40, .99], $t(98) = 4.62$, $p < .001$, $d_z = .46$.

Pronunciation gain exceeded communicative-effectiveness gain by 1.92 points on average ($SD = 1.04$), 95% CI [1.72, 2.13], $t(98) = 18.42$, $p < .001$, $d_z = 1.85$. Both skills improved, and pronunciation changed considerably more over the semester.

5. DISCUSSION

Both indicators were associated with overall gain, but the attempts–gain relationship became stronger as baseline speaking proficiency increased. Pronunciation also showed the larger gain. These findings suggest that course outcomes reflected the work students completed, the speaking resources they brought to it, and the kind of performance the tasks repeatedly rehearsed.

5.1 Practice Behaviors and Speaking Gains

The increase in teacher-rated overall speaking is consistent with Sun's report of broader speaking gains when ASR practice was paired with peer correction [5]. The present course used a different arrangement, but both settings linked technology-supported rehearsal with classroom instruction. Students practised individually on the platform and then used related language in presentations and role-play during class. This sequence gave platform rehearsal a clear course purpose and follow-up. It also required students to carry language from individual practice into classroom performance.

Students had access to the same platform, yet total attempts and repeat frequency varied

substantially. Both indicators were related to gain after baseline proficiency was included in the model. Two learners could therefore complete the same course with very different practice histories. For a teacher, the log is valuable because it reveals patterns of participation that a pretest and posttest alone leave unseen.

The two indicators captured different aspects of participation, and both remained positively associated with gain in the regression model. Treating them separately gave a fuller account of how students used the practice opportunity. Large platform datasets have also found repeated submissions to be uneven across learners and activities [10]. Here, gain was measured through parallel teacher-rated tasks with new content, linking the log patterns to performance that mattered in the course.

5.2 Baseline Proficiency and the Association between Attempts and Gain

The attempts–gain association was stronger among students with higher baseline speaking proficiency. This pattern suggests that an additional attempt had different instructional value depending on what a student could already do. Before rehearsal can refine delivery, a learner must interpret the prompt, select language, and organize a response. Students who complete these early steps more readily have more room to use later attempts for adjustment and improvement.

Variation by learner level has also appeared in meta-analytic evidence, although those comparisons were made across studies or groups [4]. In this classroom sample, students with higher baseline scores may have reached useful rehearsal sooner because they could begin shaping a response while some classmates were still interpreting the prompt and assembling basic language.

For lower-baseline learners, an attempt may include more task interpretation and language search before useful rehearsal begins. A learner can be active on the platform and still need help with what to say. For these learners, preparation is part of the practice process.

5.3 Skill-specific Gains and the Roles of Platform and Classroom Practice

Pronunciation and communicative effectiveness both improved, but at markedly different rates. The difference fits the practice conditions in this course. Repeatable oral tasks provide frequent opportunities to adjust local speech features, while communicative effectiveness depends on organizing meaning across longer and less predictable exchanges. Fewer of those demands can be rehearsed through short responses completed by one learner at a time.

Pronunciation practice is most useful when it serves intelligibility and oral communication goals [12]. A clearer sound or more stable rhythm can help a listener, yet communicative effectiveness also requires relevant content, responsiveness, and contextual judgment. Those demands become visible in role-play and extended interaction, where learners must adapt what they say as the exchange develops.

Dimension-specific results have also appeared in Chinese university research. A mobile-app study reported clearer gains in pronunciation and fluency than in vocabulary and grammar [7]. The measures differ from those used here, but the comparison reinforces a practical point: improvement is most visible in the skills that practice and assessment repeatedly target.

The two settings offered different forms of practice, and the skill pattern fits that difference. Platform tasks provided frequent individual production and focused rehearsal. Classroom presentations and role-play asked students to organize meaning for less predictable use. The larger pronunciation gain fits the denser opportunities for repeatable, local adjustment, while communicative effectiveness drew on longer, responsive performance.

6. PEDAGOGICAL IMPLICATIONS

6.1 Reading Practice Logs as Teaching Signals

Teachers can use total attempts and repeat frequency to identify practice patterns during a course. For example, few attempts may point to difficulty getting started, while many attempts with little progress can prompt a closer look at task understanding and response quality. Repeat frequency adds another clue about who revisits work. The profile gives the teacher a reason to listen to selected recordings and compare them with classroom performance. That closer look can guide the next conversation or task adjustment.

A workable routine is a brief log review at regular intervals. Teachers can focus on a short list of students with unusually low or high values, then listen to selected recordings and compare them with classroom performance. A brief review after each task cycle is manageable within ordinary teaching routines. It also gives the numbers a useful context for teaching decisions.

6.2 Supporting Learners with Lower Baseline Proficiency

Lower-proficiency learners may need support before additional attempts become useful. Before recording, teachers can clarify the task goal and model one workable response. A short phrase bank or a prompt divided into manageable stages can help students begin. The second attempt can then focus on one revision target. A brief teacher check after that attempt can confirm that students understood the feedback and know what to work on next.

Support can decrease as learners gain control. The aim is to make repeated production accessible and purposeful, then move students toward more independent responses. Baseline scores and early practice logs can help teachers decide who needs this structure. The interaction supports building differentiated preparation into a common assignment.

