

RESEARCH ARTICLE

Learner Perceptions of Multimodal Technology Resources in EFL Oral Instruction: Survey Evidence from a Chinese University

Shuqiong Fang

School of Foreign Languages, Peking University, Beijing 100871, China

Corresponding Author: Shuqiong Fang, **E-mail:** emmafang@pku.edu.cn

ABSTRACT

This study investigated 126 EFL learners' perceptions of co-deployed multimodal technology resources (video models, audio playback, ASR tools, and captioned materials) in a university oral English course. Drawing on the Technology Acceptance Model (TAM), a questionnaire measured perceived usefulness (PU), perceived ease of use (PEOU), satisfaction (SAT), and behavioral intention (BI). All four dimensions scored significantly above the scale midpoint (all $p < .001$). PU and PEOU were positively correlated ($r = .48, p < .001$). Learners with greater prior technology experience reported higher PU ($d = 0.58, p < .01$). A strong PU–SAT correlation was found ($r = .60, p < .001$). Confirmatory factor analysis supported the four-factor structure (CFI = .993, RMSEA = .031). PU predicted BI ($\beta = 0.47, p < .001$) and satisfaction partially mediated the association between PU and BI (indirect effect = 0.19, 95% CI [0.09, 0.30]). These findings provide support for the TAM's applicability to co-deployed multimodal EFL environments.

KEYWORDS

learner perceptions, multimodal technology resources, EFL oral instruction, perceived usefulness, satisfaction, technology acceptance model

ARTICLE INFORMATION

ACCEPTED: 10 August 2026

PUBLISHED: 27 September 2026

DOI: 10.32996/ijllt.2026.9.10.4

1. Introduction

The integration of multimodal technology resources has transformed English as a Foreign Language (EFL) instruction over the past decade. Digital tools such as video models, audio playback systems, automatic speech recognition (ASR) applications, and captioned video materials now provide learners with multiple channels through which to access, process, and produce linguistic input. This integration reflects a pedagogical approach informed by Mayer's (2009) cognitive theory of multimedia learning. According to this framework, learners process information through dual channels—visual and auditory—and carefully coordinated multimodal presentation reduces extraneous cognitive load while fostering deeper generative processing.

Captioned video, for instance, has been shown to improve listening comprehension and vocabulary acquisition simultaneously by reinforcing the connection between orthographic form and phonological representation (Montero Perez et al., 2013; Hsieh, 2020). ASR-based tools facilitate pronunciation development by delivering real-time, individualized feedback on segmental and suprasegmental features (Ngo et al., 2024; Jiang et al., 2021). Intelligent personal assistants and AI-mediated speaking platforms have further expanded the possibilities for low-anxiety oral practice outside the classroom, and has been associated with improvements both in speaking proficiency and willingness to communicate (Tai, 2024; Tai & Chen, 2023; Fathi et al., 2024). Systematic reviews confirm that computer-assisted language learning (CALL) positively affects learner outcomes across educational levels and instructional contexts (Macaro et al., 2012).

A substantial body of empirical research has demonstrated the effectiveness of technology-enhanced environments for L2 oral skill development. However, the vast majority of these studies have focused on single technologies in isolation, leaving the

question of how learners perceive co-deployed multimodal resources—where multiple tools are used concurrently within a single course—largely unexplored.

The global push toward digitalization in education, accelerated by the COVID-19 pandemic, has led to widespread adoption of multimodal resources in language classrooms across all educational levels. In China specifically, national policies such as the Action Plan for Educational Digitalization have promoted the integration of information technology with foreign language instruction, resulting in substantial investment in smart classrooms, online learning platforms, and AI-powered language learning tools at universities nationwide. This policy-driven adoption makes it particularly important to understand how learners perceive the multimodal resources they encounter in their courses, as favorable perceptions are a necessary condition for sustained engagement and effective use.

This policy environment creates both opportunities and challenges for technology integration. On one hand, institutional support and infrastructure investment make a wide range of multimodal resources available to instructors and students. On the other hand, the rapid pace of top-down adoption may outpace pedagogical training and empirical evidence regarding optimal resource configurations. Crucially, policy-driven technology integration has proceeded with limited input from the learners themselves, whose day-to-day experiences with these resources ultimately determine whether institutional investments translate into meaningful learning outcomes. Understanding learner perceptions is therefore not merely a theoretical concern but a practical prerequisite for evidence-based policy decisions in technology-enhanced language education.

Despite these technological advances, EFL oral instruction continues to confront persistent challenges that are particularly acute in the Asian higher education context. Learners in EFL settings typically have limited access to authentic spoken interaction in English, large class sizes that restrict individual speaking time to a few minutes per session, and elevated levels of foreign language speaking anxiety that inhibit oral participation (Zhang et al., 2024; Chen, 2024). In China, where English is a required subject from primary school through university, the College English Test Band 4 and Band 6 examinations have historically emphasized reading and listening skills, leaving oral proficiency comparatively underdeveloped among many graduates. The shift toward communication-oriented language pedagogy in recent curriculum reforms has created an urgent need for instructional approaches that can provide learners with sufficient, low-anxiety opportunities for speaking practice. Multimodal technology resources are particularly well-suited to addressing these challenges because they offer flexible, individualized practice environments that can be accessed repeatedly outside of formal class time, thereby increasing cumulative exposure to target language input while reducing affective barriers (Ebadi & Ebadijalal, 2022; Hoang et al., 2023).

The Technology Acceptance Model (TAM), originally proposed by Davis (1989) and subsequently extended by Venkatesh and Davis (2000), provides a robust theoretical framework for understanding how learners perceive and adopt educational technologies. TAM posits that two primary cognitive beliefs—perceived usefulness, defined as the degree to which a person believes that using a system will enhance their performance, and perceived ease of use, defined as the degree to which using a system will be free of effort—jointly determine behavioral intention to use the technology. Within CALL research, TAM has been extensively applied to examine acceptance of learning management systems, mobile language learning apps, and AI-powered platforms. Meta-analytic evidence supports the model's robust explanatory power across diverse educational contexts, with perceived usefulness consistently emerging as the strongest direct predictor of behavioral intention (Scherer et al., 2019). Recent extensions of the model, including the Integrated Model of Technology Acceptance (IMTA), have incorporated intrinsic motivational factors such as perceived enjoyment and positive psychology variables to better capture the full range of determinants underlying technology acceptance in language learning (Zou et al., 2025; Wang et al., 2026; Liu & Ma, 2024; Liu et al., 2025; An et al., 2023; Cai et al., 2024).

Prior empirical studies have examined learner perceptions of specific technologies in isolation, yielding valuable but fragmented insights. Research on captioned video has investigated the differential effects of full captions, keyword captions, and no captions on comprehension and vocabulary gains (Winke et al., 2010; Aldukhayel, 2021; Lai, 2024). Studies on ASR-based pronunciation training have explored how feedback type, treatment duration, and learner proficiency moderate outcomes (Bashori et al., 2024). Recent work on AI-powered chatbots and speech evaluation systems has reported favorable learner attitudes and improvements in speaking confidence (Fathi et al., 2024; Zou et al., 2023). However, comparatively little research has examined learner perceptions of a co-deployed set of multimodal technology resources within a single EFL oral communication course. Important questions remain unanswered: Which dimensions of technology acceptance are most positively evaluated when multiple resource types are combined? Does prior experience with technology moderate perceptions in such environments? And to what

extent are perceived usefulness and satisfaction aligned when resources are embedded within a coherent instructional design rather than used as supplementary tools?

Furthermore, the specific role of perceived ease of use within integrated multimodal contexts—where learners must navigate multiple tools and interfaces—remains poorly understood. Because most TAM studies in CALL have examined single-technology contexts, it is unclear whether the PU–PEOU relationship observed in simpler environments generalizes to settings where learners must coordinate across multiple resource types within a single instructional sequence. The present study addresses these gaps by applying the full TAM framework—including PU, PEOU, SAT, and BI—to a co-deployed multimodal learning environment in EFL oral instruction.

