Vision: Journal for Language and Foreign Language Learning
Vol. 15, No. 2, October 2026, pp. 133-152
ISSN 2745-9667 (p), 2541-4399 (e)
DOI: 10.21580/vjv16i133221

Deep Learning, AI Tools, and Intercultural Competence in Developing EFL Students' Speaking Performance

Titi Rokhayati,[1*] Zulia Chasanah,[2] Tengku Nazatul Shima[3]

[12]University of Muhammadiyah Purworejo, Central Java – Indonesia, [3]Universiti Teknologi MARA, Selangor – Malaysia

Article Information
Received: July 06, 2026
Revised: November 08, 2026
Accepted: November 08, 2026
Published online: November 10, 2026

Abstract

This study examined the effectiveness of a technology-integrated deep learning model combined with intercultural competence in improving university students' English speaking performance by describing their baseline speaking competence, developing and implementing the model, testing its effectiveness, and examining how its intercultural element shaped fluency, accuracy, and appropriateness. A quasi-experimental pretest–posttest control group design was used with 70 students in three intact classes, two experimental (n = 44) and one control (n = 26). The model followed a cycle of speaking practice, AI-mediated feedback, intercultural reflection, and re-performance, with YouTube, ChatGPT, ELSA Speak, and Flip each assigned to a specific phase. Speaking was scored with an analytic rubric covering fluency, accuracy, and appropriateness. At pretest, all classes performed at an intermediate level (M = 69.6–74.6). Both experimental classes made very large gains (d = 2.76 and 2.94), whereas the control class improved less (d = 1.10). The experimental group gained 6.31 points more than the control group (t(68) = 6.20, p < .001, d = 1.53). In the experimental group's posttest scores, appropriateness accounted for the largest share (37.9%), followed by accuracy (31.5%) and fluency (30.7%), reflecting the pragmatic focus of the tasks. These results suggest that pedagogical design matters more than the tools themselves, and future research should use randomised, multi-institutional designs with validated intercultural measures and delayed posttests.

Keywords: artificial intelligence; deep learning; EFL speaking performance; higher education; intercultural competence

Introduction

The rapid advancement of digital technology has fundamentally transformed educational practices, particularly in English language teaching, where technology is expected to enhance learners' communicative competence. In speaking instruction, digital tools such as artificial intelligence (AI), automatic speech recognition (ASR), and virtual communication platforms have been shown to provide authentic input, real-time feedback, and increased opportunities for interaction (Sun, 2023; Syair & Mujahid, 2025). Despite these affordances,

[*]**Corresponding Author:** Titi Rokhayati (trokhayati@umpwr.ac.id) Jl. K.H.A. Dahlan No. 3 & 6, Purworejo, Purworejo District, Purworejo Regency, Central Java, 54111

©2026 by the Authors, published by Vision: Journal for Language and Foreign Language Learning
https://journal.walisongo.ac.id/index.php/vision

students' speaking proficiency often remains at a moderate or low level, particularly in fluency, accuracy, and appropriateness. Technology integration alone is evidently insufficient without a pedagogical framework that promotes meaningful learning, critical reflection, and communicative engagement.

Research on this issue is complicated by the varied use of the term deep learning. In the computational sense, deep learning refers to neural-network systems that analyse learner speech and adapt instruction (Chai, 2026; Qu & Li, 2022). In the pedagogical sense, it refers to learning through active engagement, reflection, and contextual application (Haerazi & Irawan, 2026; Ramli & Kusmaryani, 2026). Several studies leave the intended meaning unstated (Hao, 2025; Zhong, 2024). For conceptual clarity, this study uses deep learning in its pedagogical sense: an instructional approach in which learners engage actively with meaningful speaking tasks, reflect on their performance, and apply what they learn in authentic communicative contexts. Technology integration refers to the purposeful use of digital tools, including AI-based feedback, within that instructional cycle rather than as an add-on to it. This view aligns with arguments that technology functions not merely as a supporting tool but as an integral component of instructional strategies that promote deep learning, critical thinking, and global communication competence (Popa et al., 2020; Purwanto et al., 2025).

Studies in the computational strand show what technology can contribute to speaking development. Multi-task neural-network frameworks can assess pronunciation, fluency, vocabulary-grammar complexity, and content coherence with correlations of 0.793 to 0.887 against expert ratings. Reinforcement-learning-based practice sequencing produced learning gains about 2.2 times higher than fixed curricula over eight weeks (Chai, 2026). Clustering techniques have been used to group learners by proficiency for differentiated instruction (Hao, 2025). Wearable feedback systems have reported substantial gains in fluency and pronunciation accuracy (Zheng & Zhong, 2026). Literature-based analyses confirm that such tools can improve pronunciation, fluency, and learner motivation, but their effectiveness depends on how they are integrated into instructional design (Abdelhalim & Alsehibany, 2025; Shafiee Rad & Roohani, 2026). Students also report that AI-based tools support adaptive learning and contribute to the development of digital literacy and speaking skills (Yustinah & Hartono, 2026). Much of this work, however, treats speaking chiefly as measurable performance. It says little about the teacher's role or how learners use feedback, and it reduces cultural appropriateness to a single index or leaves it out altogether. Technological sophistication, in other words, does not automatically lead to pedagogical effectiveness.

Speaking competence also has sociocultural and intercultural dimensions. English, as a global lingua franca, requires learners to negotiate meaning across diverse cultural contexts, which makes intercultural competence an essential component of effective communication (Dewi & Slamet, 2025). In this study, intercultural competence refers to the knowledge, skills, and attitudes that allow learners to interpret cultural differences and adapt their language appropriately when interacting with people from other cultural backgrounds (Zhong, 2024).

