

Learners' Perceptions of AI and Teacher-Led Speaking Feedback

Mohsen Askari^{1*} 

¹*School of Foreign Languages, Boğaziçi University, Istanbul, Türkiye*
mohsen.askari@bogazici.edu.tr

Azam Samadi Rahim² 

²*School of Foreign Languages, Boğaziçi University, Istanbul, Türkiye*
azam.rahim@bogazici.edu.tr

Received: 27/04/2026

Accepted: 27/05/2026

Published: 30/06/2026

***Corresponding author**

Citation:

Askari, M., & Rahim, A. S.
(2026). Learners'
Perceptions of AI and
Teacher-Led Speaking
Feedback. *Studies in
Linguistics and Language
Education*, 2(1), 240–255.

ABSTRACT: Artificial intelligence (AI) tools for speaking practice are increasingly used in language classrooms, raising questions about how such feedback is experienced relative to teacher-led practice. This convergent mixed-methods study examined post-course learner perceptions of AI-mediated and teacher-led speaking feedback in a university English speaking course. Fifty-three students at an English-medium university completed an end-of-course survey after eight weeks of either AI-mediated feedback (FlowSpeak) or teacher-led feedback. Four constructs were examined: competence, relatedness, expectancy, and cost. Quantitative analyses ($N = 47$) showed no statistically significant between-group differences on competence, relatedness, expectancy, or cost, although descriptively the AI group reported somewhat higher competence and expectancy and lower cost. Given the modest sample size, posttest-only design, and insufficient statistical power for small-to-medium effects, these patterns are interpreted as inconclusive rather than as evidence of equivalence. Thematic coding of brief open-ended comments ($N = 48$) suggested that learners experienced AI as a private, low-judgement space for repeated practice, while teachers and peers remained the main sources of social connection. AI appeared to reduce social-evaluative pressure for some learners while introducing effort-, time-, and technology-related burdens. In this implementation, AI-mediated speaking feedback was generally perceived as complementing rather than replacing classroom relationships, but stronger claims require larger samples, baseline measures, and richer qualitative data.

KEYWORDS: Artificial Intelligence, Higher Education, Learner Perceptions, Motivation, Speaking Feedback

*Studies in Linguistics and
Language Education (SLLE)*

© 2026 The Author(s).
Published by the University of
Oran 2 Mohamed Ben
Ahmed, Algeria.

This is an open access article
under the [CC BY license](https://creativecommons.org/licenses/by/4.0/)



1 Introduction

Artificial intelligence (AI)–mediated tools are increasingly integrated into language education, including automated feedback, adaptive practice, and conversational support for L2 learners (Bao et al., 2025; Huang et al., 2023). In speaking-focused instruction, these tools are promoted as a way to expand practice time, provide immediate feedback, and reduce routine feedback burdens for teachers. At the same time, a persistent concern argues that increased reliance on AI may erode the social and relational foundations of education — specifically, reduced teacher–student contact, weaker peer interaction, and a diminished social-emotional climate (Deci & Ryan, 2000; Ryan & Deci, 2017).

Despite the rapid expansion of AI-for-language-learning research, a specific gap remains. Relatively few studies directly compare learners' perceptions of AI-mediated and teacher-led feedback in speaking-focused courses using constructs that explicitly capture both relational experience and perceived burden (Bao et al., 2025; Huang et al., 2023). Many existing studies report broad attitudes or end-of-course impressions, which makes it difficult to determine whether AI is associated with shifts in social connection, capability beliefs, or perceived burden. This gap matters because debates about AI in language education often move quickly from perception to strong claim, even when the available evidence is based on modest samples and post-course comparisons rather than baseline-adjusted designs.

The present study addresses part of this gap by examining post-course perceptions of AI-mediated versus teacher-led feedback in a university English-speaking course. Using a convergent mixed-methods design, we conduct descriptive group comparisons on competence, relatedness, expectancy, and cost, and analyse learners' open-ended accounts of what was most helpful and what they would change. Two theoretical frameworks guide the investigation: self-determination theory (SDT; Deci & Ryan, 2000; Ryan & Deci, 2017) and the expectancy–value–cost model (Barron & Hulleman, 2015; Flake et al., 2015; Wigfield & Eccles, 2000). Together, these frameworks provide a focused lens for exploring how learners interpreted the two feedback environments, with particular attention to social connection, capability beliefs, and perceived burden.