The present study was designed to address these questions through a survey investigation of 126 EFL learners at a university in Beijing who completed a 16-week oral communication course incorporating four types of multimodal technology resources: video models demonstrating target pronunciation and discourse patterns, audio playback for self-monitoring and self-evaluation, speech recognition tools providing automated feedback on speaking accuracy, and captioned video materials for listening and vocabulary development. A questionnaire adapted from the TAM framework measured four dimensions: perceived usefulness (PU, four items), perceived ease of use (PEOU, four items), satisfaction (SAT, three items), and behavioral intention (BI, three items). The present study tested four hypotheses: (a) learners would rate all four dimensions favorably, with mean scores significantly exceeding the scale midpoint; (b) PU and PEOU would be positively correlated, consistent with the TAM nomological network; (c) learners with greater prior technology experience would report higher PU than those with less experience; and (d) a significant positive correlation would exist between PU and SAT. Beyond these hypothesized relationships, exploratory path analyses were conducted to examine the predictive relationships among TAM constructs and the potential mediating role of satisfaction. By examining these hypotheses within a naturalistic instructional setting that employed multiple multimodal technologies concurrently, this study contributes a course-level perspective on how learners perceive and value multimodal resources for oral English development, extending prior work that has focused predominantly on single technologies in isolation.

2. Literature Review

2.1 Multimodal Technology Resources in EFL Oral Instruction

Multimodal technology resources encompass digital tools that present information through multiple semiotic modes—visual, auditory, and textual—either simultaneously or in carefully sequenced combinations (Mayer, 2009). In the context of EFL oral instruction, four resource types have received particular attention. Video models provide learners with authentic or near-authentic demonstrations of target pronunciation, intonation patterns, facial expressions, and gestures, thereby offering a complete communicative model that learners can observe, analyze, and subsequently imitate (Lee, 2019). Audio playback tools allow learners to record their own speech, listen to it repeatedly, and compare their production against model recordings, a process that promotes self-monitoring and metacognitive awareness of oral performance. ASR-based speech recognition applications provide automated, real-time feedback on pronunciation accuracy at the phoneme, word, and sentence levels, enabling individualized practice at scale (Ngo et al., 2024; Jiang et al., 2021). Captioned or subtitled video materials support bottom-up listening processing by making the phonological stream visible through orthographic text, thereby reducing the cognitive load associated with decoding rapid connected speech (Montero Perez et al., 2013; Hsieh, 2020).

The theoretical rationale for multimodal input in language learning draws on Mayer's (2009) cognitive theory of multimedia learning, which proposes that the human information processing system comprises two separate channels—one for visual and one for auditory material—each with limited capacity. Learning is enhanced when information is presented through both channels in a coordinated manner that reduces extraneous processing and promotes generative processing, defined as the deep cognitive engagement required to organize and integrate new information with prior knowledge. In L2 acquisition, this theoretical perspective is complemented by the input hypothesis (Krashen, 1985), which emphasizes the role of comprehensible input in driving language development, and by noticing theory (Schmidt, 1990; Zhang & Zou, 2022), which argues that learners must consciously attend to linguistic features in the input for acquisition to occur. Multimodal presentation can facilitate noticing by directing learners' attention to specific phonological, lexical, or grammatical features through visual highlighting, simultaneous text display, or auditory emphasis, thereby increasing the likelihood that these features will be noticed and subsequently incorporated into learners' developing linguistic knowledge.

Empirical evidence for the effectiveness of individual resource types is substantial. Captioned video has been shown to enhance both listening comprehension and vocabulary acquisition, with meta-analytic findings indicating a medium-to-large overall effect

(Montero Perez et al., 2013). Hsieh (2020) found that full captioning combined with highlighted target items and L1 glosses facilitated both form recognition and meaning recall among Chinese EFL learners, and Aldukhayel (2021) extended these findings to the vlog format, demonstrating comprehension benefits across different caption conditions. ASR-based pronunciation training has attracted growing research attention; Ngo et al. (2024) reported a medium overall effect ($g = 0.69$) in a meta-analysis of 22 studies, with explicit corrective feedback yielding larger effect sizes than indirect feedback. Jiang et al. (2021) and Fang et al. (2021) demonstrated that integrating ASR technology into pre-class tasks in a flipped EFL classroom improved oral lexical and syntactic complexity. Regarding video models, Lee (2019) found that a multimodal approach combining video, audio, and speech visualization technologies produced modest improvements in speaking performance. Relatedly, Chen (2021) reported that technology-mediated task-based language teaching enhanced university students' speaking abilities in a collaborative EFL learning environment. Hou and Min (2026) and Tai (2024) showed that interaction with intelligent personal assistants providing multimodal feedback—combining narration, on-screen text, and graphics—enhanced adolescent EFL learners' speaking proficiency more markedly than single-mode audio feedback alone. Collectively, these findings suggest that combining multiple resource types within a single instructional context may yield complementary benefits, yet few studies have examined how learners perceive such integrated multimodal environments.

2.2 Technology Acceptance Model and Its Application in CALL

The Technology Acceptance Model (TAM), introduced by Davis (1989), provides a theoretically parsimonious yet empirically robust framework for predicting user acceptance of information technology. The model specifies that two belief constructs—perceived usefulness and perceived ease of use—are the primary determinants of behavioral intention to use a technology, with perceived ease of use also exerting an indirect effect through perceived usefulness. In a seminal validation study across four organizations, Davis (1989) demonstrated that these two constructs explained a substantial proportion of variance in self-reported usage, with perceived usefulness consistently emerging as the stronger predictor. Venkatesh and Davis (2000) subsequently extended the model into TAM2 by incorporating social influence processes (subjective norm, voluntariness, image) and cognitive instrumental processes (job relevance, output quality, result demonstrability), accounting for 40–60% of the variance in usefulness perceptions across four longitudinal field studies. More recent extensions have introduced additional predictors including perceived enjoyment, facilitating conditions, and technology self-efficacy, leading to the development of the Unified Theory of Acceptance and Use of Technology (UTAUT) and the Integrated Model of Technology Acceptance (IMTA).

Within CALL research, TAM has been applied to a wide range of technologies and learning contexts. Although Scherer et al. (2019) focused on teachers rather than learners, Scherer et al. (2019) conducted a comprehensive meta-analytic structural equation modeling study of 114 TAM studies involving 34,357 teachers, confirming the model's robust explanatory power while revealing that the strength of specific paths varied depending on sample characteristics and technology type. In EFL speaking specifically, Zou et al. (2023) applied the IMTA to an AI speech evaluation program and found that perceived usefulness and perceived enjoyment were significant direct predictors of behavioral intention, while perceived ease of use exerted an indirect effect through perceived usefulness and perceived enjoyment. Wang et al. (2024) examined the impact of conversational generative AI chatbots on EFL learners' willingness to communicate, and Wang et al. (2025) extended IMTA further by incorporating positive psychology variables, demonstrating that willingness to communicate with generative AI tools directly influenced both perceived usefulness and perceived ease of use, suggesting a reciprocal relationship between learner affect and technology perceptions that previous models had not captured. Liu and Ma (2024) reported that TAM constructs explained substantial variance in Chinese EFL learners' use of ChatGPT for informal digital learning of English, with perceived usefulness again emerging as the dominant predictor.

The evolution from TAM to IMTA reflects a broader recognition that technology acceptance in educational contexts is shaped not only by utilitarian beliefs about usefulness and ease of use but also by learners' intrinsic motivation, emotional states, and prior experiences with technology. In the IMTA framework, perceived enjoyment functions as a parallel pathway to behavioral intention alongside perceived usefulness, capturing the hedonic dimension of technology use that TAM's purely cognitive focus may underestimate in voluntary learning contexts (Zou et al., 2025). This expanded model is particularly relevant to multimodal learning environments, where the variety and novelty of different resource types may enhance learners' enjoyment and engagement beyond what single-technology studies have documented.

Despite the extensive application of TAM in CALL, a notable gap exists in the literature. The vast majority of TAM studies in language learning have focused on learners' acceptance of a single technology—a specific app, platform, device, or software

package—rather than on perceptions of an integrated multimodal learning environment in which multiple resource types are deployed concurrently and designed to complement one another. This distinction matters theoretically and practically. Theoretically, learners' perceptions of an integrated environment may differ from the sum of their perceptions of individual components, as synergistic effects, resource complementarity, and the overall coherence of the instructional design may shape holistic evaluations. Practically, understanding how learners perceive the multimodal environment as a whole is critical for course designers who must make decisions about resource selection, sequencing, and integration within the constraints of a single course or curriculum. The present study addresses this gap by applying the TAM framework to an integrated multimodal instructional context.

