Journal of English Language Teaching and Applied Linguistics
ISSN: 2707-756X
DOI: 10.32996/jeltal
Journal Homepage: www.al-kindipublisher.com/index.php/jeltal

JELTAL
AL-KINDI CENTER FOR RESEARCH AND DEVELOPMENT

| RESEARCH ARTICLE

Multimodal Technology Reliance Predicts English-Chinese Interpreting Outcomes: Survey Evidence from Non-English Major Undergraduates in China

Shuqiong Fang
School of Foreign Languages, Peking University, Beijing 100871, China
Corresponding Author: Shuqiong Fang, **E-mail**: emmafang@pku.edu.cn

| ABSTRACT

Multimodal technologies are increasingly embedded in interpreter training, yet learner perceptions of such technologies and their relationship with interpreting learning outcomes remain underexplored. This study surveyed 312 non-English major undergraduates (above CET-4, STEAM backgrounds) from multiple universities in Beijing who completed a compulsory English-Chinese interpreting course incorporating multimodal technology resources including video captioning, audio playback, speech recognition, and parallel multimodal materials. Using structural equation modeling, a partial mediation model was examined in which multimodal technology reliance (MTR) predicts interpreting learning outcomes (ILO) both directly and indirectly through learning engagement (LE). Results showed that MTR had a significant total effect on ILO (β = .37, p < .001), with a direct path (β = .16, p < .05) and an indirect path via LE (β = .21, p < .01) that accounted for 56.8% of the total effect. The model demonstrated good fit: CFI = .93, RMSEA = .067, SRMR = .048, chi2/df = 2.38. Common method bias was assessed using Harman's single-factor test and an unmeasured latent method construct (ULMC) approach. The first unrotated factor accounted for 44.2% of the total variance (below the 50% threshold), and the ULMC analysis showed that all substantive loadings remained significant with minimal change (most $|\Delta\lambda|$ < .10), together indicating that common method bias was not a serious concern. These findings suggest that fostering purposeful multimodal technology use may enhance interpreting learning in part by strengthening learner engagement, offering empirical guidance for designing technology-enhanced interpreting curricula that prioritize engagement as a core mechanism.

| KEYWORDS

Multimodal technology reliance, English-Chinese interpreting, interpreting learning outcomes, learning engagement, structural equation modeling, non-English major undergraduates.

| ARTICLE INFORMATION

ACCEPTED: 30 July 2026 **PUBLISHED:** 31 August 2026 **DOI:** 10.32996/jeltal.2026.8.9.5

1. Introduction

Multimodal technologies—including video captioning, speech recognition software, audio playback systems, and parallel multimedia materials—have become increasingly prevalent in language education and interpreter training over the past decade (Chan, 2023; Mayer, 2014; Moreno & Mayer, 2007). These tools offer learners multiple channels for input processing and skill practice, and their pedagogical value has been demonstrated across a range of second-language (L2) skills, including listening comprehension (Montero Perez et al., 2013), vocabulary acquisition (Ramezanali et al., 2021; Teng, 2022), and reading comprehension (Hsieh, 2020). In interpreter training specifically, technology-supported environments are thought to reduce cognitive load by allowing learners to control input pacing, revisit source materials, and compare their output with model renditions (Gile, 2009; Sweller, 2011; Tian & Yang, 2024). Despite the growing presence of multimodal technologies in interpreting classrooms, the mechanisms through which learners' reliance on these technologies translates into improved interpreting outcomes remain poorly understood. Identifying these mechanisms is critical for optimizing curriculum design and ensuring that technology integration genuinely enhances learning rather than merely increasing tool exposure.

Understanding how multimodal technology reliance influences learning is particularly important in English-Chinese interpreting, which places exceptional demands on working memory, cross-linguistic switching, and real-time information processing under time constraints (Pöchhacker, 2016). Theoretical frameworks from cognitive psychology and educational technology—such as cognitive load theory (Sweller, 2011), the cognitive theory of multimedia learning (Mayer, 2014), and the technology acceptance model (Davis, 1989; Venkatesh et al., 2003)—have been used to examine technology adoption and its cognitive effects in interpreting contexts (Yang & Wang, 2019). However, these frameworks primarily address either the cognitive affordances of technology or users' intention to adopt it; they do not fully account for the behavioral, emotional, and cognitive processes through which technology use may foster or hinder actual learning outcomes. One promising, yet underexplored, mechanism is learning engagement, a multidimensional construct encompassing behavioral, emotional, and cognitive involvement in learning activities (Fredricks et al., 2004, 2016; Henrie et al., 2015).

Prior work on engagement in technology-mediated language learning has shown that digital tools can enhance student engagement by supporting autonomy, competence, and relatedness (Chiu, 2021; Deci & Ryan, 1985; Ryan & Deci, 2000), and that engagement in turn predicts academic achievement across various L2 learning contexts (Dai & Wang, 2024; Hiver et al., 2021; Schindler et al., 2017). Technology-facilitated features such as immediate feedback, multimodal input, and self-paced practice have been linked to higher levels of both behavioral and cognitive engagement (Chen et al., 2010; Manwaring et al., 2017). In the interpreting domain, studies on computer-assisted interpreter training have largely focused on tool usability and learner perceptions (Chan, 2023; Tian & Yang, 2024), while research on interpreter training engagement has primarily investigated classroom participation or motivation in isolation (Cai et al., 2023; Liu & Liang, 2024). While prior work has explored technology use and engagement separately in language learning, few studies have specifically examined the mediating role of engagement in the relationship between multimodal technology reliance and interpreting outcomes among non-English major undergraduates.