2 Literature Review

2.1 AI-mediated feedback in language education and speaking practice

AI-mediated tools have expanded rapidly in language education, encompassing systems that provide automated feedback, conversational practice, and adaptive tasks. Reviews describe a growing research base across language skills, with increasing attention to feedback, assessment, and learner support functions (Bao et al., 2025; Huang et al., 2023). In university contexts, generative AI systems have been examined as supports for language learning, with studies reporting generally positive learner perceptions alongside concerns about accuracy, pedagogical fit, and the need for careful integration (Klimova et al., 2024).

For speaking practice specifically, AI-supported environments are positioned as offering scalable rehearsal opportunities, rapid feedback, and a private space that may feel less socially pressurised than public classroom performance. Mixed-methods evidence from AI-assisted learning contexts suggests that motivational and engagement-related experiences can co-occur with speaking performance outcomes, supporting the view that technology-mediated practice can shape learning through motivational pathways (Huang, 2025). At the same time, research indicates that technology-mediated formats can influence affective variables such as anxiety and learning attitudes (Biju et al., 2024), and that learners' uptake of feedback may differ when feedback is teacher-generated versus AI-generated (Zou, Guo, Wang, & Liu, 2025). The

literature therefore suggests that AI-mediated feedback is not uniformly beneficial or harmful; its effects depend on how learners interpret the feedback and how AI use interacts with existing classroom supports. A practical issue across these strands is that although speaking involves both individual performance and social interaction, most studies examine AI feedback in terms of proficiency outcomes or general attitudes rather than probing the specific motivational processes — capability beliefs, social connection, and perceived burden — that shape learners' willingness to speak and persist.

2.2 Self-determination theory: Competence and relatedness

Self-determination theory (SDT) proposes that autonomy, competence, and relatedness are basic psychological needs that support high-quality motivation and well-being (Deci & Ryan, 2000; Ryan & Deci, 2017). Competence refers to feeling effective and capable of meeting task demands, while relatedness refers to feeling socially connected, supported, and cared for in the learning environment. These constructs are highly relevant for speaking because learners' willingness to participate depends not only on linguistic ability but also on whether they feel safe, supported, and able to manage interactional pressure. In SDT terms, AI-mediated feedback could strengthen competence by supporting repeated practice and visible progress, while relatedness may remain anchored in teacher and peer interaction if the classroom continues to provide consistent relational support.

SDT-informed studies of GenAI-supported learning report that learner experiences can vary depending on how AI use is perceived, including whether it supports self-direction and a positive psychological experience (Wu & Wang, 2025). In writing contexts, learners' accounts often portray GenAI as a helpful companion for feedback, while raising concerns about boundaries, dependence, and the limits of non-human support (Teng, 2024). These findings suggest that AI may support competence without fulfilling the relational functions that teachers and peers provide.

Although autonomy is central in SDT, the present study does not include autonomy as a focal outcome. The decision reflects scope and alignment with the core debate motivating the paper: concerns about whether AI feedback undermines social connection and capability beliefs. Competence and relatedness map directly onto these concerns.

2.3 Expectancy-value-cost framework

Expectancy–value theory explains engagement through learners' expectancy of success and the value they attach to a task (Wigfield & Eccles, 2000). The expectancy–value–cost extension makes cost explicit, emphasising that learners may disengage when perceived costs outweigh expectancy and value (Barron & Hulleman, 2015; Flake et al., 2015). Cost is multifaceted and may include effort cost, emotional cost (e.g., anxiety, embarrassment), and opportunity costs. Cost has been described as a neglected but theoretically central component of motivation, and measuring it can clarify why learners avoid tasks that they otherwise value or believe they can succeed at (Flake et al., 2015). Practical measures have been developed to support comparative analyses of motivational profiles across learning conditions (Kosovich, Hulleman, Barron, & Getty, 2015).

In AI-mediated speaking contexts, expectancy and cost are especially relevant because AI tools can plausibly shift both confidence and burden simultaneously. AI may enhance expectancy by making progress visible through scores and feedback, while introducing new costs related to effort, time demands, and technical friction.