Existing studies frequently treat speaking as linguistic performance and overlook intercultural processes such as pragmatic adaptation, identity negotiation, and cultural interpretation, so speaking competence in EFL contexts is often understood in a fragmented manner. Pedagogical studies that address this dimension directly report encouraging results. Explicit instruction in politeness strategies, speech acts, turn-taking, and cultural norms through role-plays and problem-based tasks helped learners speak more appropriately in intercultural situations (Haerazi & Irawan, 2026). Courses built around noticing, comparing, and reflecting developed students' intercultural attitudes, knowledge, skills, and awareness (Mu & Yu, 2023; Yu & Wong, 2026). Informal digital learning of English outside class has also been associated with cultural knowledge and openness (Liu et al., 2023a). Authentic speaking tasks encourage meaningful interaction and prepare learners to communicate appropriately in multicultural settings (Jain, 2024). Promoting intercultural competence, however, calls for a multi-modal strategy involving curriculum development, teacher preparation, community involvement, and technological innovation (Eden et al., 2024). In much of the pedagogical work, technology plays only an incidental role. The studies also measure intercultural competence in different ways, often through self-report or teacher observation, which makes their findings difficult to compare.

The most consistent finding across both strands is that authentic interaction with culturally different interlocutors advances speaking and intercultural development together. Integrating technology with intercultural communication pedagogy can improve not only linguistic proficiency but also learners' confidence, willingness to communicate, and ability to negotiate meaning during intercultural interactions (Shonfeld, 2024). Generative AI and immersive environments can simulate intercultural contexts, reduce speaking anxiety, increase engagement, and expose learners to diverse cultural perspectives (Aisyah, 2025; Azahra et al., 2025; Huang et al., 2026; Qu & Chen, 2026). Comparable effects have been reported with simpler tools, including movie-based mobile learning (Chaya & Inpin, 2020), 360-degree video exchanges (Shadiev et al., 2021), and telecollaboration with native and non-native speakers (Liu et al., 2023b). Video-based exchanges through Zoom and Flip reduced speaking anxiety and raised self-efficacy and willingness to communicate (Çatan & Kayaoğlu, 2026). Together, these findings suggest that interaction design matters more than technological sophistication, although few studies isolate which component produces the improvement.

The effectiveness of technology also depends on how teachers mediate its use. Innovative instructional methods supported by information technology enhance speaking proficiency (Bakhtiyarovna et al., 2024). Speaking tools and strategy training produce larger gains than conventional instruction, especially when learners apply metacognitive strategies (Asratie et al., 2023; Pinandhita, 2025). Blended designs can move students from exam-oriented memorisation towards active oral problem-solving (Qu & Li, 2022). Technology does not by itself produce deep learning, however. Its value depends on teacher guidance and on how closely tools are tied to speaking tasks,

reflection, and communicative purposes. Some students relied too quickly on AI for ready-made answers, a pattern that shows the need for teacher mediation, digital literacy, and ethical awareness (Ramli & Kusmaryani, 2026). Teachers therefore need both technological knowledge and pedagogical strategies to design effective instruction (Robillos, 2023; Smith & Ziegler, 2023).

Several gaps follow from this body of research. First, technology integration, deep learning, and intercultural competence have mostly been examined separately rather than as an interconnected pedagogical system. Diagnostic systems measure speaking precisely but neglect reflection and culture, while reflective intercultural pedagogy uses technology only incidentally. Second, speaking and intercultural competence are usually assessed separately and with instruments of uneven validity, so the claim that the two develop together remains weakly supported. Third, most of the evidence comes from short interventions in a limited number of national settings, mainly China and other parts of Asia, and researchers have called for validation across broader populations and longer periods (Chai, 2026). Empirical models that integrate the three dimensions to develop English speaking competence holistically therefore remain limited, particularly in EFL contexts such as Indonesia, where learners face linguistic, cultural, and technological challenges at the same time. An instructional model is needed that combines technology-supported deep learning with intercultural competence to foster meaningful, reflective, and culturally appropriate communication.

This study therefore aims to describe students' English speaking competence before the implementation of a deep learning–based instructional model, and then to develop and implement a technology-integrated deep learning model combined with intercultural competence. It further analyses the effectiveness of this model in improving students' speaking performance in terms of fluency, accuracy, and appropriateness, and explains how the integration of intercultural competence within the deep learning model shapes students' speaking development. Through this approach, the study is expected to contribute theoretically and practically to the development of innovative, context-sensitive, and globally relevant English language pedagogy in the digital era.

Method

Research Design

This study used a convergent mixed-methods design within a quasi-experimental pretest–posttest nonequivalent control group design. Mixed-methods research combines the collection, analysis, and interpretation of quantitative and qualitative data to give a fuller picture of a research problem (Sardana et al., 2023). The quantitative strand had priority, as it tested whether the technology-integrated deep learning model improved students' speaking performance. The qualitative strand explained how the model worked, especially in relation to students' intercultural engagement. Both types of data were collected during the same period, analysed separately, and then merged at the interpretation stage. The study took place from December 2025 to May 2026 in an English Department in Purworejo, Indonesia.

Participants

The participants were 70 English Department students in three intact classes: Class A (22 students), Class B (22 students), and Class C (26 students). Classes A and B formed the experimental groups, and Class C served as the control group. The classes were chosen through purposive sampling because they were taking the same speaking course and had similar English proficiency. Because the classes were not randomly assigned, selection bias is possible, and the results are not treated as strict causal evidence. Pretest scores were therefore used to check that the groups were comparable at the start and to adjust the posttest comparison. A sensitivity power analysis ($\alpha = .05$, power = .80) showed that the sample could detect a medium-to-large effect ($d \approx 0.70$).

Instruments

Three instruments were used. The first was a speaking test given as a pretest and a posttest. Students' performances were recorded and scored with an analytic rubric covering fluency, accuracy, and appropriateness. Each criterion was rated on a 1–5 scale and given equal weight. Fluency covered speech rate, pausing, and continuity. Accuracy covered grammar, vocabulary, and pronunciation. Appropriateness covered politeness, register, and the adaptation of language to the interlocutor's culture. The second instrument was a questionnaire on students' experiences of deep learning and their understanding of intercultural communication. The third was an observation sheet used to record classroom activities, student engagement, and the implementation of the deep learning cycle.

All instruments were reviewed by experts for content validity. Before scoring, two raters took part in a calibration session using sample recordings. They then scored all recordings independently, without knowing each student's group. Inter-rater reliability was estimated with the intraclass correlation coefficient (ICC), and the internal consistency of the questionnaire was estimated with Cronbach's alpha.

Data Collection

Data were collected in three stages. First, all classes took the speaking pretest and completed the questionnaire. Second, the experimental groups received the technology-integrated deep learning model, while the control group received conventional instruction on the same topics and tasks.