To this end, the present study proposes and tests a partial mediation model in which multimodal technology reliance (MTR) predicts English-Chinese interpreting learning outcomes (ILO) both directly and indirectly through learning engagement (LE). A total of 312 non-English major undergraduates from multiple universities in Beijing who had completed a compulsory English-Chinese interpreting course incorporating multimodal technologies participated in the survey. Using structural equation modeling, the direct and indirect pathways from MTR to ILO were examined. It was hypothesized that MTR would predict ILO directly (H1) and through the mediation of LE (H4), with MTR positively predicting LE (H2) and LE positively predicting ILO (H3). The findings confirm these hypotheses and suggest that technology reliance contributes to interpreting learning primarily by strengthening learner engagement, highlighting the mediating role of engagement as a critical pathway for technology-enhanced interpreter education.

The Chinese higher education context provides a particularly relevant setting for examining these relationships. With the rapid expansion of English-language programs across Chinese universities, interpreting courses now reach a diverse student population with varying levels of language proficiency and technology readiness (Tian & Yang, 2024; Chan, 2023). Understanding how technology reliance operates through engagement to shape outcomes in this context can inform pedagogical practices and technology investment decisions across similar educational settings worldwide. By clarifying the mechanism through which technology reliance influences interpreting outcomes, this study aims to provide actionable insights for curriculum design and technology integration in interpreter training.

2. Literature Review

2.1 Multimodal Technology Reliance in Interpreter Education

The integration of multimodal technologies into interpreter training has accelerated considerably over the past decade. Computer-assisted interpreter training (CAIT) encompasses a range of tools including digital audio and video platforms, speech recognition software, captioning tools, and online interpreting simulations that provide learners with flexible, self-paced practice environments (Chan, 2023). These technologies are grounded in the cognitive theory of multimedia learning, which posits that learners process information more effectively when it is presented through multiple sensory channels (Mayer, 2014; Mayer et al., 2014). In interpreting specifically, multimodal input reduces the cognitive burden of real-time language processing by allowing learners to allocate attention across visual and auditory channels strategically (Moreno & Mayer, 2007; Sweller, 2011). Yang et al. (2025) recently provided empirical evidence that multimodal processing demands in simultaneous interpreting with text can be measured through ear-eye-voice span, demonstrating that the coordination of visual and auditory inputs is a key determinant of interpreting performance.

However, these studies have focused primarily on technology acceptance—the perceived ease of use and usefulness of technologies (Davis, 1989)—and on behavioral intentions to use technology (Venkatesh et al., 2003), rather than on learners' actual reliance on multimodal technologies in authentic learning tasks. Multimodal technology reliance (MTR) as conceptualized in the present study differs from technology acceptance in two fundamental respects. First, MTR captures the extent to which learners depend on specific multimodal features (captioned video, audio replay, speech recognition, and integrated multimodal tools) for task completion during interpreting learning, rather than their general attitudes toward or intentions to use technology. Second, whereas TAM and UTAUT treat technology use as voluntary and consciously evaluated, MTR reflects a habitual dependence that may develop as learners repeatedly rely on technological scaffolds to manage the cognitive demands of interpreting tasks. This distinction is important because learners may display high reliance on multimodal technologies without holding uniformly positive attitudes toward them, and vice versa. By shifting the analytical focus from acceptance to actual reliance, the present study addresses a critical gap in understanding the mechanism through which technology use translates into improved interpreting performance, a question that the existing technology acceptance frameworks are not designed to answer.

2.2 Learning Engagement in Technology-Mediated Language Learning

Learning engagement has been conceptualized as a multidimensional construct encompassing behavioral, emotional, and cognitive dimensions (Fredricks et al., 2004, 2016). Behavioral engagement refers to observable participation in learning activities such as attending class, completing assignments, and practicing skills; emotional engagement involves positive affective responses to the learning environment including interest, enjoyment, and a sense of belonging; and cognitive engagement concerns the depth of mental processing and the use of self-regulation strategies (Henrie et al., 2015; Hiver et al., 2021). In technology-mediated language learning, engagement has been identified as a key predictor of academic achievement (Schindler et al., 2017). Chen et al. (2010) demonstrated that web-based learning technology enhanced college students' engagement, which in turn predicted learning outcomes in a large-scale survey study. Chiu (2021) found that digital features supporting autonomy, competence, and relatedness as articulated by self-determination theory (Deci & Ryan, 1985; Ryan & Deci, 2000) significantly promoted student engagement in blended learning environments, with the satisfaction of these basic psychological needs mediating the relationship between technology design features and engagement outcomes.

More recently, Dai and Wang (2024) investigated EFL learners' classroom emotions in technology-based English-medium instruction and found that positive emotions such as enjoyment fostered engagement, while negative emotions such as anxiety inhibited it. Wang (2026) reported that multimodal generative AI tools enhanced learner engagement in task-based Chinese language learning by providing immediate feedback and adaptive learning paths. Manwaring et al. (2017) used experience sampling and structural equation modeling to demonstrate that student engagement mediated the relationship between course design features and learning outcomes in blended settings, with behavioral engagement emerging as the strongest mediator. From a social cognitive perspective, learners' engagement with technology is shaped by their self-efficacy beliefs and outcome expectations regarding technology use (Bandura, 1986), yet the role of these mechanisms in interpreting learning has received little empirical attention. Despite this growing body of evidence, engagement has rarely been examined as a mediating mechanism linking technology use specifically to interpreting learning outcomes. Existing engagement studies in language education have focused predominantly on general L2 skills such as reading, writing, and listening (Hiver et al., 2021), leaving interpreting—a cognitively demanding skill requiring real-time cross-linguistic processing—largely unexplored.