2.4 Previous empirical research and the remaining gap

Existing research provides suggestive but incomplete evidence regarding AI feedback and motivational consequences. Work on GenAI-supported speaking has begun to model cognitive-motivational mechanisms, including achievement emotions, that may help explain when AI-supported speaking practice feels productive versus emotionally burdensome (Wu & Liu, 2025). Comparative feedback research suggests that learners may respond differently to teacher-generated versus AI-generated feedback, with uptake shaped by perceived usefulness and credibility (Zou et al., 2025).

Two issues remain underexplored. First, many studies report general attitudes or broad outcomes, but fewer directly compare AI-mediated feedback with teacher-led feedback using constructs that capture both capability beliefs and relational experience. Second, cost is frequently discussed but less often treated as a central outcome, despite evidence that it is critical for understanding engagement and avoidance (Barron & Hulleman, 2015; Flake et al., 2015). If AI practice lowers one dimension of cost, such as social-evaluative pressure, it may still introduce other costs linked to effort, time, or technical friction. Without explicit attention to cost, these trade-offs remain invisible.

2.5 The present study and research questions

Against this background, the present study examines end-of-course perceptions of AI-mediated versus teacher-led feedback in an L2 speaking course, focusing on competence, relatedness, expectancy, and cost. A convergent mixed-methods design was chosen because it allows descriptive quantitative comparisons to be interpreted alongside learners' own accounts of what felt helpful or burdensome (Creswell & Plano Clark, 2018). Given the posttest-only design, the goal is not to establish baseline-adjusted causal effects on these motivational constructs, but to compare how learners described the two feedback environments at the end of the course. Three research questions guide the study:

1. At the end of the course, how did learners in AI-mediated and teacher-led feedback conditions report competence, relatedness, expectancy, and cost?
2. How did learners describe the most helpful aspects of their speaking practice, and what roles did social connection, competence, expectancy, and burden play in these descriptions?
3. How do learners' narratives help explain and nuance the quantitative pattern, including the possibility that AI-supported practice is experienced as complementary to classroom relationships while involving different kinds of burden?

3 Method

3.1 Research design

The study employed a convergent mixed-methods design (Creswell & Plano Clark, 2018) in which quantitative and qualitative data were collected concurrently through a single end-of-course survey. The purpose of combining methods was complementarity: quantitative data provide a descriptive comparison of the two groups on four motivational constructs, while qualitative comments illuminate how learners interpreted those experiences in their own words. Both strands were analysed separately and then integrated during interpretation.

The quantitative component used a quasi-experimental, class-level comparison of two intact groups that received different feedback conditions over eight weeks. Because the focal motivational constructs were measured only at the end of the course, the quantitative strand does not estimate change on these constructs and cannot establish baseline-adjusted group

differences. The qualitative component consisted of two open-ended survey questions eliciting learner perceptions in brief written form.

3.2 Context and participants

The study was conducted in the English preparatory school of a large English-medium university in Turkey. Participants were first-year students enrolled in a speaking-focused English course within the institution's compulsory language programme. At the beginning of the semester, students' proficiency ranged from A2 to B1 (CEFR), with most classified as A2.2 or B1.1 based on the institutional placement test.

Participants were drawn from two intact classes that were assigned to different feedback conditions. Because the classes were pre-existing and not randomly assigned, the design is quasi-experimental. Across the two classes, 53 students were enrolled (AI group $n = 30$; teacher group $n = 23$). Sample sizes differ by analysis due to missing data, which is addressed below.

For the quantitative analyses, we included only students with complete data on all four focal scales (competence, relatedness, expectancy, cost), resulting in an analytic sample of 47 students (AI $n = 27$; teacher $n = 20$). Six students were excluded because they left one or more scale items blank.

For the qualitative analyses, we included all students who provided an interpretable response to at least one open-ended item. In total, 48 students responded to each open-ended question (AI $n = 28$; teacher $n = 20$). The discrepancy between the quantitative and qualitative samples arises because one AI-group participant provided qualitative responses but was excluded from quantitative analyses due to missing scale data.

