English Linguistics Research; Vol. 15, No. 2; 2026
ISSN 1927-6028 E-ISSN 1927-6036
Published by Sciedu Press

Feedback Actionability in AI-Mediated L2 Interaction: A Repeated-Session Study of Learner Uptake

Jin Tang[1]

[1] School of Foreign Languages, Hubei University of Science and Technology, Xianning, Hubei 437100, China

Correspondence: Jin Tang, School of Foreign Languages, Hubei University of Science and Technology, Xianning, Hubei 437100, China. E-mail: tang-jin@hotmail.com

Received: October 6, 2026 Accepted: November 10, 2026 Online Published: November 16, 2026

doi:10.5430/elr.v15n2p79 URL: https://doi.org/10.5430/elr.v15n2p79

Abstract

AI systems can respond to nearly every learner turn, but how learners use that feedback during an exchange is less well documented. This study followed 48 university EFL learners across eight text-based classroom sessions using GPT-4o (384 learner-session observations), examining participation, words per turn (WPT), and researcher-coded next-turn uptake. Following each session, learners submitted ChatGPT-generated summaries of their messages, while researchers separately coded immediate revisions from the interaction records. Mean learner turns increased from 12.3 to 14.5, WPT from 29.3 to 34.6, and mean learner-session uptake from 0.36 to 0.68. Mixed-effects models showed positive session associations with WPT ($B = 0.663$, $p < .001$) and uptake (logit $B = 0.181$, $p < .001$). After accounting for the session trend, a supplementary model did not identify a distinct phase-level uptake difference following the midpoint prompt revision ($B = 0.153$, $p = .327$). The study documents increasing immediate use of AI feedback during repeated interaction, although the fixed phase order prevents separating prompt refinement from accumulated classroom experience.

Keywords: actionable feedback, artificial intelligence-mediated feedback, classroom second language interaction, feedback uptake, interactional processes, language learning, second language acquisition

1. Introduction

Feedback can support second language (L2) development when learners attend to it and use it in later production (Long, 1996; Mackey, 2012). In classroom interaction, however, a correction must be processed while the task is still moving forward. Feedback available on screen may remain difficult to act on before the next turn. This issue is especially relevant to artificial intelligence (AI)-mediated interaction, where a conversational system can respond after almost every learner turn. Frequent provision makes it necessary to examine whether learners recognise, interpret, and use feedback within the ongoing exchange.

Research on conversational agents and generative AI documents opportunities for language practice and immediate responses (Du & Daniel, 2024; Huang et al., 2022; Kohnke et al., 2023; Lai & Lee, 2024). Intervention studies have reported gains in speaking performance and willingness to communicate (Fathi et al., 2024; Tai & Chen, 2024). Shadiev et al. (2026) examined speech recognition and automatic corrective feedback in telecollaboration, reporting gains in speaking skills and confidence. Outcomes vary, however: Wang et al. (2024) reported favourable changes in willingness to communicate, anxiety, and perceived competence in one chatbot condition, while between-group differences in speaking performance were not statistically significant. Kim (2024) examined learners' responses to chatbot recasts during task completion. These studies address varied systems and outcomes. Pre–post comparisons and learner perceptions offer limited detail on how feedback use changes from one turn or session to the next. Repeated observation of the same learners can trace those changes, although it still requires care in distinguishing session progression from refinements to the feedback design.

Research on technology-supported learning emphasises the value of tracing learner activity over time (Azevedo, 2014; Winne, 2017). In L2 research, uptake is an observable response to corrective feedback, distinct from durable acquisition (Lyster & Ranta, 1997). The present study uses a narrower, repair-based measure: a revision of at least one feature addressed by an AI feedback response in the next learner turn. Words per turn provides a complementary index of turn-level elaboration, while learner turns indicate participation frequency. These indicators distinguish how

often learners participate from how they elaborate their turns and respond to feedback.

The study adopts actionability as an analytical perspective on whether feedback can inform a feasible next action. Its usability depends on the learner and task as well as the message. Feedback density, revision priorities, and alignment with the preceding turn may matter, alongside familiarity with the interaction format. The present design traces observable feedback use over repeated sessions; it does not measure actionability directly or isolate the contribution of these possible influences.

This study follows 48 learners across eight text-based, AI-mediated classroom sessions. It examines participation, turn-level elaboration, and immediate feedback uptake, alongside a later refinement of the feedback prompt. Two research questions guide the study: (RQ1) How do learner participation, turn-level elaboration, and observable feedback uptake change across repeated AI-mediated L2 interaction? (RQ2) Is the later feedback refinement associated with a further change in the level or slope of feedback uptake beyond the overall session trend? These questions concern change and association within a fixed instructional sequence, not causal intervention effects.

2. Literature Review

2.1 Feedback Provision and Learner Uptake

Interactionist accounts link L2 learning opportunities to input, feedback, attention, and modified output (Long, 1996; Mackey, 2012). Classroom research examined learners' immediate responses to corrective feedback through uptake (Lyster & Ranta, 1997). That category includes responses with successful repair and responses that still need repair. Feedback provision, learner response, and successful revision therefore represent distinct points in the exchange.