2.3 Learner Perceptions, Satisfaction, and Behavioral Intention

Beyond the core TAM constructs of perceived usefulness and perceived ease of use, learner satisfaction has emerged as a key outcome variable in technology-enhanced learning research. Satisfaction reflects learners' overall affective evaluation of the learning experience, encompassing both hedonic (enjoyment-based) and utilitarian (goal-attainment-based) dimensions. In technology acceptance research, satisfaction has been shown to mediate the relationship between perceived usefulness and continuance intention, suggesting that cognitive evaluations of instrumental value are translated into sustained behavioral intentions through the mechanism of affective satisfaction (Bhattacharjee, 2001). Within CALL, Chen (2024) and Chen et al. (2025) found that technology-enhanced language learning reduced EFL learners' public speaking anxiety and that this reduction in anxiety was associated with higher satisfaction and increased willingness to communicate. Fathi et al. (2024) reported that AI-mediated speaking activities improved both speaking skills and willingness to communicate, with learners expressing strongly positive attitudes toward the technology-mediated instructional approach. Zhang et al. (2024) similarly found that AI-mediated EFL speaking practice enhanced enjoyment and reduced foreign language speaking anxiety, with these affective changes predicting learners' willingness to communicate. These findings suggest that satisfaction may be a particularly important construct in the context of multimodal oral instruction, where the variety of tools and activities may enhance learner engagement and enjoyment.

It is also worth noting that satisfaction and behavioral intention, while related, are conceptually distinct outcomes. Satisfaction captures learners' retrospective evaluation of their experience, whereas behavioral intention reflects their forward-looking commitment to continued use. Both are important for understanding the long-term impact of multimodal technology integration: satisfaction sustains engagement in the current course context, while behavioral intention predicts whether learners will transfer their technology use to future learning contexts beyond the structured course environment.

Learner prior experience with technology has also been identified as a moderating variable in technology acceptance. Venkatesh and Davis (2000) found that direct experience with a system moderated the effect of subjective norm on behavioral intention, with the influence of social pressure diminishing as users gained firsthand experience. In educational contexts, learners with greater prior technology experience may be better equipped to recognize the usefulness of multimodal resources because they possess the digital literacy skills—such as navigating multiple interfaces, switching between tools, and troubleshooting technical issues—necessary to exploit these resources effectively (Macaro et al., 2012). Conversely, learners with limited prior experience may face higher cognitive load when engaging with multimodal learning environments, potentially attenuating their perceived usefulness of technology. Understanding the role of prior experience is particularly important in the Chinese higher education context, where students' prior exposure to technology for language learning varies widely depending on their secondary school background and personal access to digital devices.

2.4 Research Hypotheses

Drawing on the theoretical frameworks and empirical evidence reviewed above, the present study tested the following four hypotheses:

H1. EFL learners will report favorable perceptions of multimodal technology resources in oral instruction across all four measured dimensions: perceived usefulness, perceived ease of use, satisfaction, and behavioral intention, with mean scores significantly above the scale midpoint.

H2. Perceived usefulness and perceived ease of use will be positively correlated, supporting the nomological network of the Technology Acceptance Model.

H3. Learners with greater prior technology experience will report significantly higher perceived usefulness of multimodal resources than learners with less prior experience.

H4. Perceived usefulness will be positively correlated with satisfaction with multimodal technology resources.

In addition to testing these hypothesized relationships, exploratory path analyses were conducted to examine the predictive associations among TAM constructs and the potential mediating role of satisfaction.

To test these hypotheses, a cross-sectional survey design was employed. The following sections describe the participant sample, the adapted TAM questionnaire, the multimodal instructional context, and the analytical procedures used to address each research hypothesis.

3. Method

3.1 Participants

A total of 126 undergraduate EFL learners at a university in Beijing participated in this study. All participants were non-English-major undergraduates enrolled in an oral English communication course during the spring semester of the 2024–2025 academic year. The sample comprised 87 female and 39 male students (female = 69.0%), with ages ranging from 18 to 23 years ($M = 19.84$, $SD = 1.37$). All participants were native Mandarin speakers who had studied English as a foreign language for at least seven years, beginning in primary school. At the time of data collection, their English proficiency was at the intermediate level, as determined by university placement tests administered at the start of the academic year. Participants were recruited from intact class sections using convenience sampling; the course instructors invited all enrolled students to participate, and no exclusion criteria were applied beyond enrollment in the course. Incomplete questionnaires were excluded from the final dataset before statistical analysis. A prior power analysis using G*Power 3.1 (Faul et al., 2007) was conducted for the planned subgroup comparison involving prior technology experience (independent-samples t -test; $\alpha = .05$ [two-tailed], power = .80, medium effect size $d = 0.50$). The analysis indicated a minimum required sample of 128 participants, or 132 participants assuming a 60:40 subgroup allocation. The obtained sample of 126 was slightly below this target; however, it represented the complete set of eligible respondents after data screening and was retained for the planned analyses. The implications of sample size limitations are acknowledged in the Discussion section.

The course met twice weekly for 90-minute sessions over 16 weeks, focusing on oral communication skills through presentations, discussions, debates, and role-plays. Multimodal technology resources were integrated into each instructional module to provide input, practice, and feedback aligned with the weekly communicative objectives.

3.2 Instruments

The questionnaire comprised four scales adapted from previous TAM research (Davis, 1989; Venkatesh & Davis, 2000). The items were modified to refer specifically to multimodal technology resources in the context of EFL oral instruction. Perceived usefulness (PU) was measured with four items that assessed the degree to which learners believed the multimodal resources enhanced their oral English learning. A sample item was "Multimodal resources helped me improve my oral English skills." Internal consistency reliability was satisfactory ($\alpha = .85$). Perceived ease of use (PEOU) was measured with four items capturing the perceived effort required to use the multimodal tools. A sample item was "I found multimodal technology tools easy to use" ($\alpha = .82$). Satisfaction (SAT) was measured with three items assessing learners' overall affective response to the multimodal learning experience. A sample item was "I am satisfied with the multimodal resources used in this course" ($\alpha = .88$). Behavioral intention (BI) was measured with three items capturing learners' intention to continue using multimodal resources for future oral English practice. A sample item was "I intend to continue using multimodal resources for oral English practice" ($\alpha = .84$). All items were rated on a five-point Likert scale (1 = strongly disagree, 5 = strongly agree). The full questionnaire was administered in Chinese; the Chinese version was developed through a translation-back-translation procedure involving two bilingual researchers and was piloted with 15 learners from a comparable population before the main data collection, resulting in minor wording adjustments to improve clarity.

Additional examples of PEOU items included "Navigating the multimodal learning platform was straightforward" and "The technical setup for multimodal activities was simple." SAT items captured learners' overall evaluations of the multimodal learning experience, including whether the approach met their expectations for oral English learning. BI items assessed intentions to continue using multimodal resources across different future learning scenarios.

3.3 Procedure

Data were collected via an online questionnaire during the final week of a 16-week oral English communication course. Throughout the semester, learners engaged with all four multimodal resource types as integrated components of the course design rather than as supplementary or optional tools. Video models were used weekly to demonstrate speech acts, presentation structures, and discourse strategies relevant to the week's communicative task. Audio playback software was integrated into homework assignments, requiring learners to record, listen to, and self-evaluate their oral production using a structured reflection template. The iFLYTEK AI speech recognition system provided automated pronunciation scoring and feedback on accuracy at the phoneme, word, and sentence levels. Captioned video materials were used for in-class listening comprehension exercises and vocabulary-building activities, with learners able to choose between full-caption, keyword-caption, and no-caption viewing modes. Learners completed the online questionnaire outside of class time within a specified five-day window. Completion took approximately 15 minutes. Of the 156 students enrolled in the course, 132 responded to the questionnaire (response rate = 84.6%); six responses were excluded due to incomplete data, yielding a final sample of 126. Participation was voluntary and anonymous, and all participants provided digital informed consent prior to accessing the questionnaire. The study was approved by the institutional review board of the host university.