2.3 Interpreting Learning Outcomes and the Mediation Model

Assessing interpreting competence is inherently complex, involving not only linguistic accuracy but also cognitive processing efficiency, strategic decision-making, and communicative effectiveness (Gile, 2009; Pöchhacker, 2016). In educational contexts, interpreting learning outcomes are typically measured through performance-based assessments that evaluate multiple dimensions of competence. The present study operationalizes ILO as a composite of six indicators: interpreting accuracy, fluency, source language comprehension, oral expression, overall self-assessment, and end-of-term examination scores, capturing both subjective and objective aspects of learning. Recent work by Cai et al. (2023) demonstrated that psychological factors including self-efficacy, motivation, and anxiety significantly predicted interpreting competence development over a training year, suggesting that affective and motivational variables play a substantial role in interpreting skill acquisition. Yang and Wang (2019) found that technology acceptance predicted translation students' intention to use machine translation tools, which was associated with improved performance outcomes, supporting a technology-to-performance pathway.

Theoretical reasoning supports a mediation model in which multimodal technology reliance predicts interpreting learning outcomes both directly and indirectly through learning engagement. The direct path from MTR to ILO is consistent with cognitive load theory: multimodal tools that provide redundant information across channels can free cognitive resources for higher-order processing, thereby improving interpreting performance directly (Mayer et al., 2014; Sweller, 2011). The indirect path through engagement draws on self-determination theory: technology features such as self-pacing, immediate feedback, and content control may fulfill learners' basic psychological needs for autonomy and competence, thereby strengthening behavioral and cognitive engagement, which in turn promotes deeper learning (Chiu, 2021; Deci & Ryan, 1985). Schindler et al. (2017) similarly argued that computer-based technology enhances student engagement through interactive and personalized features, and that this engagement is a proximal determinant of learning outcomes. Ramezanali et al. (2021) provided meta-analytic evidence that multimodal glossing facilitated L2 vocabulary learning by engaging learners in deeper processing of word meanings, further supporting the engagement-as-mechanism perspective.

Drawing on the above theoretical frameworks, the present study proposes and tests a partial mediation model in which multimodal technology reliance predicts English-Chinese interpreting learning outcomes both directly and indirectly through learning engagement. The following four hypotheses were tested:

H1: Multimodal technology reliance positively predicts interpreting learning outcomes.
H2: Multimodal technology reliance positively predicts learning engagement.
H3: Learning engagement positively predicts interpreting learning outcomes.
H4: Learning engagement mediates the relationship between multimodal technology reliance and interpreting learning outcomes.

3. Methodology

3.1 Participants

A total of 312 non-English major undergraduates participated in this study. All participants were recruited from multiple universities in Beijing, China. They were enrolled in STEAM (science, technology, engineering, arts, and mathematics) programs and had achieved a score above CET-4 (College English Test Band 4), indicating intermediate English proficiency. The sample comprised students from a range of STEAM disciplines including engineering, computer science, information technology, and applied mathematics. Participants were enrolled in a compulsory 16-week English-Chinese interpreting course that met twice weekly for 90-minute sessions. The course incorporated multimodal technology resources, including video captioning tools for listening practice, audio playback and recording functions for self-monitoring, speech recognition software for pronunciation assessment, and parallel multimodal reference materials combining video, audio, and text for comprehensive practice. The sample size of 312 exceeds the minimum recommended threshold for structural equation modeling based on the number of observed indicators (Hair et al., 2018; Kline, 2015) and is consistent with comparable studies in technology-mediated language learning research.

3.2 Measures

All measurement instruments were adopted from established scales in the technology-enhanced learning literature and were adapted for the specific context of interpreter education. Each scale was translated into Chinese through a forward-backward translation procedure conducted by two bilingual researchers, followed by a pilot test with 30 interpreting students to confirm item clarity and appropriateness. Items were rated on a five-point Likert scale ranging from 1 (strongly disagree) to 5 (strongly agree) unless otherwise noted.

Three latent constructs were measured using established scales adapted to the interpreting learning context. All items were rated on a 5-point Likert scale (1 = strongly disagree to 5 = strongly agree). The survey was initially drafted in English, then translated into Chinese by a bilingual researcher, and back-translated into English by an independent translator to verify content validity. A pilot test with 30 students was conducted prior to the main data collection to assess item clarity and internal consistency. Table 2 presents the full list of scale items.

Multimodal Technology Reliance (MTR) was measured with four items assessing the extent to which participants depended on multimodal technological tools during their interpreting learning activities. The items captured reliance on video captions for comprehension, audio playback for self-review, speech recognition for pronunciation assessment, and multimodal materials for parallel practice. Cronbach's alpha for this scale was .82, indicating satisfactory internal consistency.

Learning Engagement (LE) was assessed using four items adapted from validated engagement scales (Fredricks et al., 2004, 2016), capturing behavioral, emotional, and cognitive dimensions of engagement in the technology-enhanced interpreting classroom. A single-factor confirmatory factor analysis was conducted to evaluate the unidimensionality of the LE scale before testing the full measurement model. The model demonstrated acceptable fit: CFI = .99, TLI = .97, RMSEA = .089 (approaching the .08 threshold; see the recommendation of Hu & Bentler, 1999), chi2/df = 3.45. All standardized factor loadings exceeded .70 (.74 to .82), and the composite reliability was .86. These results support the treatment of LE as a single latent construct representing overall learning engagement. Sample items included active participation in multimodal exercises, use of technology tools for self-directed practice, exploration of different multimodal learning approaches, and incorporation of multimodal resources into daily practice routines. Cronbach's alpha was .86.