In the experimental groups, each session followed a four-phase cycle. In the noticing phase, students watched short YouTube videos of intercultural interaction and identified expressions, politeness strategies, and cultural norms. In the practising phase, they used ChatGPT as a role-play partner in intercultural situations and asked it for feedback on their language and its cultural appropriateness. In the refining phase, they used ELSA Speak to practise pronunciation and fluency with automated feedback. In the performing and reflecting phase, they recorded speaking tasks on Flip, commented on their peers' videos, and reflected on the cultural appropriateness of their language. The teacher acted as facilitator: guiding each phase, monitoring students' use of AI so that they did not rely on ready-made answers, leading reflection, and giving feedback.

The control group learned through teacher explanation, textbook dialogues, and pair

practice, without the deep learning cycle or the four technologies. Classroom observations were used to check that the model was implemented as planned. Finally, all classes took the speaking posttest and completed the questionnaire again.

Data Analysis

Quantitative data were analysed in SPSS. Means and standard deviations were calculated for each group and each speaking dimension. Before the main tests, normality was checked with the Shapiro–Wilk test, homogeneity of variance with Levene's test, and outliers with boxplots. Within-group gains were tested with paired-samples t-tests. Differences between groups were tested with ANCOVA, using pretest scores as the covariate. Where the assumptions were not met, the Wilcoxon and Mann–Whitney tests were used instead. Results are reported with exact p-values, 95% confidence intervals, and effect sizes.

Qualitative data from observations, open-ended questionnaire responses, and student reflections were analysed with thematic analysis following Clarke and Braun's (2017) framework, which involves familiarisation, coding, theme development, and interpretation. The quantitative and qualitative findings were then compared for each speaking dimension to see how students' experiences explained or confirmed the changes in their speaking scores.

Findings and Discussion

Finding

The findings are presented in line with the four research objectives. They cover students' speaking competence before the intervention, the development and implementation of the model, its effectiveness in improving speaking performance, and the influence of intercultural competence on fluency, accuracy, and appropriateness. The quantitative results are summarised in Table 1.

Table 1

Students' Speaking Scores Before and After the Intervention

Group	n	Pretest *M*	Posttest *M*	Gain	SD of gain	95% CI of gain	*t*	*df*	*p*	*d*
Class A (experimental)	22	69.6	82.3	12.68	4.61	[10.64, 14.72]	12.96	21	< .001	2.76
Class B (experimental)	22	74.6	85.5	10.86	3.69	[9.22, 12.50]	13.80	21	< .001	2.94
Class C (control)	26	73.3	78.7	5.48	4.94	[3.48, 7.48]	5.63	25	< .001	1.10
Experimental vs. control (gain scores)	44/26	-	-	6.31	-	[4.28, 8.34]	6.20	68	< .001	1.53

Note. Within-group results are from paired-samples t-tests. The between-group result is from an independent-samples t-test on gain scores (Levene's test, p = .214).

Students' English Speaking Competence Before the Implementation of the Deep Learning–Based Instructional Model

The pretest results show that the three classes began at a similar intermediate level, with means between 69.6 and 74.6 on a 0–100 scale (Table 1). The gap between the highest and lowest class was 5.0 points, which means the classes did not start from exactly the same point. Class B had the highest mean (74.6), followed closely by Class C (73.3), with a difference of only 1.3 points between them. Class A had the lowest mean (69.6), 3.7 points below the control class and 5.0 points below Class B. When the two experimental classes are combined, their pretest mean is 72.1, which is 1.2 points below the control class. At the group level, then, the experimental and control groups began at nearly the same level. Within the experimental group, however, Class A started noticeably lower than Class B.

These differences at baseline matter for how the later results are read. A posttest-only comparison could understate the progress of a class that started lower, or overstate the progress of a class that started higher. For this reason, the effectiveness of the model was examined through gain scores, which measure how much each class improved from its own starting point rather than where it finished.

The Development and Implementation of the Technology-Integrated Deep Learning Model Combined with Intercultural Competence

The instructional model was developed by integrating speaking skill, deep learning, and intercultural competence (Bhuana et al., 2025; Putri et al., 2025; Ramadhani et al., 2025). Each component had a defined function within the model. Deep learning was the pedagogical basis, directing learning towards meaningful, reflective, mindful, and contextual activities rather than the memorisation of dialogues. AI-based tools gave automatic feedback on pronunciation, fluency, grammar, and expression. Intercultural competence directed students' attention to the cultural context of communication, so that speaking was judged not only by linguistic correctness but also by its fit with the situation and the interlocutor. Speaking skill was the performative outcome the model aimed to develop, and it was viewed as the product of cognitive, affective, social, and cultural processes.

In class, the model was carried out through four topics. Each topic combined a speaking activity, a technology, and a specific intercultural focus. For making an appointment, students carried out role-plays with ChatGPT and practised pronunciation with ELSA Speak, with the focus on the difference between direct and indirect speech. For giving opinions, students recorded videos on Flip and received AI feedback, with the focus on politeness strategies. For apologising, students took part in cross-cultural simulations with ChatGPT that addressed face-saving across cultures. For small talk, students analysed YouTube videos to examine how cultures differ in the topics they regard as suitable for casual conversation.

Across the four topics, each technology was matched to a particular phase of the learning process. YouTube provided authentic input for observation. ChatGPT provided a partner for interactive practice and simulation. ELSA Speak provided feedback on pronunciation. Flip provided a space for recording performances

and receiving feedback. The intercultural aspects addressed were all pragmatic in nature, namely directness, politeness, face, and topic choice. These are the dimensions of speaking that the appropriateness criterion was designed to capture.

The Effectiveness of the Model in Improving Students' Speaking Performance

All three classes scored higher on the posttest than on the pretest, but the size of the improvement differed clearly between the experimental and control conditions (Table 1). Class A rose from 69.6 to 82.3, a gain of 12.68 points or 18.2% of its initial mean. Class B rose from 74.6 to 85.5, a gain of 10.86 points or 14.6%. Class C, taught with conventional instruction, rose from 73.3 to 78.7, a gain of 5.48 points or 7.5%. In absolute terms, the gains in Classes A and B were about 2.3 and 2.0 times larger than the gain in Class C.