An accurate correction may receive no observable response, and a response may occur without deeper understanding. The present coding focuses on revision in the next learner turn, providing a narrower indicator of immediate feedback use than the broader uptake category in Lyster and Ranta (1997). Research on synchronous text-based computer-mediated communication provides a relevant precedent: Smith (2005) examined negotiated interaction, uptake, and lexical acquisition, while Sauro (2009) studied corrective feedback in text-chat. These studies offer a useful comparison for investigating responses to AI feedback, although the interaction partners and research contexts differ.

The distinction also matters in AI-mediated learning. Kim (2024) found that successful repair after automatic chatbot recasts was positively associated with immediate and delayed post-test scores for the target construction. In AI-assisted writing, learners evaluate feedback and accept, reject, or ignore suggestions (Chen et al., 2026; Yan & Zhang, 2024). Learners' reported preferences for specific feedback and clear explanations have also been examined in automated written corrective feedback (Elmotri et al., 2025). These writing studies concern different tasks and systems, but they illustrate why provision alone is insufficient evidence of use. The present study traces immediate revision across repeated conversational sessions.

2.2 AI-Mediated Feedback in L2 Interaction

AI-mediated L2 research has examined a wide range of systems and outcomes. Reviews cover rule-based chatbots, speech technologies, and generative tools (Du & Daniel, 2024; Huang et al., 2022; Lai & Lee, 2024; Lee & Lee, 2024). Studies of AI-supported practice have assessed speaking performance (Fathi et al., 2024; Tai & Chen, 2024), learner perceptions and task experiences (Alsalem, 2024; Jeon, 2024), and affective responses. Wang et al. (2024), for example, reported changes in willingness to communicate, anxiety, and perceived competence but no statistically significant between-group difference in speaking performance. Other settings include task-based pragmatics role-play (Sim et al., 2025) and self-directed practice (Dizon, 2026). These varied outcomes establish the breadth of AI-supported learning, but reveal less about the learner's immediate response to a particular correction.

AI-assisted writing research provides a closer view of feedback decisions: learners may incorporate, modify, or disregard suggested changes (Chen et al., 2026; Yan & Zhang, 2024). In a conversation, those decisions must be made while the learner also prepares the next message. The following turn is therefore a useful point at which to observe whether a suggested revision enters the exchange.

Immediate feedback creates both an opportunity and a competing demand on attention. A correction may be closely connected to the preceding message, yet the learner must still decide how to respond and continue the task. Chen et al. (2026) documented learners' reported difficulties with vague, repetitive, or overloaded AI writing feedback. Whether comparable demands shape next-turn revision in conversation requires direct observation of the interaction.

2.3 From Feedback Availability to Actionability

Here, actionability concerns whether feedback offers a revision that the learner can use in the ongoing exchange. Ecological accounts treat affordances as relations between an actor and an environment, rather than properties of an

object alone (Gibson, 1979; van Lier, 2004). Applied to feedback, this view draws attention to the production–feedback–response sequence instead of the correction message in isolation. Actionability serves as an interpretive perspective in this study, not an independently measured construct.

Three aspects of feedback design are relevant to that sequence: density (how many issues a response raises), prioritisation (which issue is foregrounded), and local alignment (how directly a suggestion addresses the preceding learner turn). Research on feedback content and usefulness (Hattie & Timperley, 2007; Shute, 2008; Wisniewski et al., 2020) and on processing demands in instruction (Paas & van Merriënboer, 1994) offers reasons to examine these features. It does not establish which combination, if any, improves uptake in AI conversation.

Usability also depends on the learner. Students interpret feedback, judge its relevance, and decide whether to act on it (Carless & Boud, 2018; Han & Hyland, 2015). Across repeated sessions, learners may become more familiar with the system's responses and with revising while a conversation continues. Such adaptation provides one possible explanation for changes in feedback use over time.

2.4 Repeated Interaction and Longitudinal Change

Repeated sessions allow changes in feedback use to be observed as learners gain experience with an interaction format. Work on technology-supported learning stresses sequential activity and trace data as sources of evidence (Azevedo, 2014; Winne, 2017). In a longitudinal conversation-analytic case study of one learner, Choi and Oh (2026) documented changes in L2 turn-taking with ChatGPT. That case concerns conversational practices rather than feedback uptake, but it shows the value of examining exchanges over time. The present study follows uptake across a larger classroom sequence.

A refinement introduced midway through a classroom sequence complicates interpretation because a later difference may coincide with an ongoing session trend. The refinement after Session 4 was part of iterative classroom practice, without random assignment. RQ2 therefore concerns possible additional level and slope differences beyond the primary linear trend, neither of which can be interpreted as a causal effect of the prompt change.

3. Method

3.1 Participants and Context

Participants were 48 undergraduates enrolled in a compulsory English course at a university in central China. All were non-English majors, drawn from intact classes within the regular curriculum.

Each learner completed eight classroom sessions of AI-mediated interaction, giving 384 session-level observations (48 learners × 8 sessions). Learners interacted with the AI system in writing during these sessions. The study was reviewed and approved by the relevant institutional ethics committee (Approval No. 202503019), and informed consent was obtained from all participants before data collection.