Interpreting Learning Outcomes (ILO) was operationalized as a composite measure reflecting six indicators: self-reported interpreting accuracy, interpreting fluency, source language comprehension, oral expression ability, an overall improvement assessment (reverse-coded), and end-of-term interpreting test scores obtained from the course's standardized examination. The first five items were assessed on a 5-point Likert scale, while the sixth was an objective performance indicator. Cronbach's alpha for the six indicators was .86, supporting the reliability of the composite measure.

3.3 Procedure

The study was conducted over a 16-week semester in two phases. In the first phase (Weeks 1–8), all participants attended regular interpreting courses covering consecutive and sight interpretation skills. Multimodal technologies were introduced progressively: video annotation tools in Week 2, speech analysis software in Week 4, and multimodal corpus tools in Week 6. In the second phase (Weeks 9–16), participants were expected to integrate these tools into their independent practice routines. The questionnaire was administered in Week 12, and the interpreting test was conducted during Week 16.

Data were collected during the final two weeks of the 16-week interpreting course. Participants completed the MTR and LE scales through an online survey platform along with demographic information and self-reported weekly technology usage hours for interpreting practice. Weekly technology usage hours was entered as a control variable in the structural model to account for the potential confounding effect of overall technology exposure on the hypothesized relationships. The survey took approximately 15 minutes to complete. ILO scores were obtained from the course's standardized end-of-term interpreting assessment, which was conducted under controlled examination conditions. Participation was voluntary, and informed consent was obtained from all respondents prior to data collection. The study received ethical approval from the relevant institutional review boards at all participating universities.

3.4 Data Analysis

Model fit was assessed using multiple indices to reduce the risk of Type I and Type II errors associated with reliance on a single criterion (Hu & Bentler, 1999). The chi-square statistic was reported but was not used as the sole arbiter of model fit due to its sensitivity to sample size. Instead, the CFI and TLI were selected as incremental fit indices, while the RMSEA and SRMR were used as absolute fit indices. Mediation effects were evaluated using bias-corrected bootstrapping, which does not assume multivariate normality and provides greater statistical power than the Sobel test (MacKinnon et al., 2004). All analyses were conducted using IBM SPSS 27.0 and AMOS 28.0 with maximum likelihood estimation.

Descriptive statistics and reliability analyses were conducted using SPSS 27.0. Structural equation modeling was performed with AMOS 28.0 using maximum likelihood estimation. Following the two-step approach recommended by Hair et al. (2018), a confirmatory factor analysis (CFA) was first conducted to evaluate the measurement model, assessing factor loadings, composite reliability, and discriminant validity. The structural model was then specified to test the hypothesized mediation pathways. Model fit was evaluated using the chi-square statistic, the comparative fit index (CFI), the Tucker-Lewis index (TLI), the root mean square error of approximation (RMSEA), and the standardized root mean square residual (SRMR). Acceptable fit was defined as CFI >= .90, TLI >= .90, RMSEA <= .08, and SRMR <= .08 (Byrne, 2016; Kline, 2015). To test the significance of the indirect effect, bias-corrected bootstrap confidence intervals based on 5,000 resamples were used. Common method bias was assessed using Harman's single-factor test.

4. Results/Findings

4.1 Preliminary Analyses

Before testing the structural model, data screening was conducted to verify assumptions of normality, linearity, and homoscedasticity. Skewness values for all items fell within the acceptable range of ±2, and kurtosis values were within ±7, supporting approximate univariate normality (Curran et al., 1996). No multivariate outliers were identified using Mahalanobis distance ($p < .001$). Common method bias was assessed using two approaches. First, Harman's single-factor test showed that the first unrotated factor accounted for 44.2% of the variance, below the 50% threshold. Second, an unmeasured latent method construct (ULMC) was added to the measurement model; all substantive standardized loadings remained significant with negligible change (most $|\Delta\lambda| < .10$), and the method factor accounted for only about 9.7% of the variance on average. Together these results indicate that common method bias was not a serious concern (Podsakoff et al., 2003).

All scales demonstrated acceptable internal consistency, with Cronbach's alpha values exceeding the .70 threshold (Hair et al., 2018): MTR (alpha = .82), LE (alpha = .86), and ILO (alpha = .86). The measurement model was evaluated prior to testing the structural relationships. Common method bias was further evaluated using Harman's single-factor test (44.2%) and the ULMC approach, both of which indicated that method bias was unlikely to substantially affect the data (see Section 4.1). Table 1 summarizes the demographic characteristics of the sample.

Table 1
Participant Demographics (N = 312)

Characteristic	Category	n	%
Gender	Male	168	53.8
	Female	144	46.2
Year of study	Freshman	78	25.0
	Sophomore	134	42.9
	Junior	72	23.1
	Senior	28	9.0
Discipline	Engineering	142	45.5
	Computer Science / IT	98	31.4
	Sciences	47	15.1
	Arts / Design	25	8.0
Weekly tech use	< 2 hours	45	14.4
for interpreting	2-5 hours	156	50.0
	> 5 hours	111	35.6

Note. STEAM = Science, Technology, Engineering, Arts, and Mathematics.