The paired-samples t-tests showed that the improvement in each class was statistically significant ($p < .001$), so the pretest–posttest difference within each class was unlikely to be due to chance. The effect sizes show that the magnitude of change was very different across conditions. Class A ($d = 2.76$) and Class B ($d = 2.94$) showed very large effects, while Class C showed a large but considerably smaller effect ($d = 1.10$). The confidence intervals point the same way. The upper limit of the control class's interval (7.48) is below the lower limits of both experimental classes (10.64 and 9.22). The intervals do not overlap, which indicates that the experimental gains were consistently higher than the control gain, not only higher on average. The variability of the gains also differed. Class B had the smallest standard deviation of gains (3.69), so its students improved by relatively similar amounts. Class A (4.61) and Class C (4.94) showed more variation between individual students. Class B therefore combined the highest posttest mean with the most uniform progress.

The independent-samples t-test compared the gain scores of the combined experimental group with those of the control group. Levene's test showed that the two groups had equal variances ($p = .214$), so the standard form of the test was appropriate. On average, the experimental group gained 6.31 points more than the control group ($t(68) = 6.20$, $p < .001$, 95% CI [4.28, 8.34], $d = 1.53$). The entire confidence interval lies above zero, and even its lower limit is 4.28 points. The effect size of 1.53 shows that the difference between the groups was large in practical terms as well as statistically significant.

The results also show a change in the relative positions of the classes. Before the intervention, Class A was 3.7 points below the control class. After the intervention, it was 3.6 points above it, a shift of 7.3 points. Class B widened its lead over the control class from 1.3 points to 6.8 points. Combined, the experimental group moved from 1.2 points below the control group on the pretest to 5.2 points above it on the posttest. The class that started lowest made the largest gain, so the difference between the two experimental classes narrowed from 5.0 points to 3.2 points. Because the classes were intact groups and not randomly assigned, these results establish a clear association between the instructional model and greater improvement in speaking performance. They do not establish definitive causal proof.

The Influence of Integrating Intercultural Competence in the Deep Learning Model on Students' Speaking Fluency, Accuracy, and Appropriateness

The performance of the experimental group (n = 44) was broken down into its three speaking components to examine where the improvement was concentrated. The group's mean speaking score was 80. Appropriateness accounted for 37.9% of the total score, accuracy for 31.5%, and fluency for 30.7%. The three components carried equal weight in the rubric, so if students had performed equally on all three, each would have accounted for about 33.3% of the total. Appropriateness exceeded this level by 4.6 percentage points, while accuracy and fluency fell 1.8 and 2.6 percentage points below it. In other words, students' scores were relatively stronger in appropriateness than in the two linguistic components.

This distribution matches the content of the intervention. Every activity in the model had an explicit pragmatic and intercultural focus: directness, politeness strategies, face-saving, and topic choice. The appropriateness criterion assessed exactly these aspects of speaking. The component with the highest relative score is therefore the one most directly addressed by the intercultural element of the model. Accuracy and fluency accounted for similar shares (a difference of only 0.8 percentage points), which shows that students' linguistic performance was balanced across these two components.

Some limits apply to how far these component results can be interpreted. The component percentages are reported only for the experimental group and only after the intervention. As a result, they show the relative strength of the three components at the end of the programme. They do not show how much each component changed from the pretest, or how the experimental group compared with the control group on each component. Component-level pretest data and control-group data would be needed to establish how much of the advantage in appropriateness can be attributed specifically to the intercultural element of the model.

Discussion

Students' Speaking Competence Before the Intervention

The pretest results placed all three classes at an intermediate level, with means between 69.6 and 74.6. This means students began with a working command of spoken English but had not yet reached the level expected for communicative performance in higher education. The pattern is consistent with Astuti and Charneidy (2025), who argue that EFL learners often show declarative knowledge of language forms but find it difficult to transfer that knowledge into authentic communicative performance. It also matches recent studies reporting that many EFL learners continue to struggle with communicative performance because classroom instruction tends to prioritise linguistic accuracy over meaningful interaction and experiential learning (Cheng et al., 2024; Sayed et al., 2024). As English functions as a global lingua franca, learners are expected to negotiate meaning across diverse cultural contexts (Dewi & Slamet, 2025). An intermediate score on a test that includes appropriateness therefore points to room for growth beyond grammatical control alone.

The baseline also differs from much of the earlier literature in one respect. Studies of deep-

learning diagnostic systems mostly describe initial proficiency through computational profiling, such as stratifying learners into proficiency tiers for differentiated instruction (Hao, 2025) or analysing pronunciation, fluency, complexity, and coherence automatically (Chai, 2026). The present study measured baseline performance with a rater-scored analytic rubric that included appropriateness alongside fluency and accuracy. This gave the starting point a pragmatic dimension that diagnostic studies seldom include. The 5.0-point gap between Class A and Class B also shows that intact classes enrolled in the same course may still differ at the outset, a point that needs attention when the effects of interventions are compared across groups.

The Development and Implementation of the Model

The model combined deep learning as pedagogy, AI-based tools, and intercultural competence within one instructional cycle of speaking practice, AI-mediated feedback, intercultural reflection, and re-performance. This cycle is consistent with constructivist learning theory, which treats learning as an active process of building meaning through interaction, reflection, and revision (Sholehhudin & Puspitasari, 2026). Its design resembles the reflective intercultural pedagogy reported in earlier studies. Haerazi and Irawan (2026) integrated politeness strategies, speech acts, turn-taking, and cultural norms into role-plays and problem-based tasks. Mu and Yu (2023) built an intercultural speaking course around noticing, comparing, and reflecting. In the present model, the activity topics drew on the same pragmatic features: directness in making appointments, politeness and hedging when giving opinions, face-saving when apologising, and cultural differences in topic choice during small talk.

The model differs from earlier work in where technology sits within the pedagogy. In diagnostic studies, technology is the core of the intervention, and the main outcomes are the system's agreement with expert ratings and the efficiency of adaptive practice sequencing (Chai, 2026; Zheng & Zhong, 2026). In much reflective intercultural pedagogy, by contrast, technology plays a minor role (Mu & Yu, 2023; Yu & Wong, 2026). The present model assigned each tool to a specific phase of the cycle: YouTube for authentic input, ChatGPT for simulated interaction and feedback, ELSA Speak for pronunciation practice, and Flip for recorded performance and peer feedback. This arrangement follows the position that the effectiveness of AI tools depends on how they are integrated into instructional design (Abdelhalim & Alsehibany, 2025; Shafiee Rad & Roohani, 2026). It also follows the argument that technology does not produce deep learning by itself and needs teacher guidance that links tools to speaking tasks, reflection, and communicative purpose (Ramli & Kusmaryani, 2026).