6.3 Matching Practice Settings to Speaking Goals

Course designers can match each setting to the speaking work it handles well. Platform practice can provide frequent rehearsal, preparation for short presentations, and focused pronunciation work. Classroom time can extend that preparation through role-play, unexpected follow-up questions, and negotiation of meaning. A role-play may begin with familiar course language, but the follow-up question requires students to adapt it. The sequence links individual preparation with communication that develops in response to another speaker.

Assessment should reflect the same division. Pronunciation can be sampled through focused tasks, while communicative effectiveness needs contextualized performance with a clear purpose and an interlocutor. When the goal is workplace communication, platform tasks and classroom scenarios should share topics and language resources. Students then meet the same content under increasingly demanding speaking conditions.

7. LIMITATIONS

The single-group design describes semester change and associations among the measured variables. A comparison condition would be needed to estimate the contribution of AI-supported practice separately from other course components. The sample came from one vocational college and one program. Replication across institutions, majors, and course types would test the stability of the observed patterns.

Platform indicators summarized the semester as a whole. Task-level logs and recordings would allow future research to examine what students change from one attempt to the next and how practice varies across tasks. The assessment covered overall speaking, pronunciation, and communicative effectiveness. Additional measures of fluency, interactional responsiveness, and workplace-task performance could extend the outcome profile.

8. CONCLUSION

This study connected platform-recorded practice behaviors with teacher-rated parallel speaking tasks in a 16-week vocational English course. Overall speaking increased across the semester, and both total attempts and repeat frequency were positively associated with gain. Baseline proficiency strengthened the association between attempts and gain. Pronunciation and communicative effectiveness both improved, with a much larger gain in pronunciation.

For course evaluation, practice logs, baseline scores, and skill-specific assessments serve different purposes. Logs reveal how students participated, while baseline scores help identify where structured support may be useful. Skill-specific assessments then show whether the practice provided by the course matches its speaking goals.

Course design should coordinate platform and classroom work around a shared speaking goal. Repeated individual production can prepare students for classroom tasks that require longer responses, role-play, and meaning negotiation. The sequence moves practice from rehearsal toward responsive use.

ACKNOWLEDGMENTS

Project: 2025 Hubei Provincial Education Science Planning Project: Research on AI-Enabled

Smart Classroom Development in Higher Vocational Education Based on Deep Learning Theory, 2025GB521. (项目：2025 年湖北省教育科学规划一般课题“基于深度学习理论的 AI 赋能高职智慧课堂构建研究”成果，项目编号 2025GB521。)

REFERENCES

[1] General Office of the Ministry of Education of the People’s Republic of China. Notice on issuing the English and Information Technology Curriculum Standards for Higher Vocational Education (2021) [Internet]. 2021. Available from: https://www.moe.gov.cn/srcsite/A07/moe_737/s3876_qt/202104/t20210409_525482.html

[2] Liu Y, Rahman FBA, Zain FBM. A systematic literature review of research on automatic speech recognition in EFL pronunciation. Cogent Education. 2025;12(1):2466288. https://doi.org/10.1080/2331186X.2025.2466288

[3] Ge N, Zheng C. A systematic review of intelligent chatbots supporting English speaking instruction. Foreign Language Education. 2025;46(3):56–63. https://doi.org/10.16362/j.cnki.cn61-1023/h.2025.03.013

[4] Ngo TT-N, Chen HH-J, Lai KK-W. The effectiveness of automatic speech recognition in ESL/EFL pronunciation: A meta-analysis. ReCALL. 2024;36(1):4–21. https://doi.org/10.1017/S0958344023000113

[5] Sun W. The impact of automatic speech recognition technology on second language pronunciation and speaking skills of EFL learners: A mixed methods investigation. Frontiers in Psychology. 2023;14:1210187. https://doi.org/10.3389/fpsyg.2023.1210187

[6] Fathi J, Rahimi M, Derakhshan A. Improving EFL learners’ speaking skills and willingness to communicate via artificial intelligence-mediated interactions. System. 2024;121:103254. https://doi.org/10.1016/j.system.2024.103254

[7] Ma M, Noordin N, Razali AB. Improving EFL speaking performance among undergraduate students with an AI-powered mobile app in after-class assignments: An empirical investigation. Humanities and Social Sciences Communications. 2025;12:370. https://doi.org/10.1057/s41599-025-04688-0

[8] Wu J, Zhou W, Cao C. An empirical study on generative artificial intelligence empowering speaking instruction. China Educational Technology. 2024;(4):105–111.

[9] Inceoglu S, Chen W-H, Lim H. Monitoring student behavior in autonomous automatic speech recognition-based pronunciation practice. System. 2024;124:103387. https://doi.org/10.1016/j.system.2024.103387

[10] Nickolai D. Quantifying the impact of ASR-based instruction: What does the iSpraak platform learner data show? The EuroCALL Review. 2024;31(1):16–23. https://doi.org/10.4995/eurocall.2024.20221

[11] Derwing TM, Munro MJ. Putting accent in its place: Rethinking obstacles to communication. Language Teaching. 2009;42(4):476–490. https://doi.org/10.1017/S026144480800551X

[12] Levis JM. Intelligibility, oral communication, and the teaching of pronunciation. Cambridge: Cambridge University Press; 2018. https://doi.org/10.1017/9781108241564