Descriptive statistics revealed that the mean scores for MTR ($M = 3.68$, $SD = 0.80$), LE ($M = 3.52$, $SD = 0.80$), and ILO ($M = 3.34$, $SD = 0.81$) were above the mid-point of the five-point scale, indicating generally positive perceptions among participants. Bivariate correlations showed that MTR was positively correlated with LE ($r = .52$, $p < .001$) and with ILO ($r = .47$, $p < .001$), and LE was positively correlated with ILO ($r = .55$, $p < .001$). These correlation patterns provided preliminary support for the hypothesized mediation model and justified further structural analysis.

4.2 Measurement Model

The measurement model consisted of three latent factors (MTR, LE, and ILO) with their respective indicator items. Each item was specified to load only on its intended latent construct, and all latent factors were allowed to correlate freely in the measurement phase. Model fit was evaluated using the chi-square statistic and several approximate fit indices, with acceptable fit defined as CFI ≥ .90, TLI ≥ .90, RMSEA ≤ .08, and SRMR ≤ .08 (Hu & Bentler, 1999; Kline, 2015).

A confirmatory factor analysis was conducted to evaluate the measurement model before testing the structural model. The three-factor model (MTR, LE, ILO) demonstrated acceptable fit to the data: CFI = .93, TLI = .91, RMSEA = .062 (90% CI [.048, .076]), SRMR = .048, chi2/df = 2.38. All factor loadings were significant ($p < .001$) and exceeded the recommended threshold of .50 (Hair et al., 2018). Standardized factor loadings ranged from .68 (ILO5_rev) to .81 (LE1), supporting convergent validity. Discriminant validity was assessed using the heterotrait-monotrait ratio of correlations (HTMT). The HTMT values were 0.629 for the MTR–LE pair, 0.580 for MTR–ILO, and 0.636 for LE–ILO, all below the recommended threshold of 0.85 (Henseler et al., 2015), confirming adequate discriminant validity. Composite reliability (CR) values were .83 for MTR, .86 for LE, and .86 for ILO, all exceeding the recommended threshold of .70 (Hair et al., 2018). The full list of items and their factor loadings is presented in Table 2.

Table 2

Scale Items and Factor Loadings

Construct	Item	Loading
MTR (alpha = .82)	MTR1: I rely on video captions to support interpreting comprehension	.78
	MTR2: I use audio playback to review my interpreting output	.75
	MTR3: I use speech recognition tools to assess pronunciation accuracy	.71
	MTR4: I use multimodal materials (video+audio+text) for parallel practice	.74
LE (alpha = .86)	LE1: I actively participate in multimodal exercises in interpreting class	.81
	LE2: I use technology tools for self-directed interpreting practice after class	.79
	LE3: I actively try different multimodal learning approaches	.76
	LE4: I incorporate multimodal resources into daily interpreting practice	.80
ILO (alpha = .86)	ILO1: My interpreting accuracy has improved noticeably	.77
	ILO2: My interpreting fluency has improved	.74
	ILO3: My source language comprehension has improved	.73
	ILO4: My oral expression ability has improved	.76
	ILO5: I am not sure my interpreting skills have improved substantially (R)	.68
	ILO6: End-of-term interpreting test score	.72

Note. MTR = Multimodal Technology Reliance; LE = Learning Engagement; ILO = Interpreting Learning Outcomes. (R) = reverse-coded item. All factor loadings were significant at $p < .001$.

4.3 Structural Model and Hypothesis Testing

The structural model was specified to test the hypothesized mediation relationships. The model demonstrated acceptable fit (CFI = .93, RMSEA = .067, SRMR = .048), consistent with recommended thresholds (Byrne, 2016; Kline, 2015). Table 3 presents the standardized path coefficients and effect decomposition. The results showed that MTR positively predicted LE ($\beta = .52$, $p < .001$), supporting H2. LE positively predicted ILO ($\beta = .40$, $p < .001$), supporting H3. The direct path from MTR to ILO was also significant ($\beta = .16$, $p < .05$), supporting H1.

The structural model demonstrated acceptable fit to the data: chi-square = 347.82, df = 146, chi-square/df = 2.38, CFI = .93, TLI = .91, RMSEA = .067 (90% CI [.055, .069]), SRMR = .048. All factor loadings were significant and exceeded .60, confirming that the latent constructs were adequately measured by their respective indicators. The model explained 32.4 percent of the variance in LE (R-squared = .324) and 28.7 percent of the variance in ILO (R-squared = .287). These R-squared values include the contribution of weekly technology usage hours, which was included as a control variable in the structural model. Both values indicate moderate explanatory power and are consistent with effect sizes reported in prior technology-mediated learning research.

4.4 Mediation Analysis

To test the mediation hypothesis (H4), bias-corrected bootstrap confidence intervals based on 5,000 resamples were used. The total effect of MTR on ILO was significant ($\beta = .37$, $p < .001$). The indirect effect via LE was significant ($\beta = .21$, $p < .01$, 95% CI [.09, .34]), confirming that LE partially mediated the relationship between MTR and ILO. The indirect effect accounted for 56.8% of the total effect, indicating that more than half of the predictive relationship between MTR and ILO was transmitted through learning engagement. The direct effect remained significant after controlling for the mediator ($\beta = .16$, $p < .05$), consistent with a partial mediation model. These results provide support for H4.

Table 3

Standardized Path Coefficients and Effect Decomposition

Path	beta	p	Effect size
MTR -> LE	.52	< .001	Large
LE -> ILO	.40	< .001	Medium-large
MTR -> ILO (direct)	.16	< .05	Small
MTR -> ILO (indirect)	.21	< .01	Medium
MTR -> ILO (total)	.37	< .001	Medium

Note. N = 312. Bootstrap resamples = 5,000. Model fit: CFI = .93, RMSEA = .067, SRMR = .048, chi2/df = 2.38.