The model's intercultural focus is in line with studies reporting that AI-assisted learning environments support intercultural communication competence, learning adaptability, and communicative flexibility (Cheng et al., 2024), and that technology-assisted intercultural tasks encourage pragmatic awareness and contextual sensitivity (Cao, 2025; Li et al., 2026). Unlike immersive and telecollaborative designs, however, the present model did not involve real interaction with interlocutors from other cultures. Huang et al.

(2026) used AI-enhanced immersive environments, and Liu et al. (2023b) used telecollaboration with native and non-native speakers. In the present study, intercultural contact was simulated through ChatGPT and represented through YouTube videos. This difference matters, because earlier studies identify authentic interaction with culturally different interlocutors as the most consistently reported driver of both speaking and intercultural development.

The Effectiveness of the Model in Improving Speaking Performance

Both experimental classes made significantly larger gains than the control class. The between-group difference in gain scores was 6.31 points (d = 1.53), and the confidence intervals of the experimental and control gains did not overlap. This result supports earlier findings that technology-supported speaking instruction produces greater improvement than conventional teaching. Asratie et al. (2023) reported that speaking technology tools improved fluency, coherence, lexical range, and pronunciation more than conventional instruction. Pinandhita (2025) found large effects (d > 0.8) when digital tools were combined with strategy training among Indonesian undergraduates. The effect in the present study is of a similar order to the reinforcement-learning-based training reported by Chai (2026), where d = 1.34 over eight weeks. Bakhtiyarovna et al. (2024) likewise argue that innovative instructional methods supported by information technology enhance the development of speaking proficiency. Because the present study and Pinandhita (2025) were both conducted in Indonesian higher education, the agreement between them extends evidence that has so far come mainly from Chinese settings.

The results can be explained through the principles of deep learning pedagogy, namely meaningful engagement, critical reflection, iterative revision, and contextualised learning (Li et al., 2026). Speaking is increasingly understood as a socially situated act shaped by cultural norms and pragmatic considerations. Learners are therefore expected to produce language and also to evaluate, reflect on, and rebuild their performance through continued interaction and self-monitoring (Singh, 2025; Zhang, 2024). The re-performance stage put this into practice, asking students to revise their utterances with attention to pronunciation, fluency, grammar, politeness, indirectness, and cultural appropriateness. This mechanism is in line with findings that automated feedback and scaffolded interaction help learners develop metacognitive awareness and communicative confidence through continuous practice and self-evaluation (Qu & Chen, 2026). It is also in line with studies showing that AI-assisted oral assessment supports learner autonomy and speaking performance (Sayed et al., 2024), and that large language model–supported speaking systems contribute to fluency, engagement, and sustained interaction (Xia et al., 2024).

The present study differs from some earlier work in which outcomes it measured. Several studies report affective gains alongside speaking improvement. Exchanges based on Zoom and Flip reduced speaking anxiety and increased self-efficacy and willingness to communicate (Çatan & Kayaoğlu, 2026). Generative AI simulations reduced anxiety and increased engagement (Azahra et al., 2025; Qu & Chen, 2026). Kurnia et al. (2026) explained that intercultural sensitivity

was the strongest predictor of willingness to communicate (WTC) among Indonesian EFL learners. Perceived AI usefulness had a smaller but still significant effect, and intercultural sensitivity did not mediate between AI usefulness and WTC. The authors add that this null mediation result reflects the study's cross-sectional design and measurement limits, so it shouldn't be read as a general pattern. The present study did not measure anxiety, confidence, or willingness to communicate, so it cannot confirm whether the speaking gains came with similar affective changes. The control class also improved significantly (d = 1.10), which shows that conventional instruction produced measurable progress over the same period. The contribution of the model therefore lies in the additional gain beyond this level rather than in improvement as such. Finally, Class A started lowest and made the largest gain, which narrowed its gap with Class B from 5.0 to 3.2 points. Earlier studies have tended to report group means without examining whether interventions reduce differences between learners with different starting levels, so this pattern adds a dimension to the existing evidence.

The Influence of Intercultural Competence on Fluency, Accuracy, and Appropriateness

Appropriateness accounted for the largest share of the experimental group's total score (37.9%), followed by accuracy (31.5%) and fluency (30.7%). With equal weighting, each component would account for about 33.3%, so appropriateness was the only component above that level. This pattern agrees most closely with Haerazi and Irawan (2026), who found that explicit intercultural pragmatic instruction helped learners speak more appropriately and effectively in intercultural situations. It also agrees with Yu and Wong (2026), who reported that embedding intercultural communicative competence into oral English instruction produced greater gains than regular instruction. In both cases, as in the present study, the strongest outcomes appeared in the dimension most directly addressed by the intercultural content of the instruction.

The pattern differs from studies that relied mainly on computational deep learning. Zheng and Zhong (2026) reported the largest gains in pronunciation accuracy (78%) and fluency (72%), with cultural sensitivity measured as a separate index (0.71). Chai (2026) evaluated pronunciation, fluency, complexity, and coherence but did not include appropriateness among the scored dimensions. Where technology is designed mainly to diagnose and correct linguistic form, improvement tends to concentrate in linguistic components. Where intercultural pragmatics is built into the tasks, as in the present study, the relative strength shifts towards appropriateness. This contrast supports the argument in earlier work that technological sophistication does not automatically lead to pedagogical effectiveness. It also shows that the outcomes of technology-integrated instruction depend on the content the technology is used to deliver.

Accuracy and fluency accounted for similar shares, a difference of only 0.8 percentage points, which shows that the emphasis on appropriateness did not come at the cost of linguistic performance. This result is in line with studies reporting that technology-mediated intercultural communication environments improve communicative flexibility and

contextual interactional competence (Liu et al., 2025; Xia et al., 2024), and that AI-based tools support adaptive learning and the development of speaking skills (Yustinah & Hartono, 2026). It also matches the findings of Gui-Hua and Yang (2025), who reported that AI-driven tools improved cultural knowledge, attitudinal openness, and communication skills together.