3.2 AI-Mediated Interaction Tasks

Each session followed a recurring production–feedback–response format using OpenAI's GPT-4o model through the ChatGPT interface in June 2025. At the start of the session, learners pasted the teacher-assigned prompt template into the chat interface and then interacted individually with the system in written English. The same classroom procedure was used across the eight sessions.

Learners were asked to produce at least eight turns per session. At the end of the exchange, they requested a closing summary and then entered a standardised instruction asking ChatGPT for a JSON-formatted summary. The requested fields included student_turns (number of student-sent messages), avg_words_per_turn (average English words per student message), and corrections_total (AI corrective-feedback count), together with feedback-type counts and examples. Learners submitted the JSON output through a designated form. The instructor introduced the task and managed classroom time but did not provide turn-level corrective feedback. The classroom guide specified an approximately six-to-seven-minute interaction target; this was not a recorded measure of active typing time. Accordingly, typing speed was not estimated.

3.3 Feedback Design Across the Two Phases

The eight sessions formed two consecutive feedback phases. In Sessions 1–4, the AI was instructed to respond to grammar, word choice, and expression and continue the conversation. In Sessions 5–8, the prompt was revised to prioritise one or two issues, provide a concise improved version, and align the feedback more closely with the preceding learner turn; the tone was also adjusted. The preserved teacher prompt materials specify the timing of correction, the amount of explanation, and the tone of feedback. For example, one immediate-feedback module

stated: “Please correct my errors immediately after each of my responses and give me suggestions for improvement.” Other modules requested either a brief correction or fuller explanations with examples, using neutral or encouraging language. These examples illustrate the original prompt wording; the present analysis retains the phase-level comparison rather than estimating separate effects of individual prompt components. The same broad task format and minimum turn requirement were used throughout.

The refinement aimed to make feedback easier to use. Phase 2 bundled changes in density, prioritisation, local alignment, provision of an improved version, and tone. All learners experienced Phase 1 before Phase 2, with no random assignment. Consequently, a phase contrast cannot isolate the effect of any individual feedback feature from cumulative experience, session order, or task-related change.

3.4 Measures and Coding

Three interactional indicators were analysed. At the end of each session, students submitted a ChatGPT-generated JSON summary requested under a standardised instruction. The instruction defined student_turns as the number of messages sent by the student and avg_words_per_turn as the average English words per student message. The retained learner-session dataset contained a turn-count field (turns) and a learner-word total (learner_words); WPT for the present analysis was calculated as learner_words divided by turns. A learner turn therefore refers to a student-sent message as counted in the session summary, rather than an entire learner–AI exchange. Turns indexed participation frequency and WPT indexed contribution length; neither indicator was interpreted as evidence of language proficiency gains.

Feedback uptake was coded separately by researchers using the saved interaction transcripts. A feedback instance was one AI feedback response represented in the uptake coding, irrespective of the number of suggestions in that response. Observable uptake was recorded when the next learner turn revised at least one feature targeted by the feedback. For each learner-session, the researcher-coded revision count was divided by the researcher-coded feedback-instance count to calculate the uptake proportion. The corrective-feedback count returned in the ChatGPT JSON summary was a separate descriptive item; it was not used as the denominator in the uptake model. Feedback instances and learner turns therefore represent different units.

To assess coding consistency, approximately 20% of the material was independently double-coded. The original coding summary reported 91% simple percentage agreement across the double-coded material. A chance-corrected agreement coefficient was not available from the retained summary records. Uptake is interpreted as observable next-turn revision, not a direct measure of lasting language learning.

3.5 Data Analysis

Analyses used 384 learner-session observations from 48 learners (Baayen et al., 2008). Learner turns were described by session rather than fitted as a primary outcome. WPT was fitted with a linear mixed-effects model (LMM). For uptake, the recorded revision count served as binomial successes and the feedback-instance count as trials in a logit-link generalized linear mixed-effects model (GLMM). Mean uptake proportions in Table 1 are unweighted averages of the 48 learner-session proportions at each session, whereas the GLMM uses the underlying counts.

Primary models included Session (1–8) as a continuous predictor and a learner-specific random intercept. Supplementary models added Phase (0 = Sessions 1–4; 1 = Sessions 5–8). For uptake, an additional segmented model included an indicator of the level change at Session 5 and a post-transition slope-change term; thus, it tested both possible departures from the overall session trend. Model comparisons used Akaike information criterion (AIC). Session was coded 1–8 without centering in all reported coefficient tables; the intercept therefore refers to an extrapolated Session 0, not the observed Session 1. Fixed-effect estimates are reported with standard errors, t or z statistics, 95% confidence intervals, and p values.

Models were re-estimated from the retained session-level data using Python 3.13, with statsmodels 0.14.6 for linear mixed-effects models and numerical maximum likelihood with Gauss–Hermite quadrature for binomial mixed-effects models. Marginal and conditional R^2 were computed on a latent-variable scale as the fixed-effect variance divided by the total model variance, or fixed plus random-effect variance divided by the same total, respectively. For the logit GLMM, the distribution-specific variance was $\pi^2/3$; these estimates are not interchangeable with R^2 values obtained under other definitions. For visualisation, population-averaged fitted uptake probabilities were obtained by integrating over the estimated normal learner intercepts, and their 95% confidence intervals were estimated by simulating from the fitted parameter covariance matrix.