5. Discussion

5.1 Direct and Indirect Pathways

The direct path from MTR to ILO aligns with the technology-mediated learning literature, which increasingly recognizes that the effectiveness of educational technology depends on how learners interact with tools rather than on the tools themselves (Kimmons et al., 2025). The engagement pathway identified in this study provides empirical support for this perspective within the specific domain of interpreter education, extending previous work conducted in general language learning contexts (Lai, 2019; Hiver et al., 2021).

This study set out to examine whether learning engagement serves as an indirect pathway linking multimodal technology reliance and interpreting learning outcomes among non-English major undergraduates. As the ILO measure was based on self-reported indicators, these findings reflect learners' perceived interpreting outcomes rather than objectively assessed performance. The results provide clear support for the hypothesized partial mediation model. Multimodal technology reliance predicted interpreting learning outcomes both directly (β = .16) and indirectly through learning engagement (β = .21), with the indirect pathway accounting for 56.8% of the total effect. These findings carry both theoretical and pedagogical implications for technology-enhanced interpreter education.

The significant direct path from MTR to ILO indicates that multimodal technologies confer cognitive benefits that are independent of engagement. This finding aligns with the cognitive theory of multimedia learning (Mayer, 2014; Moreno & Mayer, 2007) and cognitive load theory (Sweller, 2011), which predict that providing information through multiple sensory channels reduces extraneous cognitive load and frees processing resources for interpreting task performance. When learners can rely on captioning, speech recognition feedback, and replay functions, they process source material more efficiently and produce higher-quality interpretations. This direct effect, although modest in magnitude, suggests that the functional affordances of multimodal technologies meaningfully support interpreting performance at the cognitive level, irrespective of how engaged the learner may be.

The indirect path through engagement was stronger than the direct path and accounted for the majority of the total effect. This finding is consistent with self-determination theory (Deci & Ryan, 1985; Ryan & Deci, 2000): multimodal technologies may foster engagement by supporting learners' autonomy in pacing their practice, generating immediate feedback that builds competence, and creating interactive environments that satisfy relatedness needs. The results corroborate previous work on technology and engagement in general L2 learning (Chiu, 2021; Dai & Wang, 2024; Henrie et al., 2015) while extending this finding to the specific domain of English-Chinese interpreting, a skill context that has received limited empirical attention. The prominence of the indirect path implies that the primary value of multimodal technologies in interpreting education may not lie in their direct cognitive scaffolding but in their capacity to activate and sustain learner engagement, which in turn drives deeper learning and improved outcomes.

The strong relationship between MTR and LE suggests that students who depend more on multimodal tools also report higher levels of behavioral, emotional, and cognitive involvement. This is consistent with the proposition that technology features such as interactive drills, multimodal input, and automated feedback function as engagement catalysts (Schindler et al., 2017; Wang, 2026). The subsequent strong link between LE and ILO reinforces the well-established finding that engaged learners achieve better outcomes across educational contexts (Fredricks et al., 2004, 2016; Hiver et al., 2021). These results indicate that engagement is not merely a byproduct of technology use but a substantive mechanism through which technology exerts its influence on learning.

5.2 Practical Implications

The findings also carry practical implications for interpreting curriculum development at Chinese universities. The significant direct path from multimodal technology reliance to interpreting outcomes suggests that simply providing access to technology resources can benefit student performance, validating the ongoing investment in interpreting laboratories and digital language platforms. However, the substantial indirect effect through learning engagement indicates that technology resources should be accompanied by pedagogical strategies designed to maximize learner engagement. For example, instructors can implement structured reflective journals in which students document their technology use patterns and evaluate how different tools contribute to their learning; they can also design peer-observation activities where students share effective technology strategies with classmates. These engagement-focused pedagogical practices can enhance the impact of existing technology resources without requiring additional financial investment, which is particularly valuable under the resource constraints faced by many interpreting programs.

5.3 Limitations and Future Directions

Several limitations should be considered when interpreting these findings. First, the cross-sectional design precludes causal inferences. Although the mediation model was theory-driven and the direction of effects was hypothesized a priori, causal relationships cannot be conclusively established without longitudinal or experimental data. Additionally, the present study tested a single partial mediation model without comparing it against theoretically plausible alternative specifications, such as full mediation or reversed mediation. Future research should employ longitudinal designs with multiple waves of data collection, experimental studies that manipulate specific technology features, and model comparison techniques to determine which theoretical model best fits the data. Second, the sample comprised non-English major undergraduates from STEAM programs in Beijing universities, which limits generalizability to other learner populations, proficiency levels, and educational contexts. The study also did not examine potential moderators such as prior technology experience, self-regulated learning skills, or language proficiency, which may influence the strength of the mediated pathways. Cross-institutional and cross-cultural comparisons are needed to establish the generalizability of the mediation model. Third, the ILO measure relied primarily on self-reported indicators rather than objective measures of interpreting performance. While the measures demonstrated acceptable reliability and validity, self-report data may introduce shared method variance and do not fully capture actual interpreting competence. Although the Harman single-factor test (44.2%), ULMC analysis, and HTMT assessment collectively supported the adequacy of the measurement model, future studies should incorporate more objective performance assessments—such as rater-scored interpreting tests—to reduce reliance on self-report data. Future research should also cross-validate the factor structure across different learner populations to ensure measurement robustness.