One difference in measurement should be noted. Earlier studies measured intercultural competence as a separate construct. Mu and Yu (2023) examined intercultural attitudes, knowledge, skills, and awareness. Zhong (2024) treated intercultural competence as a composite of knowledge, skill, and attitude. Liu et al. (2023a) modelled cultural knowledge and openness through a large-scale survey. The present study examined the influence of intercultural competence through the appropriateness component of a speaking rubric. This approach links intercultural competence directly to observable speaking performance. However, it does not measure intercultural competence itself, and the component percentages are available only for the experimental group after the intervention. The results therefore show where the experimental group's performance was relatively strongest, but not how much each component changed or how it compared with the control group.

Implications and Future Research

The findings support the view that speaking competence is multidimensional, involving linguistic, pragmatic, sociocultural, and reflective dimensions at once, and that the outcomes of technology-integrated instruction depend more on pedagogical design than on the tools themselves. In practice, EFL speaking courses can combine AI-simulated role-plays on culturally sensitive topics, analysis of authentic video interaction, pronunciation practice with automated feedback, and recorded re-performance with peer feedback to develop linguistic accuracy and contextual appropriateness together. Because students may rely on AI too quickly for ready-made answers (Ramli & Kusmaryani, 2026), teachers need to guide and monitor AI use actively, and teacher preparation should develop both technological knowledge and pedagogical strategies (Robillos, 2023; Smith & Ziegler, 2023). The study was limited to 70 students in intact classes at one institution and to posttest component data from the experimental group, and it examined intercultural competence only through speaking appropriateness. Future research could therefore use random assignment, larger samples, and several institutions. It could also collect pre- and post-intervention component scores for both groups, administer validated intercultural competence instruments, and measure affective outcomes such as anxiety and willingness to communicate. Other useful directions include comparing AI-simulated interaction with authentic telecollaboration and using delayed posttests with integrated qualitative data to examine whether the gains last over time.

Conclusion

Before the intervention, students' English speaking competence was at an intermediate level, with pretest means between 69.6 and 74.6. They had a working command of spoken English but still had room to develop fluency, accuracy, and especially appropriateness. The instructional model was carried out as a cycle of

noticing, practising, refining, and performing and reflecting. YouTube provided authentic input, ChatGPT acted as a partner for simulated intercultural interaction and feedback, ELSA Speak supported pronunciation practice, and Flip was used to record performances, give peer feedback, and reflect. Each activity was linked to a pragmatic and intercultural focus: directness, politeness, face-saving, and topic choice.

Both experimental classes improved considerably more than the control class. On average, the experimental group gained 6.31 points more than the control group, a large effect (d = 1.53). The class that started lowest made the largest gain, which narrowed the gap between the classes. Within the experimental group, appropriateness accounted for the largest share of the total speaking score (37.9%), followed by accuracy (31.5%) and fluency (30.7%). Because each component carried equal weight, this pattern shows that students' performance was relatively strongest in appropriateness, the component most directly addressed by the intercultural focus of the instruction. Fluency and accuracy accounted for similar shares, which shows that the emphasis on appropriateness did not come at the cost of linguistic performance.

These findings support the view that speaking competence is multidimensional, involving linguistic, pragmatic, sociocultural, and reflective dimensions at once. They also suggest that the outcomes of technology-integrated instruction depend on pedagogical design and task content more than on the technology itself. In practice, EFL speaking courses can combine AI-simulated role-plays on culturally sensitive topics, analysis of authentic video interaction, pronunciation practice with automated feedback, and recorded re-performance with peer feedback, so that linguistic accuracy and contextual appropriateness develop together. For this to work, teachers need to keep an active role in guiding and monitoring students' use of AI. Teacher preparation programmes should therefore develop both technological knowledge and the pedagogical strategies needed to build intercultural content into technology-supported speaking instruction.

Conflict of Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Generative AI Statement

The authors used Grammarly solely for language editing and proofreading to improve the clarity and readability of the manuscript. All AI-assisted output was critically reviewed and revised by the authors, who take full responsibility for the accuracy, integrity, and final content of the manuscript.

Author Contributions

TR: Conceptualization, Methodology, Investigation, Formal analysis, Data curation, Writing – original draft, Writing – review & editing, Supervision, Project administration. ZC: Resources, Writing – original draft, Writing – review & editing. TNS: Writing – review & editing.

References

Abdelhalim, S. M., & Alsehibany, R. A. (2025). Integrating AI-powered tools in EFL

pronunciation instruction: Effects on accuracy and L2 motivation. Computer Assisted Language Learning. Advance online publication. https://doi.org/10.1080/09588221.2025.2534015

Aisyah, S. (2025). The role of immersive technology and AI-based adaptive learning in enhancing intercultural competence of EFL students. International Journal of Innovation and Thinking, 2(8). https://doi.org/10.71364/ijit.v2i8.73

Asratie, M., Wale, B. D., & Aylet, Y. T. (2023). Effects of using educational technology tools to enhance EFL students' speaking performance. Education and Information Technologies, 1–21. https://doi.org/10.1007/s10639-022-11562-y

Astuti, W., & Charneidy, M. A. (2025). Digital Intensive English untuk meningkatkan kompetensi speaking mahasiswa PPL–KKN internasional. Al-Umm: Jurnal Pengabdian dan Pengembangan Masyarakat, 3(1), 25–34. https://doi.org/10.53398/alumm.v3i1.586

Azahra, D. N., Nabhan, M. G. A., Ismail, H. R. N., Mahaputri, R. A., & Rahayu, Y. M. (2025). English public speaking anxiety: Intercultural communicative competence perspective. Jurnal Pendidikan Indonesia, 6(7). https://japendi.publikasiindonesia.id/index.php/japendi/article/view/8446

Bakhtiyarovna, S., Qizi, Y. B., Yuldashovna, Z., Anvarovna, E., Rustamova, Z., & Qizi, Z. (2024). Requirements to improve students' speaking skills through information technology. SPAST Reports, 1(7). https://doi.org/10.69848/sreports.v1i7.5087