3.4 Data Analysis

Descriptive statistics (means, standard deviations) and bivariate Pearson correlations were computed for all study variables. One-sample t-tests were conducted to compare the mean ratings of each dimension against the scale midpoint of 3.0, providing a formal test of whether learner perceptions were significantly favorable (H1). An independent-samples t-test compared PU scores between learners classified as having higher versus lower prior technology experience based on a median split of their self-reported experience ratings. An additional independent-samples t-test examined whether gender differences existed in perceived usefulness as an exploratory research question; this analysis used the conventional $\alpha = .05$ because it was not part of the confirmatory hypothesis set. Cohen's d was calculated as a standardized effect size measure for all group comparisons, with values of 0.20, 0.50, and 0.80 interpreted as small, medium, and large effects, respectively. Pearson's r was used to assess bivariate correlations, with coefficients of .10, .30, and .50 interpreted as small, medium, and large effects, respectively. Linear regression analysis was conducted to examine whether perceived usefulness predicted behavioral intention, with standardized regression coefficients (β) reported. Prior to hypothesis testing, confirmatory factor analysis (CFA) was conducted using *semopy* in Python to evaluate the fit of the proposed four-factor measurement model. Nested model comparisons were conducted using chi-square difference tests to compare the proposed four-factor model with an alternative single-factor model. Harman's single-factor test was also conducted to assess the potential influence of common method bias (CMB).

Statistical analyses, including descriptive statistics, reliability analysis, correlations, and hypothesis testing, were conducted using SPSS version 27.0, whereas CFA was conducted separately using *semopy*. To address the risk of inflated Type I error arising from multiple significance tests, a Bonferroni correction was applied to the four one-sample t-tests used for Hypothesis 1 (adjusted $\alpha = .0125$). The remaining hypothesis tests—the independent-samples t-test for Hypothesis 3, the bivariate correlation for Hypothesis 4, the regression analyses, and the exploratory gender comparison—were each tested at the conventional .05 level without correction, either because they addressed distinct research questions with separate *a priori* predictions (H3, H4) or because they served a descriptive or exploratory purpose (regression, gender comparison). Scale reliability was assessed using Cronbach's alpha. Mediation analysis was conducted to examine whether satisfaction mediated the relationship between perceived usefulness and behavioral intention. The indirect effect was estimated using bootstrapping with 5,000 resamples, and a 95% confidence interval was calculated. An indirect effect was considered significant when the confidence interval did not include zero.

4. Results

4.1 Descriptive Statistics and Preliminary Analysis

Table 1 presents the means, standard deviations, internal consistency coefficients, and bivariate correlations for the four study variables. All four dimensions received mean ratings above the scale midpoint of 3.0, indicating generally favorable perceptions of the multimodal technology resources. Perceived usefulness received the highest mean rating ($M = 4.12$, $SD = 0.67$), followed by satisfaction ($M = 4.03$, $SD = 0.71$), behavioral intention ($M = 3.92$, $SD = 0.75$), and perceived ease of use ($M = 3.85$, $SD = 0.72$). The relative ordering of these means is noteworthy: learners perceived the multimodal resources as more useful than easy to use, a pattern that is consistent with TAM research showing that usefulness perceptions tend to be more salient than ease-of-

use perceptions in task-relevant contexts (Davis, 1989; Scherer et al., 2019). Standard deviations ranged from 0.67 to 0.75, indicating moderate variability across all four dimensions without floor or ceiling effects. Scale reliability was satisfactory for all measures, with Cronbach's alpha values ranging from .82 (PEOU) to .88 (SAT).

Table 1. Descriptive Statistics, Reliability, and Bivariate Correlations Among Study Variables (N = 126)

Variable	M	SD	α	1	2	3	4
1. PU	4.12	0.67	.85	—			
2. PEOU	3.85	0.72	.82	.48***	—		
3. SAT	4.03	0.71	.88	.60***	.46***	—	
4. BI	3.92	0.75	.84	.55***	.41***	.58***	—

Note. PU = perceived usefulness; PEOU = perceived ease of use; SAT = satisfaction; BI = behavioral intention. *** $p < .001$ (two-tailed). All one-sample t -tests (H1) against the scale midpoint of 3.0 were significant with large effect sizes: PU, $d = 1.67$; SAT, $d = 1.45$; BI, $d = 1.23$; PEOU, $d = 1.18$ (all $p < .001$, Bonferroni-adjusted $\alpha = .0125$).

One-sample t -tests confirmed that the mean of each dimension significantly exceeded the scale midpoint of 3.0 (all $p < .001$), supporting Hypothesis 1. Specifically, perceived usefulness was significantly above the midpoint, followed by satisfaction, behavioral intention, and perceived ease of use. These results indicate that learners held favorable perceptions of the multimodal technology resources across all measured dimensions, with large effect sizes: Cohen's d values ranged from 1.18 (PEOU) to 1.67 (PU).

As hypothesized, perceived usefulness and perceived ease of use were positively and significantly correlated, $r(124) = .48$, $p < .001$, supporting Hypothesis 2. This finding is consistent with the nomological network of the TAM, which proposes that these two perceptual constructs, while conceptually distinct, share meaningful variance because systems that are easier to use are generally perceived as more useful (Davis, 1989; Venkatesh & Davis, 2000). The magnitude of this correlation ($r = .48$) falls within the range typically reported in meta-analytic syntheses of TAM studies in education, broadly consistent with the .50–.60 estimate reported in Scherer et al.'s (2019) meta-analysis of teacher samples. Although Scherer et al. focused on teacher populations, the direction and approximate magnitude of this relationship are broadly consistent with student-sample TAM studies, which typically report PU–PEOU correlations in the .40–.60 range. Examination of the full correlation matrix revealed that all four constructs were significantly and positively intercorrelated, with coefficients ranging from $r = .41$ (PEOU–BI) to $r = .60$ (PU–SAT), providing additional support for the expected pattern of associations among the TAM constructs in this EFL context. The pattern of correlations indicates that PU is the dimension most strongly associated with both satisfaction and behavioral intention, reinforcing its central role in technology acceptance.

The complete correlation matrix provides additional insights into the pattern of relationships among the four TAM dimensions. PU exhibited the strongest correlations with other constructs (ranging from $r = .48$ with PEOU to $r = .60$ with SAT), consistent with its theoretical role as the central evaluative belief in the TAM framework. PEOU, while significantly correlated with all other constructs, showed the weakest associations ($r = .41$ –.48), suggesting that ease of use, while relevant, may play a less central role in shaping learners' overall evaluations of multimodal learning environments than perceived usefulness. SAT and BI were strongly correlated ($r = .58$), indicating that learners who were satisfied with the multimodal experience were also likely to express intentions to continue using similar resources.

4.2 Confirmatory Factor Analysis and Common Method Bias

Prior to hypothesis testing, confirmatory factor analysis (CFA) was conducted to evaluate the fit of the proposed four-factor measurement model comprising perceived usefulness, perceived ease of use, satisfaction, and behavioral intention. The CFA yielded an excellent model fit: $\chi^2(71) = 79.36$, $p = .23$; CFI = .993; RMSEA = .031; GFI = .938. All factor loadings were statistically significant (all $p < .001$), and the factor intercorrelations (latent correlations) ranged from $\phi = .51$ (PEOU–SAT) to $\phi = .67$ (SAT–BI), indicating that the four constructs were related yet empirically distinguishable.

A single-factor CFA model in which all 14 items loaded on a common factor fit the data poorly ($\chi^2[77] = 479.94$, CFI = .664, RMSEA = .205), and a chi-square difference test confirmed that the four-factor model fit significantly better ($\Delta\chi^2[6] = 400.58$, $p < .001$). Harman's single-factor test further indicated that the first unrotated factor accounted for 50.7% of the total variance (PCA upper bound). Although this result does not provide definitive evidence regarding the presence or absence of common

method bias, it suggests that common method bias cannot be completely ruled out and warrants caution in interpreting the observed relationships. Taken together, these results support the multidimensionality and construct validity of the TAM measurement model in the present multimodal EFL context.