For WPT, sensitivity checks included a learner-specific session random slope (Barr et al., 2013) and a quadratic session term. The phase analyses were exploratory because all learners encountered the two phases in the same order;

non-significant estimates were not treated as evidence of equivalence or no effect.

4. Results

4.1 Descriptive Changes Across Sessions

All 48 learners contributed eight session observations, with no missing learner-sessions. Mean learner turns increased from 12.31 (SD = 2.45) in Session 1 to 14.48 (SD = 2.32) in Session 8 (about 18%) and were described rather than modelled. Over the same period, WPT rose from 29.33 to 34.60, and mean learner-session uptake increased from 0.36 to 0.68. Recorded feedback-instance counts ranged from 5 to 11 per learner-session, with no zero-denominator observations. Table 1 reports the session means, standard deviations, and mean coded feedback-instance counts.

WPT rose early, remained broadly similar in Sessions 5–6, and increased again in later sessions. Uptake showed a larger observed increase between Sessions 4 and 5, followed by a small decline in Session 6 and further increases. These observations do not establish whether time, task familiarity, or the combined prompt refinement explains the change.

Table 1. Descriptive Statistics of Interactional Processes Across Sessions

Session	Turns (M, SD)	Words per Turn (M, SD)	Uptake (M, SD)	Feedback instances (M)
1	12.3 (2.5)	29.3 (5.1)	0.36 (0.12)	6.52
2	12.4 (2.5)	31.2 (4.4)	0.41 (0.13)	6.54
3	13.0 (2.5)	32.7 (4.6)	0.45 (0.11)	7.42
4	13.4 (2.4)	33.0 (4.5)	0.48 (0.11)	7.52
5	13.6 (2.4)	33.4 (4.9)	0.58 (0.10)	7.44
6	14.3 (2.5)	33.3 (4.2)	0.58 (0.11)	8.31
7	14.4 (2.3)	34.5 (4.3)	0.61 (0.11)	8.40
8	14.5 (2.3)	34.6 (5.9)	0.68 (0.09)	8.48

Note. Values are learner-session means (M) and standard deviations (SD). Uptake is the unweighted mean of the 48 learner-session uptake proportions in each session; the binomial model uses the corresponding revision and feedback-instance counts. "Feedback instances" gives the mean number of researcher-coded AI feedback responses per learner-session, distinct from learner turns and the ChatGPT-reported corrective-feedback count.

4.2 Longitudinal Change in Words per Turn

WPT increased across sessions. The random-intercept LMM estimated a positive session association (B = 0.663, SE = 0.070, t = 9.51, 95% CI [0.527, 0.800], $p < .001$), or approximately 0.66 additional words per turn per session. Marginal R^2 was .093 and conditional R^2 was .604 under the stated variance decomposition. Allowing learner-specific session slopes yielded a similar estimate (B = 0.663, 95% CI [0.514, 0.813]). A supplementary quadratic model indicated modest curvature (Session²: B = −0.100, SE = 0.034, t = −2.90, p = .004; AIC = 2084.83 versus 2091.11 for the linear model), consistent with slower increases during later sessions. Figure 1 presents the observed means.

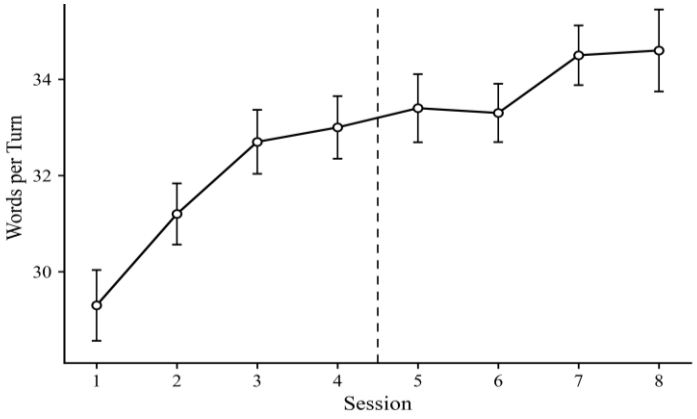


Figure 1. Observed Words per Turn Across Sessions

Note. Points are observed learner-session WPT means; error bars show ±1 standard error across the 48 learners per session. The dashed vertical line marks the transition between the two consecutive phases. These descriptive observations are not adjusted causal effects.

4.3 Longitudinal Change in Feedback Uptake

Observable feedback uptake increased across the eight sessions. The random-intercept binomial GLMM estimated a positive session association on the logit scale (B = 0.181, SE = 0.017, z = 10.55, 95% CI [0.147, 0.215], p < .001; odds ratio = 1.198 per session). Under the latent-variable variance definition specified above, marginal R^2 was .048 and conditional R^2 was .076. Figure 2 compares observed learner-session uptake means with population-averaged predictions and 95% confidence intervals calculated from the fitted model.