Bhuana, G. P., Argawati, N. O., & Aprilianti, D. L. (2025). Integrasi deep learning dan teknologi: Upaya meningkatkan keterampilan speaking and listening. Abdimas Siliwangi, 8(3), 718–729. https://doi.org/10.22460/as.v8i3.28013

Cao, L. T. X. (2025). The effects of technology-assisted tasks on students' affect and autonomy: A case study in EFL speaking classes in Vietnam. Asian-Pacific Journal of Second and Foreign Language Education, 10(1), Article 52. https://doi.org/10.1186/s40862-025-00361-w

Çatan, A. G., & Kayaoğlu, M. (2026). Exploring the impact of technology-enhanced intercultural exchanges on EFL learners' speaking performance and awareness of English as a lingua franca. Journal of English as a Lingua Franca, 15, 27–52. https://doi.org/10.1515/jelf-2025-0020

Chai, R. (2026). Deep learning-based intelligent diagnosis and adaptive training system for university English oral proficiency. Scientific Reports, 16. https://doi.org/10.1038/s41598-026-51608-6

Chaya, P., & Inpin, B. (2020). Effects of integrating movie-based mobile learning instruction for enhancing Thai university students' speaking skills and intercultural communicative competence. English Language Teaching, 13(7), 27. https://doi.org/10.5539/elt.v13n7p27

Cheng, L., Gao, Y., Mao, H., & Peng, Y. (2024). Metapragmatic awareness development in Chinese children: A conversational competence perspective. Journal of Pragmatics, 231, 47–60.

https://doi.org/10.1016/j.pragma.2024.07.013

Clarke, V., & Braun, V. (2017). Thematic analysis. The Journal of Positive Psychology, 12(3), 297–298. https://doi.org/10.1080/17439760.2016.1262613

Dewi, O. T. S., & Slamet, J. (2025). Formative ELF-based assessment of spoken communication via learning management system in a culturally diverse classroom. Discover Education, 4(1), Article 516. https://doi.org/10.1007/s44217-025-00970-0

Eden, C. A., Chisom, O. N., & Adeniyi, I. S. (2024). Cultural competence in education: Strategies for fostering inclusivity and diversity awareness. International Journal of Applied Research in Social Sciences, 6(3), 383–392. https://doi.org/10.51594/ijarss.v6i3.895

Haerazi, H., & Irawan, L. A. (2026). Deep learning model integrated with intercultural pragmatic competences in improving EFL learners' speaking skills. Jurnal Ilmiah Global Education, 7(2). https://doi.org/10.55681/jige.v7i2.6308

Hao, L.-L. (2025). Optimization of English teaching mode in colleges and universities based on deep learning in the improvement of intercultural communication skills. Journal of Combinatorial Mathematics and Combinatorial Computing. https://doi.org/10.61091/jcmcc127a-487

Huang, Q.-Y., Chen, Z.-J., & Wu, H. (2026). Effects of AI-enhanced immersive learning environments on English proficiency, intercultural competence, and learner engagement. Frontiers in Education. https://doi.org/10.3389/feduc.2026.1774203

Jain, P. (2024). Authentic materials for intercultural communication: Develop speaking materials that encourage learners to engage in authentic cross-cultural communication effectively in diverse global contexts. Creative Saplings, 3(4), 1–14. https://doi.org/10.56062/gtrs.2024.3.4.563

Kurnia, A., Basthomi, Y., Widiati, U., Cahyono, B. Y., & Tengku Sharif, T. I. S. (2026). Intercultural sensitivity, perceived AI usefulness, and willingness to communicate among Indonesian EFL learners. Vision: Journal for Language and Foreign Language Learning, 15(1), 29–48. https://doi.org/10.21580/vjv16i133953

Li, X.-M. (2021). Promoting English learners' intercultural competence through technology-enhanced model. In 2021 2nd International Conference on Information Science and Education (ICISE-IE) (pp. 1592–1595). IEEE. https://doi.org/10.1109/icise-ie53922.2021.00352

Li, Z., Gu, C., & Bahari, A. (2026). Comparing AI-enhanced instructional designs for EFL postgraduate proficiency: A multisite randomized controlled trial with exploratory ICC outcomes. Computer Assisted Language Learning. Advance online publication. https://doi.org/10.1080/09588221.2026.2658598

Liu, G., Ma, C., Bao, J., & Liu, Z. (2023). Toward a model of informal digital learning of English and intercultural competence: A large-scale structural equation modeling approach. Computer Assisted Language Learning, 38(3), 342–368. https://doi.org/10.1080/09588221.2023.2191652

Liu, G.-Z., Fathi, J., & Rahimi, M. (2023). Enhancing EFL learners' intercultural communicative effectiveness through telecollaboration with native and non-native speakers of English. Computer Assisted Language Learning, 38, 97–127. https://doi.org/10.1080/09588221.2022.2164778

Liu, Z., Lin, G., Tan, H. L., Zhang, H., Lu, Y., Gao, X., Yin, S. X., Sun, H., Goh, H. H., & Wong, L. H. (2025). SingaKids: A multilingual multimodal dialogic tutor for language learning (arXiv:2506.02412). arXiv. https://doi.org/10.48550/arXiv.2506.02412

Ma, G., & Yang, X. (2025). Fostering intercultural competence through AI-driven tools: A case study of linguistic and cultural adaptability in Chinese EFL education. Digital Technologies Research and Applications, 4(1), 158–169. https://doi.org/10.54963/dtra.v4i1.1131

Mu, Y.-F., & Yu, B. (2023). Developing intercultural competence in college business English students: A study of innovative teaching in China. International Journal of Intercultural Relations. https://doi.org/10.1016/j.ijintrel.2022.101747

Pinandhita, F. (2025). Integrating digital tools and learning strategies to enhance speaking proficiency: Evidence from a mixed-methods study in Indonesian EFL universities. EDUKASIA: Jurnal Pendidikan dan Pembelajaran, 6(2). https://doi.org/10.62775/edukasia.v6i2.1674

Popa, D., Repanovici, A., Lupu, D., Norel, M., & Coman, C. (2020). Using mixed methods to understand teaching and learning in COVID 19 times. Sustainability, 12(20), 8726. https://doi.org/10.3390/su12208726