4.3 Regression Analysis and Mediation

To examine the relationships among TAM constructs and their predictive associations with behavioral intention and satisfaction, regression-based path analyses were conducted. The standardized coefficients and explained variance are presented in Figure 1.

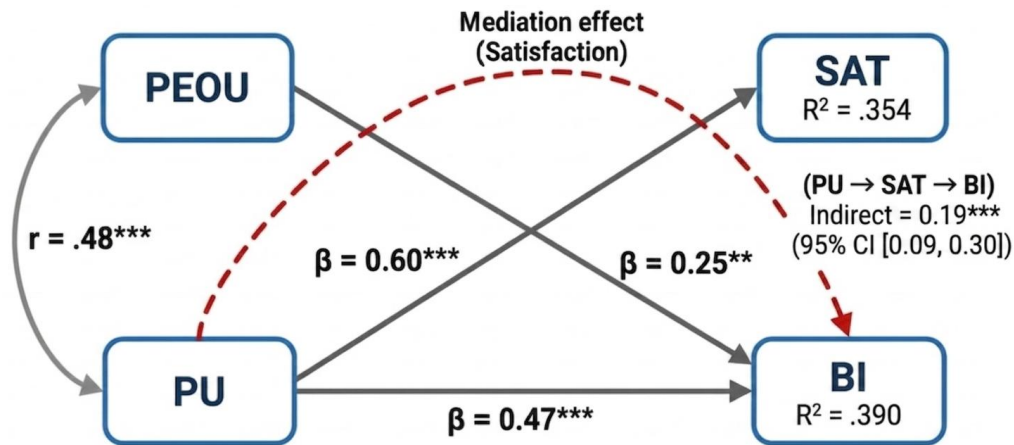


Figure 1. Regression-based model of the Technology Acceptance Model predicting satisfaction and behavioral intention ($N = 126$).

Note. Standardized regression coefficients (β) are shown for directional relationships, and Pearson's correlation coefficient (r) is shown for the association between perceived ease of use (PEOU) and perceived usefulness (PU). R^2 values indicate explained variance. The indirect effect was estimated using 5,000 bootstrap resamples. * $p < .05$, ** $p < .01$, *** $p < .001$.

To further test the predictive relationships specified by the TAM framework, two linear regression analyses were conducted. In the first model, behavioral intention was regressed on perceived usefulness and perceived ease of use. PU emerged as a significant positive predictor of BI ($\beta = 0.47$, $t = 5.88$, $p < .001$), and PEOU also contributed uniquely ($\beta = 0.25$, $t = 3.07$, $p = .003$). In the mediation model, satisfaction was also a significant predictor of behavioral intention ($\beta = 0.32$, $p < .001$). The model explained a substantial proportion of variance in BI (unadjusted $R^2 = .390$), confirming the centrality of perceived usefulness as the stronger direct determinant of behavioral intention in multimodal learning contexts. In the second model, satisfaction was regressed on perceived usefulness, which emerged as a strong and significant predictor ($\beta = 0.60$, $t = 8.24$, $p < .001$) with PU accounting for 35% of the variance in SAT (unadjusted $R^2 = .354$). These path results are consistent with the TAM's theoretical claim that cognitive evaluations of instrumental value drive both intended future use and affective satisfaction. To further examine the mechanism underlying these relationships, a mediation analysis tested whether satisfaction mediated the effect of perceived usefulness on behavioral intention. The indirect effect of PU on BI through SAT was significant (indirect effect = 0.19, bootstrap 95% CI [0.09, 0.30]), with satisfaction accounting for approximately 33% of the total effect. The direct effect of PU on BI remained significant after satisfaction was included in the model ($\beta = 0.47$, $p < .001$), indicating partial rather than full mediation. This result provides empirical support for the theoretical proposition that satisfaction serves as a partial mediator of the PU–BI relationship in multimodal learning contexts (Bhattacharjee, 2001). However, because PU, SAT, and BI were measured at a single time point, the direction of the mediation path represents a theoretical assumption grounded in the TAM framework rather than an empirically tested causal sequence. Longitudinal designs with temporally separated measurements are needed to establish the directionality of these relationships.

4.4 Individual Differences: Prior Experience and Gender

To test Hypothesis 3, participants were divided into higher prior experience ($n = 61$) and lower prior experience ($n = 65$) groups based on a median split of their self-reported prior technology experience scores (median = 3 on a five-point scale). An independent-samples t -test revealed that learners in the higher prior experience group reported significantly greater perceived

usefulness ($M = 4.34$, $SD = 0.72$) than those in the lower prior experience group ($M = 3.92$, $SD = 0.73$), $t(124) = 3.24$, $p < .01$, Cohen's $d = 0.58$. This result supports Hypothesis 3. The effect size ($d = 0.58$) represents a medium-to-large effect according to conventional benchmarks, indicating that the difference between the two groups was not only statistically significant but also practically meaningful. Levene's test for equality of variances was not significant ($p > .05$), confirming that the homogeneity of variance assumption was met.

As an exploratory analysis, an independent-samples t -test examined whether male and female learners differed in their perceptions of multimodal technology usefulness. The results revealed no significant gender difference in perceived usefulness, $t(124) = 1.12$, $p = .26$. Male learners ($M = 4.21$, $SD = 0.62$) and female learners ($M = 4.07$, $SD = 0.69$) reported comparable levels of perceived usefulness, suggesting that the multimodal resources were equally well received across genders. This null finding is consistent with prior CALL research that has generally found small or non-significant gender effects on technology acceptance (Scherer et al., 2019).

Taken together, these results indicate that prior technology experience, rather than gender, was the primary individual-difference factor shaping learners' perceived usefulness of the multimodal resources.

5. Discussion

This survey of 126 EFL learners at a Chinese university examined perceptions of co-deployed multimodal technology resources within a university oral English course. Drawing on the Technology Acceptance Model, the study assessed four dimensions of learner perception: perceived usefulness, perceived ease of use, satisfaction, and behavioral intention. Confirmatory factor analysis (CFA) supported the four-factor measurement model with excellent fit indices ($CFI = .993$, $RMSEA = .031$), and four hypotheses were tested concerning the favorability of these perceptions, the relationships among core TAM constructs, differences associated with prior technology experience, and the association between usefulness and satisfaction. All four hypotheses received empirical support, and several findings merit further discussion in relation to the existing literature.

5.1 Favorable Perceptions Across All Dimensions

The finding that mean scores for PU, PEOU, SAT, and BI all significantly exceeded the scale midpoint (H1) is broadly consistent with the growing body of research reporting positive learner attitudes toward technology-enhanced language learning (Fathi et al., 2024; Chen, 2024). However, the present findings extend this literature in an important way: they demonstrate that favorable perceptions characterize not just a single technology but a multimodal learning environment in which four distinct resource types are deployed concurrently. This is a nontrivial extension because learners must navigate multiple tools, interfaces, and activity types within a single course, which could potentially increase cognitive load or create fragmentation. This interpretation connects to a well-established cognitive mechanism in second language acquisition: the Noticing Hypothesis, which posits that conscious attention to linguistic input is a prerequisite for intake (Schmidt, 1990). The theoretical bridge between noticing and perceived usefulness is that multimodal resources may theoretically enhance the salience of linguistic input through coordinated visual and auditory cues, which could help explain why learners recognized their instrumental value. Within the TAM framework, this increased salience can be understood as strengthening the cognitive belief that the technology effectively supports learning goals, which is the core definition of perceived usefulness. The fact that perceptions remained uniformly favorable suggests that the combination was experienced as coherent rather than disjointed, possibly because the resources were embedded within a consistent pedagogical framework rather than used as a disjointed collection of optional tools.

Two specific findings within this overall pattern warrant interpretive attention. First, PU received the highest mean rating ($M = 4.12$, $SD = 0.67$), suggesting that learners clearly recognized the instrumental value of multimodal resources for developing their oral English skills. This primacy of perceived usefulness is consistent with the TAM framework's claim that usefulness is the most salient determinant of technology acceptance in task-oriented contexts (Davis, 1989) and with meta-analytic evidence showing that PU is the strongest predictor of behavioral intention across educational settings (Scherer et al., 2019).