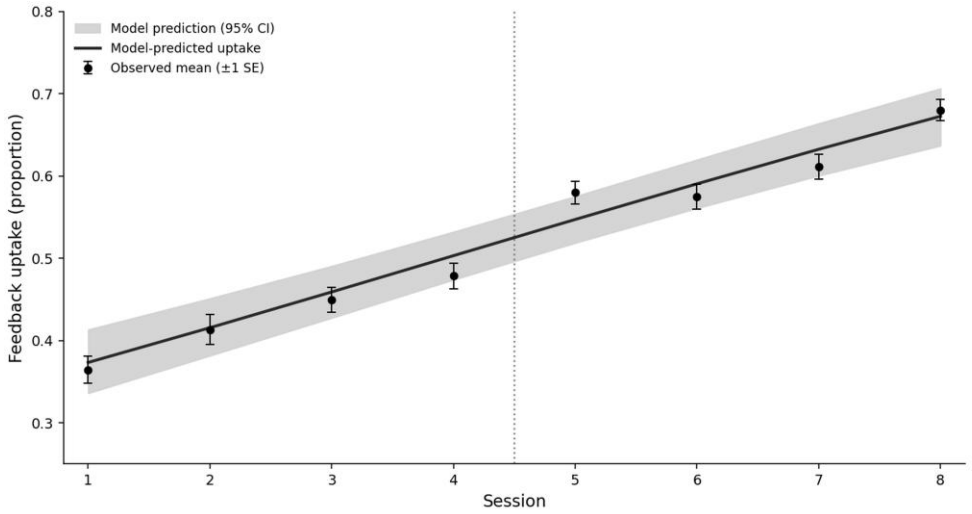


Figure 2. Observed Feedback Uptake and Model-Based Predictions Across Sessions

Note. Points show observed mean learner-session uptake proportions (±1 SE; 48 learners per session). The solid line shows population-averaged predictions from the fitted binomial GLMM, integrated over the learner random-intercept distribution; the shaded region shows a 95% confidence interval based on the fitted parameter covariance. The vertical dotted line separates Sessions 1–4 from Sessions 5–8. The curve is based on the fitted model, not rounded coefficients.

Table 2. Primary Session Effects From Random-Intercept Mixed-Effects Models

Outcome	Predictor	B (SE)	t/z	95% CI	p
WPT	Intercept	29.774 (0.623)	47.83	[28.554, 30.994]	< .001
WPT	Session	0.663 (0.070)	9.51	[0.527, 0.800]	< .001
Uptake (logit)	Intercept	−0.711 (0.100)	−7.10	[−0.907, −0.515]	< .001
Uptake (logit)	Session	0.181 (0.017)	10.55	[0.147, 0.215]	< .001

Note. Test statistics are t values for the WPT linear mixed-effects model (LMM) and z values for the uptake binomial logit generalized linear mixed-effects model (GLMM), computed from unrounded estimates and standard errors. Each model included 384 learner-session observations nested in 48 learners. AIC: WPT 2091.11; uptake 1054.87. Latent-variable marginal/conditional R^2: WPT .093/.604; uptake .048/.076. The two R^2 measures do not have an identical residual-variance interpretation across LMM and GLMM.

4.4 Supplementary Phase and Segmented Analyses

When both session and instructional phase were included, session remained associated with uptake (B = 0.151, SE = 0.035, 95% CI [0.084, 0.219], p < .001), whereas the additional phase level coefficient was imprecise (B = 0.153, SE = 0.156, 95% CI [−0.153, 0.458], p = .327). A supplementary WPT model likewise gave an uncertain phase coefficient (B = −1.096, SE = 0.652, 95% CI [−2.375, 0.183], p = .093), while the session coefficient remained positive. Full estimates are given in Table 3.

A segmented uptake model estimated the level change at Session 5 (B = 0.152, SE = 0.160, 95% CI [−0.162, 0.466], p = .343) and the post-transition change in slope (B = −0.002, SE = 0.069, 95% CI [−0.137, 0.134], p = .982). Adding phase parameters did not improve AIC relative to the primary session model (session-only 1054.87; session-plus-phase 1055.91; segmented 1057.91). Phase and session were strongly related by construction, and these wide confidence intervals do not establish absence of a phase effect. The estimates describe associations within the fixed classroom sequence, not effects attributable to the bundled refinement.

Table 3. Supplementary Session and Phase Models

Model	Predictor	B (SE)	t/z	95% CI	p
WPT + Phase	Session	0.872 (0.142)	6.13	[0.593, 1.151]	< .001
WPT + Phase	Phase	−1.096 (0.652)	−1.68	[−2.375, 0.183]	.093
Uptake + Phase	Session	0.151 (0.035)	4.37	[0.084, 0.219]	< .001
Uptake + Phase	Phase	0.153 (0.156)	0.98	[−0.153, 0.458]	.327
Uptake, segmented	Session	0.152 (0.050)	3.04	[0.054, 0.250]	.002
Uptake, segmented	Level at S5	0.152 (0.160)	0.95	[−0.162, 0.466]	.343
Uptake, segmented	Slope change	−0.002 (0.069)	−0.02	[−0.137, 0.134]	.982

Note. Test statistics are t values for WPT and z values for uptake, computed before display rounding. All models include learner random intercepts and Session coded 1–8. Phase is 0 for Sessions 1–4 and 1 for Sessions 5–8. Uptake uses binomial counts on the logit scale; "Level at S5" and "Slope change" are the segmented-model additions. WPT + Phase AIC = 2090.30; uptake + Phase AIC = 1055.91; uptake segmented AIC = 1057.91. Estimates are exploratory and are not causal effects.