Purwanto, J., Nurhidayati, N., Faizah, U., Rifki, I., & Permataningtyas, D. (2025). Pengembangan model pembelajaran berbasis deep learning untuk peningkatan keterampilan berbicara peserta didik SMP Muhammadiyah Purworejo. JURRIBAH: Jurnal Riset Rumpun Ilmu Bahasa, 4(1). https://doi.org/10.55606/jurribah.v4i1.4744

Putri, E. U., Stianingsih, L., Isfahani, R., Mahyati, Cholik, M. W., & Afipah, P. (2025). Asesmen bahasa Inggris berbasis pembelajaran mendalam (deep learning) di sekolah menengah atas. PT Revormasi Jangkar Philosophia. https://revormasi.com/

Qu, C., & Li, Y. (2022). Oral English auxiliary teaching system based on deep learning. Advances in Multimedia, 2022. https://doi.org/10.1155/2022/4109663

Qu, Z., & Chen, L. (2026). Simulating the "cultural other": The impact of GenAI interlocutors on EFL learners' intercultural communicative competence, speaking anxiety, and learning engagement. Frontiers in Psychology, 17, Article 1799695. https://doi.org/10.3389/fpsyg.2026.1799695

Ramadhani, I. S., Khotimah, K., Putri, R. D., & Arisya, S. (2025). AI tools in improving speaking skills: A literature review. Journal in Teaching and Education Area, 2(3), 349–362. https://doi.org/10.69673/n7ebwk13

Ramli, R., & Kusmaryani, W. (2026). Technology integration for promoting deep learning in EFL contexts. Script Journal: Journal of Linguistics and English Teaching, 11(2). https://doi.org/10.24903/sj.v11i2.2370

Robillos, R. J. (2023). Improving students' speaking performance and communication engagement through technology-mediated pedagogical approach. International Journal

of Instruction, 16(1), 551–572. https://doi.org/10.29333/iji.2023.16131a

Sardana, N., Shekoohi, S., Cornett, E. M., & Kaye, A. D. (2023). Qualitative and quantitative research methods. In A. D. Kaye, R. D. Urman, E. M. Cornett, & A. Edinoff (Eds.), Substance use and addiction research: Methodology, mechanisms, and therapeutics (pp. 65–69). Academic Press. https://digitalscholar.lsuhsc.edu/som_facpubs/862

Sayed, B. T., Bani Younes, Z. B., Alkhayyat, A., Adhamova, I., & Teferi, H. (2024). To be with artificial intelligence in oral test or not to be: A probe into the traces of success in speaking skill, psychological well-being, autonomy, and academic buoyancy. Language Testing in Asia, 14(1), Article 49. https://doi.org/10.1186/s40468-024-00321-0

Setiyono, J., Sholehhudin, M., & Puspitasari, A. (2026). Meaning construction in Indonesian language and literature education students' speaking during academic discussions in higher education. Jurnal Pendidikan Edutama, 13(1), 58–67. https://doi.org/10.30734/jpe.v13i1.5247

Shadiev, R., Yu, J., & Sintawati, W. (2021). Exploring the impact of learning activities supported by 360-degree video technology on language learning, intercultural communicative competence development, and knowledge sharing. Frontiers in Psychology, 12. https://doi.org/10.3389/fpsyg.2021.766924

Shafiee Rad, H., & Roohani, A. (2026). Fostering L2 learners' pronunciation and motivation via affordances of artificial intelligence. Computers in the Schools, 43(1), 19–40. https://doi.org/10.1080/07380569.2024.2330427

Shonfeld, M. (2024). The impact of technology on intercultural education. Social Pedagogy/Pedagogika Społeczna, 91(1), 137–151. https://pedagogikaspoleczna.com/1-2024/

Singh, B. B. (2025). Pedagogical strategies: Teaching methods and approaches to promote student engagement and understanding. Bharati International Journal of Multidisciplinary Research and Development, 3(12). https://bijmrd.com/index.php/volume3-issue12/pedagogical-strategies-teaching-methods-and-approaches-to-promote-student-engagement-and-understanding/

Smith, G. F., & Ziegler, N. (2023). Engagement in technology-mediated TBLT. In C. Lambert, S. Aubrey, & G. Bui (Eds.), The role of the learner in task-based language teaching: Theory and research methods (pp. 91–109). Routledge. https://doi.org/10.4324/9781003227267-8

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

Syair, S. N., & Mujahid, K. (2025). Artificial intelligence as an assisted-language learning tool: A systematic review of its efficacy in enhancing speaking skills among EFL learners. IDEAS: Journal on English Language Teaching and Learning, Linguistics and Literature, 13(2), 7955–7981. https://doi.org/10.24256/ideas.v13i2.9004

Xia, Y., Shin, S.-Y., & Kim, J.-C. (2024). Cross-cultural intelligent language learning system (CILS): Leveraging AI to facilitate language learning

strategies in cross-cultural communication. Applied Sciences, 14(13), 5651. https://doi.org/10.3390/app14135651

Yu, S., & Wong, B.-E. (2026). Preparing Chinese higher vocational students to succeed in a globalized and multicultural world: Enhancing their oral English competence and intercultural competence. Acta Psychologica, 263, 106253. https://doi.org/10.1016/j.actpsy.2026.106253

Yustinah, Y., & Hartono, B. (2026). Enhancing digital literacy and speaking skills of students based on artificial intelligence through deep learning. Asian Journal of Social and Humanities, 4(4), 2602–2616. https://doi.org/10.59888/ajosh.v4i4.670

Zhang, L. (2024). Video-production-integrated spoken English pedagogy: An action research into the effects of video production on Chinese EFL (English as a foreign language) learners' spoken attainment [Doctoral dissertation, Brunel University London]. Brunel University Research Archive. https://bura.brunel.ac.uk/handle/2438/29664

Zheng, H., & Zhong, D.-M. (2026). Optimization of methods for improving English cross-cultural communication skills based on deep learning and the internet of things. Discover Internet of Things, 6. https://doi.org/10.1007/s43926-026-00374-y

Zhong, D.-M. (2024). Cultivation of intercultural communication competence of English teacher training students in colleges and universities in ethnic areas under the background of deep learning. Applied Mathematics and Nonlinear Sciences, 9. https://doi.org/10.2478/amns-2024-1980