The strong PU–SAT correlation ($r = .60$) reinforces the central TAM proposition that cognitive evaluations of a technology's instrumental value are closely and positively linked to affective responses (Davis, 1989). The magnitude of this relationship is notable—perceived usefulness explained over one-third of the variance in satisfaction in the regression model—and is consistent with the TAM's theoretical prediction that perceived instrumentality is closely linked to affective outcomes. This finding aligns with the broader evidence in educational psychology showing that perceived instrumentality—the belief that current activities are valuable for achieving future goals—is a robust predictor of both engagement and satisfaction in learning contexts.

Second, PEOU received the lowest mean ($M = 3.85$, $SD = 0.72$), suggesting that while learners found the tools reasonably accessible, there was room for improvement in the user interface or instructional scaffolding. This pattern mirrors findings from Zou et al. (2023), who reported that PEOU exerted only an indirect effect on BI through PU, suggesting that ease of use may be a necessary but not sufficient condition for technology acceptance in multimodal learning contexts.

This pattern of results may be interpreted as serving a dual function. On one hand, it constitutes a replication of well-established TAM relationships—the primacy of PU over PEOU, the moderate PU-PEOU correlation, and the link between cognitive evaluation and affective response have all been documented across educational technology contexts (Davis, 1989; Scherer et al., 2019). On the other hand, the present findings extend these patterns to an understudied context: a course-level multimodal resource ecology in EFL oral instruction, where multiple tools are encountered simultaneously rather than in isolation. The extension is nontrivial because the cognitive and navigational demands of a multi-tool environment could plausibly have attenuated or altered these relationships. At the same time, this claim should be interpreted with caution: because the study did not include a single-technology comparison condition, it cannot isolate the unique contribution of resource co-deployment from the simple presence of any one technology. Future research should adopt component-control designs that systematically vary the number and combination of resource types. Such designs are needed to establish whether co-deployment produces added perceptual benefits beyond single-tool implementations. The fact that the observed TAM nomological pattern under these conditions provides preliminary survey evidence that the model is applicable to co-deployed multimodal designs in EFL speaking instruction. The present findings thus make a dual contribution: they replicate core TAM relationships in a novel instructional ecology, and they extend the model by identifying prior technology experience as a meaningful moderator that conditions how learners evaluate the usefulness of co-deployed multimodal resources.

5.2 TAM Nomological Network

The positive PU-PEOU correlation ($H2$, $r = .48$) supports the foundational assumption of the TAM that these two constructs are related but distinct, sharing variance attributable to the fact that systems perceived as easier to use are also perceived as more useful (Davis, 1989). At the same time, the magnitude of this correlation ($r = .48$) falls at the lower end of the range reported by Scherer et al. (2019) in their meta-analytic synthesis of TAM studies in education, where the estimated PU-PEOU correlation is approximately .50–.60. It should be noted, however, that the Scherer et al. meta-analysis drew primarily on teacher samples, whereas the present study surveyed learners. Because teacher and student populations may differ in their TAM parameter estimates, direct comparisons should be treated as approximate. One possible explanation is that when learners navigate multiple resource types, the perceived ease of each individual tool may contribute less to overall usefulness perceptions, because usefulness depends more on how well the resources collectively support learning goals than on the ease of any single interface. This interpretation, while speculative, highlights the need for further research into how the number and diversity of resources affect the structure of the TAM nomological network. The broader pattern of significant intercorrelations among all four constructs (range $r = .41$ –.60) further supports the construct validity of the TAM measurement model when applied to multimodal learning environments. Notably, the PU-BI correlation ($r = .55$) and the SAT-BI correlation ($r = .58$) in the present study were comparable in magnitude, suggesting that both cognitive evaluations (PU) and affective responses (SAT) contribute meaningfully to learners' intentions to continue using multimodal resources, a finding that aligns with the theoretical proposition that satisfaction mediates the PU-BI relationship (Bhattacharjee, 2001). The significant indirect effect of PU on BI through SAT (indirect effect = 0.19, bootstrap 95% CI [0.09, 0.30]) identified in the present study provides empirical support for this theoretically specified mediation pathway, suggesting that satisfaction may carry a meaningful portion of the association between perceived usefulness and learners' intentions to continue using multimodal resources.

5.3 The Role of Prior Technology Experience

The significant effect of prior technology experience on PU ($H3$, $d = 0.58$) indicates that prior technology experience is an important individual-difference factor associated with technology acceptance in CALL contexts, echoing the experience moderation effect originally documented by Venkatesh and Davis (2000). Learners who entered the course with greater familiarity with digital tools appeared to recognize and exploit the affordances of the multimodal resources more readily, reporting PU scores that were, on average, 0.42 points higher on a five-point scale than those of less experienced peers—a practically meaningful difference. This finding carries implications for instructional practice. Educators cannot assume uniform receptivity to multimodal resources across all learners. Those with limited prior experience may require targeted onboarding. Guided tutorials, step-by-step task cards, or peer modeling can help develop the digital skills necessary to benefit from multimodal learning environments. Without such support, technology-enhanced instruction may inadvertently widen rather than narrow

individual differences in learning outcomes, as learners with greater prior technology experience may be better positioned to perceive value from the available resources. However, this finding should be treated as preliminary because prior technology experience was measured with a single self-report item rather than a validated multi-dimensional scale. Single-item measures cannot be assessed for internal consistency and may capture only a narrow facet of what constitutes technology experience—such as frequency of use—while missing other relevant dimensions including digital self-efficacy, diversity of tool exposure, and attitudes toward technology. Replication with validated multi-item instruments is needed before this moderation effect can be considered robust.

The absence of a significant gender difference in PU is consistent with prior CALL research showing that gender effects on technology acceptance tend to be small and context-dependent (Scherer et al., 2019). In the present study, male and female learners reported statistically equivalent perceptions, suggesting that multimodal resources in EFL oral instruction are equally accessible and appealing across genders. This finding also supports the interpretation that prior technology experience—rather than gender per se—is the relevant individual difference variable in this context. The interaction between gender and prior experience, which could not be examined in the present study due to the imbalanced gender distribution (69% female), would be a worthwhile avenue for future research with more balanced samples.

5.4 Limitations and Directions for Future Research

Several limitations should be acknowledged when interpreting these findings. First, the data were collected at a single university using convenience sampling, and the sample was predominantly female (69.0%). The generalizability of the results to other institutional contexts, proficiency levels, age groups, and demographic profiles requires further investigation through multi-site replication studies with more diverse and balanced samples. Second, the study relied exclusively on self-report measures collected at a single time point. While self-reported perceptions are a valid and widely used indicator of technology acceptance (Davis, 1989), this design is susceptible to common method bias (CMB). Although the Harman's single-factor test did not provide definitive evidence regarding the presence or absence of CMB, the possibility that shared method variance influenced the observed associations cannot be completely ruled out. Future research should incorporate behavioral trace data—such as system log analytics, task completion rates, and patterns of resource use—alongside pre-post measures of oral proficiency to establish whether favorable perceptions translate into measurable learning gains. Third, prior technology experience was assessed by a single self-report item, and participants were divided into higher and lower experience groups using a median split—a method that reduces statistical power, discards within-group variance, and can produce biased estimates under non-normal distributions. Future studies should assess prior experience using validated multi-item scales and analyze it as a continuous moderating variable through regression-based approaches to preserve statistical power and estimation accuracy.

The final sample of 126 was slightly below the target estimated through an a priori G*Power analysis conducted for the planned subgroup comparison. Although the achieved sample size was close to the prespecified target, findings involving small effects should be interpreted with caution. Future studies with larger samples would provide greater statistical power for detecting small effects and examining individual differences.

Fourth, the cross-sectional design precludes causal inference regarding the relationships among the TAM constructs. Although the observed PU–SAT correlation is consistent with the theoretical prediction that perceived usefulness drives satisfaction, the reverse causal path—satisfied learners subsequently perceiving greater usefulness—is equally plausible, as is the possibility of reciprocal effects over time. Longitudinal designs with repeated measures at multiple time points are needed to establish the temporal precedence and directionality of these relationships. Fifth, the present study did not examine the impact of individual resource types on perceptions or learning outcomes separately. While the composite, course-level assessment adopted here reflects the ecological reality that resources operate as a co-deployed system, future component-analysis designs that systematically vary resource availability could identify the unique and interactive contributions of video models, audio playback, ASR, and captioned materials. Sixth, the study did not include a measure of actual technology usage, precluding analysis of the perception–behavior link that lies at the heart of the TAM framework.