5. Discussion

5.1 Participation, Output Length, and Recorded Uptake

RQ1 concerned changes in participation, output length, and observable uptake across repeated AI-mediated interaction. Learner turns increased from 12.3 to 14.5 (approximately 18%) and were analysed descriptively. Over the same sequence, WPT increased, and a larger proportion of coded feedback instances was followed by next-turn revision. Together, these measures describe complementary aspects of participation, turn elaboration, and immediate feedback use.

The post-session JSON summaries counted student-sent messages. Because the task required at least eight learner turns, the turn count reflects a participation requirement as well as what learners chose to contribute. Mean counts exceeded the minimum in the first session and increased by approximately 18% by the last. WPT and uptake had no equivalent minimum, giving a different view of how the interaction developed.

A longer message may also incorporate wording suggested by ChatGPT. WPT and uptake need not vary independently: revising an expression can lengthen a turn, and a longer turn can contain feedback-derived phrasing. This possible overlap matters when interpreting their parallel increases, although the measures capture different features of the recorded interaction.

5.2 Increasing Feedback Uptake Across Repeated Interaction

Uptake showed the clearest descriptive change. Feedback was available from Session 1, yet the mean proportion of instances followed by observable uptake rose from 0.36 to 0.68, and session significantly predicted uptake. The response remained the coding unit even when it contained several suggestions. The increase therefore means that learners revised at least one relevant feature after a larger share of AI responses. It gives evidence of more frequent immediate use at that response level, while leaving open how fully each response was used.

Increasing familiarity with the recurring exchange is one plausible explanation for the rising uptake pattern. Repeated practice may help learners recognise a correction, select a revision, and continue the task. Research on sequential learning activity provides a basis for considering such changes over time (Azevedo, 2014; Winne, 2017). Choi and Oh (2026) likewise documented changing conversational practices during repeated interaction with ChatGPT, offering a related perspective on adaptation to an AI interlocutor.

Evidence from chatbot recasts has linked successful repair to performance on a target construction (Kim, 2024), while AI writing studies document differences in which suggestions learners accept, reject, or ignore (Chen et al., 2026; Yan & Zhang, 2024). The present records show immediate revision; whether those revisions persist in later independent performance remains open.

5.3 Interpreting the Additional Phase Level Difference

The supplementary analyses addressed whether the later prompt revision coincided with an additional change beyond the overall session trend. After Session was included, the uptake phase coefficient was imprecise (B = 0.153, 95% CI [−0.153, 0.458]), as was the post-transition slope change in the segmented model (B = −0.002, 95% CI [−0.137, 0.134]). Mean uptake rose between Sessions 4 and 5, but these estimates did not identify a distinct adjusted level or slope change. The confidence intervals also leave room for differences that this study could not estimate precisely.

The revised prompt combined fewer correction demands, clearer priorities, closer alignment with the preceding

learner turn, and an adjustment in tone. These are plausible design choices in light of feedback research (Hattie & Timperley, 2007; Shute, 2008; Wisniewski et al., 2020) and reported difficulties with overloaded AI feedback (Chen et al., 2026). Yet every learner encountered the revised prompt after Session 4, making phase and session closely correlated ($r \approx .87$). The sequence cannot distinguish continued practice from the bundled refinement or identify the contribution of any individual prompt feature.

5.4 Actionability as a Process Perspective

Counting the corrections an AI system offers tells us little about what learners do with them. In this study, feedback was available throughout the eight sessions, but a growing share of coded feedback responses was followed by revision in the next learner turn. Following the production–feedback–response sequence makes that change visible. Actionability names the question of whether feedback can be put to use at that point in the exchange; it is not a separately validated mechanism. Future studies could examine individual feedback features alongside learners' revisions and later independent performance to identify when immediate use extends beyond the task.

6. Pedagogical Implications

In short text-chat tasks, feedback that raises several issues at once may be difficult to use before the next response. Teachers designing AI prompts might begin by specifying one priority issue, requesting a concise suggested revision, and allowing the conversation to continue. These are practical options for classroom trials; the present study did not isolate their individual effects.

Teachers can also review a short sequence of learner turns alongside the AI feedback rather than judging a prompt by the number of corrections it produces. When a suggestion repeatedly receives no observable revision, its specificity, relevance, or fit with the task may warrant attention. Looking at the learner's next message provides a concrete basis for revising the task or prompt during repeated practice.

7. Limitations

The findings should be interpreted in light of the study design and the scope of the measures. The phase-level prompt refinement was introduced at a fixed point in the eight-session sequence, making it difficult to distinguish its contribution from accumulated practice and other session-related changes. Moreover, the refinement involved several feedback features, whose individual effects could not be isolated in the present analysis.

The measures captured immediate text-based interaction rather than broader language development. Turns and words per turn described participation frequency and output length, while observable uptake reflected next-turn revisions following AI feedback, which may include the incorporation of AI-provided wording. These indicators therefore should not be interpreted as direct evidence of gains in spoken proficiency or sustained language learning. Future research could compare feedback configurations under counterbalanced conditions and combine interactional measures with independent assessments of language development.