Future research should employ longitudinal designs to track how the relationships among MTR, LE, and ILO evolve over the course of interpreter training. Experimental studies that manipulate specific technology features could clarify which aspects of multimodal tools most effectively drive engagement and learning outcomes. Additionally, comparative studies across different language pairs, interpreting modalities (consecutive vs. simultaneous), and instructional contexts (face-to-face vs. online) would extend the generalizability of the model. Finally, investigating the role of teacher scaffolding and peer interaction in conjunction with technology use could provide a richer picture of the social and contextual factors that shape engagement in technology-enhanced interpreting classrooms.

6. Conclusion

The present study contributes to the growing body of research on technology-enhanced interpreter education by demonstrating that learning engagement serves as a significant indirect pathway linking multimodal technology reliance to English-Chinese interpreting learning outcomes. Using structural equation modeling with 312 non-English major undergraduates, the results showed that MTR predicted ILO both directly ($\beta = .16$, $p < .05$) and indirectly through LE ($\beta = .21$, $p < .01$), with the indirect path accounting for 56.8% of the total effect. The model explained 32.4% of the variance in LE and 28.7% of the variance in ILO. These findings suggest that the primary mechanism through which multimodal technologies support interpreting learning is by strengthening learner engagement rather than by improving cognitive processing alone. The concept of multimodal technology reliance, grounded in social cognitive theory, reflects learners' self-efficacy beliefs regarding technology use and their outcome expectations for interpreting skill development (Bandura, 1986). This perspective is particularly relevant in EFL contexts where students may depend on multimodal resources to bridge the gap between classroom learning and real-world interpreting demands. By providing empirical support for engagement as a statistically supported indirect pathway, the study offers a theoretically grounded framework that can guide both future research and pedagogical practice.

The findings offer actionable guidance for interpreter educators and curriculum designers. The prominence of the engagement pathway suggests that technology integration strategies should prioritize features that foster active learner involvement rather than merely providing access to digital tools. For example, video captioning tools should be paired with interactive comprehension checks; speech recognition software should be used for self-monitoring exercises that promote reflective practice; and multimodal materials should be structured to encourage cross-modal comparison and deeper cognitive processing. Furthermore, the mediating role of engagement implies that interpreting programs should consider engagement as a key outcome metric rather than focusing solely on technology adoption rates or user satisfaction. Incorporating structured self-reflection exercises, personalized feedback on engagement patterns, and peer accountability structures may be particularly valuable for non-English major students who may have limited intrinsic motivation for interpreting skill development and benefit from structured engagement supports.

Building on the findings of the present study, several directions for future research emerge. Longitudinal designs with multiple waves of data collection would allow researchers to track how the relationships among MTR, LE, and ILO evolve over the course of interpreter training and to examine the potential reciprocal effects among these variables over time. Experimental and quasi-experimental studies that manipulate specific technology features—such as captioning formats, feedback modalities, or practice structures—would clarify which aspects of multimodal tools most effectively drive engagement and performance gains. Cross-institutional and cross-cultural comparisons, extending beyond STEAM programs in Beijing universities, would test the generalizability of the mediation model across diverse learner populations and educational contexts.

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 prior to 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. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

[1] Bandura, A. (1986). *Social foundations of thought and action: A social cognitive theory*. Prentice Hall.

[2] Byrne, B. M. (2016). *Structural equation modeling with AMOS: Basic concepts, applications, and programming* (3rd ed.). Routledge.

[3] Cai, R., Lin, J., & Dong, Y. (2023). Psychological factors and interpreting competence in interpreting students: A developmental study. *The Interpreter and Translator Trainer*, *17*(2), 246-263. https://doi.org/10.1080/1750399X.2023.2182590

[4] Chan, V. (2023). Research on computer-assisted interpreter training: A review of studies from 2013 to 2023. *SN Computer Science*, *4*(5), Article 648. https://doi.org/10.1007/s42979-023-02072-w

[5] Chen, P. S., Lambert, A. D., & Guidry, K. R. (2010). Engaging online learners: The impact of web-based learning technology on college student engagement. *Computers & Education*, *54*(4), 1222–1232. https://doi.org/10.1016/j.compedu.2009.11.008

[6] Chiu, T. K. F. (2021). Digital support for student engagement in blended learning based on self-determination theory. *Computers in Human Behavior*, *124*, Article 106909. https://doi.org/10.1016/j.chb.2021.106909

[7] Curran, P. J., West, S. G., & Finch, J. F. (1996). The robustness of test statistics to nonnormality and specification error in confirmatory factor analysis. *Psychological Methods*, *1*(1), 16-29. https://doi.org/10.1037/1082-989X.1.1.16

[8] Dai, K., & Wang, Y. (2024). Enjoyable, anxious, or bored? Investigating Chinese EFL learners' classroom emotions and their engagement in technology-based EMI classrooms. *System*, *123*, Article 103339. https://doi.org/10.1016/j.system.2024.103339

[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] Deci, E. L., & Ryan, R. M. (1985). *Intrinsic motivation and self-determination in human behavior*. Springer. https://doi.org/10.1007/978-1-4899-2271-7

[11] Fredricks, J. A., Blumenfeld, P. C., & Paris, A. H. (2004). School engagement: Potential of the concept, state of the evidence. *Review of Educational Research*, *74*(1), 59–109. https://doi.org/10.3102/00346543074001059

[12] Fredricks, J. A., Filsecker, M., & Lawson, M. A. (2016). Student engagement, context, and adjustment: Addressing definitional, measurement, and methodological issues. *Learning and Instruction*, *43*, 1–4. https://doi.org/10.1016/j.learninstruc.2016.02.002

[13] Gile, D. (2009). *Basic concepts and models for interpreter and translator training (Rev. ed.)*. John Benjamins. https://doi.org/10.1075/btl.8

[14] Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2018). *Multivariate data analysis* (8th ed.). Cengage Learning.