Future research should also consider extending the scope of investigation beyond the TAM constructs measured here. The IMTA framework, which incorporates perceived enjoyment alongside PU and PEOU, may offer a more complete picture of technology acceptance in multimodal learning contexts where the variety of resource types may generate enjoyment effects that a purely utilitarian model underestimates. Additionally, studies examining the relationship between learner perceptions and actual learning outcomes—measured through pre-post oral proficiency assessments—would address the critical question of whether

favorable perceptions translate into improved performance. Such research would help educators make evidence-based decisions about which multimodal configurations yield the strongest combination of positive learner perceptions and measurable skill development.

A further methodological consideration concerns the measurement of prior technology experience. The single-item measure used in this study, while practical for survey administration, may not capture the multidimensional nature of technology experience, which includes frequency of use, diversity of tools used, digital self-efficacy, and attitudes toward technology. Future research employing validated multi-item scales of digital literacy and technology readiness could provide more precise estimates of the moderating effect of prior experience on technology acceptance in multimodal learning contexts and could identify which specific dimensions of experience are most consequential for learner perceptions.

5.5 Practical Implications

Despite these limitations, the findings offer differentiated guidance for three stakeholder groups. For EFL instructors, the uniformly favorable perceptions across all four TAM dimensions suggest that multimodal technology resources can be incorporated into oral English courses without compromising learner satisfaction, provided that each resource is introduced with a clear explanation of its pedagogical purpose. The significant role of prior technology experience further implies that instructors should not assume uniform digital readiness: brief orientation sessions at the start of the semester, peer tutoring pairings, or step-by-step screencast tutorials accompanying each resource type can help less experienced learners benefit equally (Shi & Shakibaei, 2025).

For course designers, the strong PU–SAT correlation ($r = .60$) underscores that perceived usefulness is the primary lever for learner satisfaction. Design efforts should therefore prioritize the pedagogical relevance and goal alignment of each resource over expanding the number of tools available. A graduated-release instructional model—moving from instructor-guided use early in the semester to autonomous learner selection of resources later—may help bridge the experience gap while fostering self-regulated learning. For teacher education programs, the findings carry a specific implication: pre-service teachers should receive hands-on training in selecting, sequencing, and scaffolding multimodal resources during their practicum training. Given that prior technology experience is associated with differences in perceptions, future teachers who enter the profession with strong digital literacy may be better positioned to design multimodal learning experiences that their students will perceive as useful and satisfying. Incorporating technology integration modules into TESOL and EFL methodology courses could help bridge this readiness gap before teachers enter the classroom. These training needs align with the Technological Pedagogical Content Knowledge (TPACK) framework (Koehler & Mishra, 2009), which underscores that effective technology integration requires teachers to develop not only technical proficiency but also the pedagogical knowledge to align specific tools with communicative learning objectives—a competence that, as the present findings suggest, may influence how learners perceive the usefulness of multimodal resources in their oral skill development.

For institutional decision-makers, the findings support continued investment in multimodal technology infrastructure, but also suggest that hardware and software purchases should be accompanied by training budgets for both instructors and students. Without such complementary investments, technology-enhanced instruction may inadvertently widen achievement gaps between digitally literate and less experienced learners.

At the program level, the findings also carry implications for curriculum sequencing. Rather than introducing all four resource types simultaneously, programs may benefit from a phased rollout in which simpler, single-tool activities precede multi-tool integrative tasks. This approach would allow learners to develop competence and comfort with each resource type before facing the combined cognitive demands of a fully multimodal environment, potentially reducing the ease-of-use barriers reflected in the relatively lower PEOU ratings observed in this study.

6. Conclusion

This study contributes survey-based evidence suggesting that the TAM framework offers a useful framework for understanding learner perceptions when multiple technology resources are co-deployed within a university EFL oral communication course. The central contributions are twofold. First, the core TAM relationships—the primacy of PU over PEOU, the moderate PU–PEOU correlation, and the strong PU–SAT link—are observed in a multi-resource instructional ecology, providing preliminary evidence for the model's applicability across different instructional contexts. Within this pattern, satisfaction partially mediates the PU–BI relationship (indirect effect = 0.19, 95% CI [0.09, 0.30]), suggesting a possible mechanism through which cognitive evaluations may be associated with behavioral intentions. Second, prior technology experience was associated with differences in perceived

usefulness ($d = 0.58$), identifying prior technology experience as a meaningful individual-difference variable that conditions how learners evaluate the usefulness of multimodal resources in CALL contexts. As multimodal technologies continue to proliferate in language classrooms worldwide, understanding how learner background relates to technology perceptions will remain essential for designing equitable and effective instruction.

Future research should extend the present findings by employing longitudinal designs that track how learner perceptions evolve over the course of a semester, experimental designs that isolate the effects of individual resource types, and qualitative methods that capture the richness of learners' subjective experiences with multimodal technology. Such multi-method, multi-time-point approaches would provide a more complete understanding of the dynamic relationships among learner perceptions, technology engagement, and oral skill development, ultimately supporting the design of more effective and engaging multimodal learning environments for EFL oral instruction.

Statements and Declarations

Funding: This research was funded by the Undergraduate Teaching Reform Project of Peking University (Grant No. 7100904793).

Ethical Approval: This study was approved by the Institutional Review Board of Peking University. Informed consent was obtained from all participants before data collection.

Conflicts of Interest: The author declares no conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers.

References

- [1] Aldukhayel, D. (2021). The effects of captions on L2 learners' comprehension of vlogs. *Language Learning & Technology*, 25(2), 178–191. <https://doi.org/10.64152/10125/73439>
- [2] An, X., Chai, C. S., Li, Y., Zhou, Y., & Yang, B. (2023). Modeling students' perceptions of artificial intelligence assisted language learning. *Computer Assisted Language Learning*, 38(5–6), 987–1008. <https://doi.org/10.1080/09588221.2023.2246519>
- [3] Bashori, M., van Hout, R., Strik, H., & Cucchiarini, C. (2024). "I Can Speak": Improving English pronunciation through automatic speech recognition-based language learning systems. *Innovation in Language Learning and Teaching*, 18(5), 443–461. <https://doi.org/10.1080/17501229.2024.2315101>
- [4] Bhattacharjee, A. (2001). Understanding information systems continuance: An expectation-confirmation model. *MIS Quarterly*, 25(3), 351–370. <https://doi.org/10.2307/3250921>
- [5] Cai, Q., Lin, Y., & Yu, Z. (2024). Factors influencing learner attitudes towards ChatGPT-assisted language learning in higher education. *International Journal of Human-Computer Interaction*, 40(22), 7112–7126. <https://doi.org/10.1080/10447318.2023.2261725>
- [6] Chen, K. (2021). The effects of technology-mediated TBLT on enhancing the speaking abilities of university students in a collaborative EFL learning environment. *Applied Linguistics Review*, 12(2), 331–352. <https://doi.org/10.1515/applirev-2018-0126>
- [7] Chen, Y. (2024). Effects of technology-enhanced language learning on reducing EFL learners' public speaking anxiety. *Computer Assisted Language Learning*, 37(4), 789–813. <https://doi.org/10.1080/09588221.2022.2055083>
- [8] Chen, Y., Zhi, Y., & Derakhshan, A. (2025). Integrating artificial intelligence (AI) into the English as a foreign language classroom: Exploring its impact on students' achievement emotions and willingness to communicate (WTC). *European Journal of Education*, 60(3), e70157. <https://doi.org/10.1111/ejed.70157>
- [9] Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly*, 13(3), 319–340. <https://doi.org/10.2307/249008>
- [10] Ebadi, S., & Ebadijalal, M. (2022). The effect of Google Expeditions virtual reality on EFL learners' willingness to communicate and oral proficiency. *Computer Assisted Language Learning*, 35(8), 1975–2000. <https://doi.org/10.1080/09588221.2020.1854311>