8. Conclusion

Across eight classroom sessions, the same 48 EFL learners sent more messages, produced longer messages on average, and revised their next turns after a growing proportion of coded AI feedback responses. Mixed-effects models showed positive associations between session and both WPT and uptake. Although uptake rose around the midpoint, the supplementary analyses did not identify a distinct phase-level change after accounting for the overall session trend.

The study's contribution lies in following feedback through to the learner's subsequent response. The number of corrections offered would not, by itself, show whether learners used them during the exchange. Recording next-turn revisions alongside participation and output length provides a more detailed account of repeated AI-supported practice. These findings establish a behavioural pattern; determining the relative contributions of prompt design and practice, and whether immediate revisions carry over to independent performance, will require further evidence.

Acknowledgments

The author thanks the students and the course instructor who participated in this classroom-based study for their cooperation and engagement.

Author Contributions

Jin Tang (ORCID: https://orcid.org/0009-0003-6134-3234) was responsible for conceptualization, methodology, formal analysis, investigation, original draft preparation, review and editing, and project administration.

Consent for Publication

The author has read and approved the submitted version of the manuscript.

Funding

This research was supported by the 2025 Provincial Teaching Reform Research Project for Undergraduate Education in Hubei Province, Synergising Generative AI and Teacher Support to Foster L2 Flow Experience: A Practice-Oriented Study (Grant No. 2025439).

Declaration of Competing Interest

The author declares no conflict of interest.

Informed Consent

Informed consent was obtained from all participants.

Ethics Approval

This study was reviewed and approved by the Academic Ethics and Research Integrity Committee of Hubei University of Science and Technology (Approval No. 202503019).

Data Availability Statement

The de-identified learner-session data underlying the reported analyses are not publicly deposited but may be shared by the corresponding author upon reasonable request, subject to ethical and institutional requirements.

Use of Artificial Intelligence

GPT-4o was used as the AI interaction system in the study, and ChatGPT-generated session summaries formed part of the data collection procedure described in the Methods. During manuscript preparation, ChatGPT (OpenAI) was also used to assist with language editing, formatting, and checking statistical reporting. The author independently conducted and verified all literature selection, eligibility decisions, coding, analyses, interpretations, and final writing decisions. AI-generated suggestions were reviewed and validated by the author, who takes full responsibility for the accuracy and content of the manuscript.

Supplementary Materials

No supplementary materials are submitted with this manuscript.

Provenance and peer review

Not commissioned; externally double-blind peer reviewed.

Data sharing statement

Data sharing is limited to the de-identified learner-session dataset described above.

Open access

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (http://creativecommons.org/licenses/by/4.0/).

Copyrights

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.

References

Alsalem, M. S. (2024). EFL students' perception and attitude towards the use of ChatGPT to promote English speaking skills in the Saudi context. *Arab World English Journal, 15*(4), 73-84. https://doi.org/10.24093/awej/vol15no4.5

Azevedo, R. (2014). Issues in dealing with sequential and temporal characteristics of self- and socially-regulated learning. *Metacognition and Learning, 9*(2), 217-228. https://doi.org/10.1007/s11409-014-9123-1

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language, 59*(4), 390-412. https://doi.org/10.1016/j.jml.2007.12.005

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language, 68*(3), 255-278. https://doi.org/10.1016/j.jml.2012.11.001

Carless, D., & Boud, D. (2018). The development of student feedback literacy: Enabling uptake of feedback. *Assessment & Evaluation in Higher Education, 43*(8), 1315-1325. https://doi.org/10.1080/02602938.2018.1463354

Chen, Z., Zhu, X., Lu, Q., & Wei, W. (2026). L2 students' barriers in engaging with form and content-focused AI-generated feedback in revising their compositions. *Computer Assisted Language Learning, 39*(3), 715-735. https://doi.org/10.1080/09588221.2024.2422478

Choi, M. G., & Oh, S.-Y. (2026). Developing L2 turn-taking with ChatGPT: A longitudinal conversation analytic study. *System, 138*, Article 103959. https://doi.org/10.1016/j.system.2025.103959

Dizon, G. (2026). ChatGPT as a tool for self-directed foreign language learning. *Innovation in Language Learning and Teaching, 20*(2), 334-350. https://doi.org/10.1080/17501229.2024.2413406

Du, J., & Daniel, B. K. (2024). Transforming language education: A systematic review of AI-powered chatbots for English as a foreign language speaking practice. *Computers and Education: Artificial Intelligence, 6*, Article 100230. https://doi.org/10.1016/j.caeai.2024.100230

Elmotri, B., Harizi, R., Boujlida, A., Elyasa, Y. M., Garrouri, S. S., Amri, F., Malik, F. H., & Al-humari, M. A. (2025). The impact of AI-generated feedback explicitness (generic vs. specific) on EFL students' use of automated written corrective feedback. *Arab World English Journal, 16*(1), 384-402. https://doi.org/10.24093/awej/vol16no1.24

Fathi, J., Rahimi, M., & Derakhshan, A. (2024). Improving EFL learners' speaking skills and willingness to communicate via artificial intelligence-mediated interactions. *System, 121*, Article 103254. https://doi.org/10.1016/j.system.2024.103254

Gibson, J. J. (1979). The ecological approach to visual perception. Houghton Mifflin.