[15] Henrie, C. R., Halverson, L. R., & Graham, C. R. (2015). Measuring student engagement in technology-mediated learning: A review. *Computers & Education*, *90*, 36–53. https://doi.org/10.1016/j.compedu.2015.09.005

[16] Hiver, P., Al-Hoorie, A. H., & Mercer, S. (Eds.). (2021). *Student engagement in the language classroom*. Multilingual Matters.

[17] 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

[18] Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, *6*(1), 1-55. https://doi.org/10.1080/10705519909540118

[19] Kimmons, R., Veletsianos, G., & Trust, T. (2025). *Judicious AI use to improve existing OER*. *The AI + Open Education Initiative*, MIT Open Learning. https://aiopeneducation.pubpub.org/pub/bl09sl1d

[20] Kline, R. B. (2015). *Principles and practice of structural equation modeling* (4th ed.). Guilford Press.

[21] Lai, C. (2019). Learning beliefs and autonomous language learning with technology beyond the classroom. *Language Awareness*, *28*(4), 291-309. https://doi.org/10.1080/09658416.2019.1675679

[22] Liu, Y., & Liang, J. (2024). Multidimensional comparison of Chinese-English interpreting outputs from human and machine: Implications for interpreting education in the machine-translation age. *Linguistics and Education*, *80*, Article 101273. https://doi.org/10.1016/j.linged.2024.101273

[23] MacKinnon, D. P., Lockwood, C. M., & Williams, J. (2004). Confidence limits for the indirect effect: Distribution of the product and resampling methods. *Multivariate Behavioral Research*, *39*(1), 99-128. https://doi.org/10.1207/s15327906mbr3901_4

[24] Manwaring, K. C., Larsen, R., Graham, C. R., Henrie, C. R., & Halverson, L. R. (2017). Investigating student engagement in blended learning settings using experience sampling and structural equation modeling. *The Internet and Higher Education*, *35*, 21–33. https://doi.org/10.1016/j.iheduc.2017.06.002

[25] Mayer, R. E. (2014). Cognitive theory of multimedia learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (2nd ed., pp. 43-71). Cambridge University Press. https://doi.org/10.1017/CBO9781139547369.005

[26] Mayer, R. E., Lee, H., & Peebles, A. (2014). Multimedia learning in a second language: A cognitive load perspective. *Applied Cognitive Psychology*, *28*(5), 653–660. https://doi.org/10.1002/acp.3050

[27] 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

[28] Moreno, R., & Mayer, R. (2007). Interactive multimodal learning environments. *Educational Psychology Review*, *19*(3), 309–326. https://doi.org/10.1007/s10648-007-9047-2
[29] Pöchhacker, F. (2016). *Introducing interpreting studies* (2nd ed.). Routledge.
[30] Podsakoff, P. M., MacKenzie, S. B., Lee, J.-Y., & Podsakoff, N. P. (2003). Common method biases in behavioral research: A critical review of the literature and recommended remedies. *Journal of Applied Psychology*, *88*(5), 879-903. https://doi.org/10.1037/0021-9010.88.5.879
[31] Ramezanali, N., Uchihara, T., & Faez, F. (2021). Efficacy of multimodal glossing on second language vocabulary learning: A meta-analysis. *TESOL Quarterly*, *55*(1), 105–133. https://doi.org/10.1002/tesq.579
[32] Ryan, R. M., & Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist*, *55*(1), 68–78. https://doi.org/10.1037/0003-066X.55.1.68
[33] Schindler, L. A., Burkholder, G. J., Morad, O. A., & Marsh, C. (2017). Computer-based technology and student engagement: A critical review of the literature. *International Journal of Educational Technology in Higher Education*, *14*(1), Article 25. https://doi.org/10.1186/s41239-017-0063-0
[34] Sweller, J. (2011). Cognitive load theory. In J. Mestre & B. Ross (Eds.), *The psychology of learning and motivation* (Vol. 55, pp. 37-76). Academic Press. https://doi.org/10.1016/B978-0-12-387691-1.00002-8
[35] Teng, M. F. (2022). Incidental L2 vocabulary learning from viewing captioned videos: Effects of learner-related factors. *System*, *105*, Article 102736. https://doi.org/10.1016/j.system.2022.102736
[36] Tian, S., & Yang, W. (2024). Modeling the use behavior of interpreting technology for student interpreters: An extension of UTAUT model. *Education and Information Technologies*, *29*(9), 10499–10528. https://doi.org/10.1007/s10639-023-12225-2
[37] Venkatesh, V., Morris, M. G., Davis, G. B., & Davis, F. D. (2003). User acceptance of information technology: Toward a unified view. *MIS Quarterly*, *27*(3), 425–478. https://doi.org/10.2307/30036540
[38] Wang, Y. (2026). Exploring student engagement with multimodal generative AI in task-based Chinese language learning. *Computer Assisted Language Learning*. *Computer Assisted Language Learning*, 1–26. https://doi.org/10.1080/09588221.2026.2645428
[39] Yang, Y., & Wang, X. (2019). Modeling the intention to use machine translation for student translators: An extension of Technology Acceptance Model. *Computers & Education*, *133*, 116–126. https://doi.org/10.1016/j.compedu.2019.01.015
[40] Yang, S., Li, D., & Lai, V. L. C. (2025). Multimodal processing in simultaneous interpreting with text: Evidence from ear-eye-voice span and performance. *PLoS ONE*, *20*(7), Article e0326527. https://doi.org/10.1371/journal.pone.0326527