- [11] Fang, W.-C., Yeh, H.-C., Luo, B.-R., & Chen, N.-S. (2021). Effects of mobile-supported task-based language teaching on EFL students' linguistic achievement and conversational interaction. *ReCALL*, 33(1), 71–87. <https://doi.org/10.1017/S0958344020000208>
- [12] Fathi, J., Rahimi, M., & Derakhshan, A. (2024). Improving EFL learners' speaking skills and willingness to communicate via artificial intelligence-mediated interactions. *System*, 121, 103254. <https://doi.org/10.1016/j.system.2024.103254>
- [13] Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, 39(2), 175–191. <https://doi.org/10.3758/BF03193146>
- [14] Hoang, D. T. N., McAlinden, M., & Johnson, N. F. (2023). Extending a learning ecology with virtual reality mobile technology: Oral proficiency outcomes and students' perceptions. *Innovation in Language Learning and Teaching*, 17(3), 491–504. <https://doi.org/10.1080/17501229.2022.2070626>
- [15] Hou, Z., & Min, S. (2026). Dialogue-based computer-assisted language learning systems for second language speaking development: A three-level meta-analysis. *ReCALL*, 38(1), 40–56. <https://doi.org/10.1017/s0958344025100268>
- [16] Hsieh, Y. (2020). Effects of video captioning on EFL vocabulary learning and listening comprehension. *Computer Assisted Language Learning*, 33(5–6), 567–589. <https://doi.org/10.1080/09588221.2019.1577898>
- [17] Jiang, M. Y.-C., Jong, M. S.-Y., Lau, W. W.-F., Chai, C.-S., & Wu, N. (2021). Using automatic speech recognition technology to enhance EFL learners' oral language complexity in a flipped classroom. *Australasian Journal of Educational Technology*, 37(2), 110–131. <https://doi.org/10.14742/ajet.6798>
- [18] Koehler, M. J., & Mishra, P. (2009). What is technological pedagogical content knowledge (TPACK)? *Contemporary Issues in Technology and Teacher Education*, 9(1), 60–70.
- [19] Krashen, S. (1985). *The input hypothesis: Issues and implications*. Longman.
- [20] Lai, C.-J. (2024). Examining the impact of multimodal task design on English oral communicative competence in fourth-grade content-language integrated social studies: A quasi-experimental study. *Asian-Pacific Journal of Second and Foreign Language Education*, 9, 64. <https://doi.org/10.1186/s40862-024-00289-7>
- [21] Lee, Y.-J. (2019). Integrating multimodal technologies with VARK strategies for learning and teaching EFL presentation: An investigation into learners' achievements and perceptions of the learning process. *Australian Journal of Applied Linguistics*, 2(1), 17–31. <https://doi.org/10.29140/ajal.v2n1.118>
- [22] Liu, G., Darvin, R., & Ma, C. (2025). Exploring AI-mediated informal digital learning of English (AI-IDLE): A mixed-method investigation of Chinese EFL learners' AI adoption and experiences. *Computer Assisted Language Learning*, 38(7), 1632–1660. <https://doi.org/10.1080/09588221.2024.2310288>
- [23] Liu, G., & Ma, C. (2024). Measuring EFL learners' use of ChatGPT in informal digital learning of English based on the technology acceptance model. *Innovation in Language Learning and Teaching*, 18(2), 125–138. <https://doi.org/10.1080/17501229.2023.2240316>
- [24] Macaro, E., Handley, Z., & Walter, C. (2012). A systematic review of CALL in English as a second language: Focus on primary and secondary education. *Language Teaching*, 45(1), 1–43. <https://doi.org/10.1017/S0261444811000395>
- [25] Mayer, R. E. (2009). *Multimedia learning* (2nd ed.). Cambridge University Press.
- [26] Montero Perez, M., Van Den Noortgate, W., & Desmet, P. (2013). Captioned video for L2 listening and vocabulary learning: A meta-analysis. *System*, 41(3), 720–739. <https://doi.org/10.1016/j.system.2013.07.013>
- [27] Ngo, T. T.-N., Chen, H. H.-J., & Lai, K. K.-W. (2024). The effectiveness of automatic speech recognition in ESL/EFL pronunciation: A meta-analysis. *ReCALL*, 36(1), 4–21. <https://doi.org/10.1017/s0958344023000113>
- [28] Scherer, R., Siddiq, F., & Tondeur, J. (2019). The technology acceptance model (TAM): A meta-analytic structural equation modeling approach to explaining teachers' adoption of digital technology in education. *Computers & Education*, 128, 13–35. <https://doi.org/10.1016/j.compedu.2018.09.009>
- [29] Schmidt, R. (1990). The role of consciousness in second language learning. *Applied Linguistics*, 11(2), 129–158. <https://doi.org/10.1093/applin/11.2.129>

- [30] Shi, W., & Shakibaei, G. (2025). Insights into the effectiveness of artificial intelligence-integrated speaking instruction in enhancing speaking skills and social-emotional competence as well as reducing demotivation and shyness. *European Journal of Education*, 60(3), e70174. <https://doi.org/10.1111/ejed.70174>
- [31] Tai, T.-Y., & Chen, H. H.-J. (2023). The impact of Google Assistant on adolescent EFL learners' willingness to communicate. *Interactive Learning Environments*, 31(3), 1485–1502. <https://doi.org/10.1080/10494820.2020.1841801>
- [32] Tai, T.-Y. (2024). Effects of intelligent personal assistants on EFL learners' oral proficiency outside the classroom. *Computer Assisted Language Learning*, 37(5–6), 1281–1310. <https://doi.org/10.1080/09588221.2022.2075013>
- [33] Venkatesh, V., & Davis, F. D. (2000). A theoretical extension of the technology acceptance model: Four longitudinal field studies. *Management Science*, 46(2), 186–204. <https://doi.org/10.1287/mnsc.46.2.186.11926>
- [34] Wang, C., Du, Y., & Zou, B. (2026). Learners' acceptance and use of multimodal artificial intelligence (AI)-generated content in AI-mediated informal digital learning of English. *International Journal of Applied Linguistics*, 36(1), 927–940. <https://doi.org/10.1111/ijal.12827>
- [35] Wang, C., Zou, B., Du, Y., & Wang, Z. (2024). The impact of different conversational generative AI chatbots on EFL learners: An analysis of willingness to communicate, foreign language speaking anxiety, and self-perceived communicative competence. *System*, 127, 103533. <https://doi.org/10.1016/j.system.2024.103533>
- [36] Wang, C., Li, X., & Zou, B. (2025). Revisiting integrated model of technology acceptance among the generative AI-powered foreign language speaking practice: Through the lens of positive psychology and intrinsic motivation. *European Journal of Education*, 60(1), e70054. <https://doi.org/10.1111/ejed.70054>
- [37] Winke, P., Gass, S., & Sydorenko, T. (2010). The effects of captioning videos used for foreign language listening activities. *Language Learning & Technology*, 14(1), 65–86. <https://doi.org/10.64152/10125/44203>
- [38] Zhang, C., Meng, Y., & Ma, X. (2024). Artificial intelligence in EFL speaking: Impact on enjoyment, anxiety, and willingness to communicate. *System*, 121, 103259. <https://doi.org/10.1016/j.system.2024.103259>
- [39] Zhang, R., & Zou, D. (2022). A state-of-the-art review of the modes and effectiveness of multimedia input for second and foreign language learning. *Computer Assisted Language Learning*, 35(9), 2790–2816. <https://doi.org/10.1080/09588221.2021.1896555>
- [40] Zou, B., Du, Y., Wang, Z., Chen, J., & Zhang, W. (2023). An investigation into artificial intelligence speech evaluation programs with automatic feedback for developing EFL learners' speaking skills. *SAGE Open*, 13(3), 1–8. <https://doi.org/10.1177/21582440231193818>
- [41] Zou, B., Lyu, Q., Han, Y., Li, Z., & Zhang, W. (2025). Exploring students' acceptance of an artificial intelligence speech evaluation program for EFL speaking practice: An application of the Integrated Model of Technology Acceptance. *Computer Assisted Language Learning*, 38(5–6), 1366–1391. <https://doi.org/10.1080/09588221.2023.2278608>