Han, Y., & Hyland, F. (2015). Exploring learner engagement with written corrective feedback in a Chinese tertiary EFL classroom. *Journal of Second Language Writing, 30*, 31-44. https://doi.org/10.1016/j.jslw.2015.08.002

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research, 77*(1), 81-112. https://doi.org/10.3102/003465430298487

Huang, W., Hew, K. F., & Fryer, L. K. (2022). Chatbots for language learning—Are they really useful? A systematic review of chatbot-supported language learning. *Journal of Computer Assisted Learning, 38*(1), 237-257. https://doi.org/10.1111/jcal.12610

Jeon, J. (2024). Exploring AI chatbot affordances in the EFL classroom: Young learners' experiences and perspectives. *Computer Assisted Language Learning, 37*(1-2), 1-26. https://doi.org/10.1080/09588221.2021.2021241

Kim, R. (2024). Effects of learner uptake following automatic corrective recast from artificial intelligence chatbots on the learning of English caused-motion construction. *Language Learning & Technology, 28*(2), 109-133. https://doi.org/10.64152/10125/73574

Kohnke, L., Moorhouse, B. L., & Zou, D. (2023). ChatGPT for language teaching and learning. *RELC Journal, 54*(2), 537-550. https://doi.org/10.1177/00336882231162868

Lai, W. Y. W., & Lee, J. S. (2024). A systematic review of conversational AI tools in ELT: Publication trends, tools, research methods, learning outcomes, and antecedents. *Computers and Education: Artificial Intelligence, 7*, Article 100291. https://doi.org/10.1016/j.caeai.2024.100291

Lee, H., & Lee, J. H. (2024). The effects of AI-guided individualized language learning: A meta-analysis. *Language Learning & Technology, 28*(2), 134-162. https://doi.org/10.64152/10125/73575

Long, M. H. (1996). The role of the linguistic environment in second language acquisition. In W. C. Ritchie & T. K. Bhatia (Eds.), *Handbook of second language acquisition* (pp. 413-468). Academic Press. https://doi.org/10.1016/B978-012589042-7/50015-3

Lyster, R., & Ranta, L. (1997). Corrective feedback and learner uptake: Negotiation of form in communicative classrooms. *Studies in Second Language Acquisition, 19*(1), 37-66. https://doi.org/10.1017/S0272263197001034

Mackey, A. (2012). Input, interaction, and corrective feedback in L2 learning. Oxford University Press.

Paas, F. G. W. C., & van Merriënboer, J. J. G. (1994). Variability of worked examples and transfer of geometrical problem-solving skills: A cognitive-load approach. *Journal of Educational Psychology, 86*(1), 122-133. https://doi.org/10.1037/0022-0663.86.1.122

Sauro, S. (2009). Computer-mediated corrective feedback and the development of L2 grammar. *Language Learning & Technology, 13*(1), 96-120. https://doi.org/10.64152/10125/44170

Shadiev, R., Feng, Y., Zhussupova, R., & Altinay, F. (2026). Effects of speech-enabled corrective feedback technology on EFL speaking skills, anxiety and confidence. *Computer Assisted Language Learning, 39*(4), 878-914. https://doi.org/10.1080/09588221.2024.2430761

Shute, V. J. (2008). Focus on formative feedback. *Review of Educational Research, 78*(1), 153-189. https://doi.org/10.3102/0034654307313795

Sim, S., Kim, Y., & Ku, K. (2025). Exploring generative AI as a roleplay interlocutor in L2 task-based pragmatics learning: Comparing ChatGPT-learner and technology-mediated peer interactions. *TESOL Quarterly, 59*(S1), S86-S116. https://doi.org/10.1002/tesq.70010

Smith, B. (2005). The relationship between negotiated interaction, learner uptake, and lexical acquisition in task-based computer-mediated communication. *TESOL Quarterly, 39*(1), 33-58. https://doi.org/10.2307/3588451

Tai, T. Y., & Chen, H. H. J. (2024). Navigating elementary EFL speaking skills with generative AI chatbots: Exploring individual and paired interactions. *Computers & Education, 220*, Article 105112. https://doi.org/10.1016/j.compedu.2024.105112

van Lier, L. (2004). *The ecology and semiotics of language learning: A sociocultural perspective*. Kluwer Academic Publishers. https://doi.org/10.1007/1-4020-7912-5

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*, Article 103533. https://doi.org/10.1016/j.system.2024.103533

Winne, P. H. (2017). Learning analytics for self-regulated learning. In C. Lang, G. Siemens, A. F. Wise, & D. Gašević (Eds.), *Handbook of learning analytics* (pp. 241-249). Society for Learning Analytics Research. https://doi.org/10.18608/hla17.021

Wisniewski, B., Zierer, K., & Hattie, J. (2020). The power of feedback revisited: A meta-analysis of educational feedback research. *Frontiers in Psychology, 10*, Article 3087. https://doi.org/10.3389/fpsyg.2019.03087

Yan, D., & Zhang, S. (2024). L2 writer engagement with automated written corrective feedback provided by ChatGPT: A mixed-method multiple case study. *Humanities and Social Sciences Communications, 11*, Article 1086. https://doi.org/10.1057/s41599-024-03543-y