3.3 Procedure

Both groups followed the same course syllabus, used the same core materials, and prepared for the same speaking assessment over an eight-week teaching period. The key difference concerned how feedback was delivered during speaking practice.

In the AI group, students regularly completed speaking tasks through FlowSpeak, an AI-based speaking feedback tool. The system provided automated scores and text-based feedback on dimensions such as fluency, vocabulary, grammar, and pronunciation, and it allowed repeated attempts. In the teacher group, students completed comparable speaking tasks during class and received feedback directly from the course teacher during or after these activities. Students in this group did not use FlowSpeak.

At the end of the eight-week period, students completed an online survey during class time. Participation was voluntary and had no impact on course grades. Prior to analysis, data were anonymised and participants were identified only via codes (FS-xxx for the AI group; TF-xxx for the teacher group).

Ethics and consent. The study procedures were reviewed and approved by the university's human research ethics committee. Students received an information sheet explaining the purpose of the study, what participation involved, and that participation was voluntary and unrelated to course grades. All participants provided informed consent before completing the survey. Survey responses were anonymised prior to analysis and stored on a secure, access-restricted drive available only to the research team.

3.4 Instruments

3.4.1 Motivational scales

The end-of-course survey included composite scales measuring four focal constructs: competence, relatedness, expectancy, and cost. These constructs were selected because they map directly onto the motivational concerns at the centre of debates about AI feedback (Deci & Ryan, 2000; Ryan & Deci, 2017; Barron & Hulleman, 2015; Flake et al., 2015; Wigfield & Eccles, 2000). All items used a 7-point Likert-type response format (1 = strongly disagree; 7 = strongly agree). Scale scores were computed as the mean of the relevant items, so higher scores indicate higher levels of the construct. Each scale contained three items. Internal consistency estimates were computed on participants with complete item-level data ($n = 48$):

Competence ($\alpha = .70$). Items assessed perceived capability and improvement in speaking, aligned with SDT's competence construct (Deci & Ryan, 2000; Ryan & Deci, 2017). Example: "Over the course of this study, I felt I was getting better at speaking English."

Relatedness ($\alpha = .76$). Items captured perceived social connection and support in the speaking component of the course, aligned with SDT's relatedness construct (Ryan & Deci, 2017). Example: "I felt supported by my classmates when working on speaking tasks."

Expectancy ($\alpha = .81$). Items assessed expectations of success in speaking, adapted from expectancy–value–cost measurement approaches (Barron & Hulleman, 2015; Kosovich et al., 2015; Wigfield & Eccles, 2000). Example: "I am confident that I can do well in speaking tasks in this course."

Cost ($\alpha = .63$). Items captured perceived burdens associated with speaking practice, including time/effort and emotional strain (Barron & Hulleman, 2015; Flake et al., 2015). Items were phrased so that higher agreement reflected higher cost. Example: "This course took too much time from my other responsibilities."

The cost reliability estimate ($\alpha = .63$) is below the conventional .70 threshold but falls within the range considered acceptable for short scales (three items) in exploratory research (Taber, 2018). Cronbach's alpha is sensitive to scale length and tends to underestimate reliability for scales with few items (Cortina, 1993; Taber, 2018). Given this constraint, the cost measure was retained for group-level comparisons, and findings involving cost are interpreted with appropriate caution.

3.4.2 Open-ended questions

The survey included two open-ended prompts: (1) "What was the most helpful part of the speaking practice (AI or teacher) for you? Why?" and (2) "What would you like to change or improve about the speaking practice in this course?" Responses ranged from short phrases to multi-sentence comments. Forty-eight students responded to each prompt (AI $n = 28$; teacher $n = 20$). Because many responses were brief, the qualitative strand is used primarily to contextualise and explain the quantitative pattern rather than to claim deep thematic saturation. Responses were analysed as written, and any non-English segments were translated into English for reporting purposes while preserving the original meaning.

3.5 Data analysis

3.5.1 Quantitative analysis

A post hoc power analysis was conducted to contextualise the null findings. Using G*Power (Faul et al., 2007) for an independent-samples t-test (two-tailed, $\alpha = .05$), the achieved power for detecting a medium effect ($d = 0.50$) with the present sample ($n_1 = 27$, $n_2 = 20$) was

approximately .41 — well below the conventional .80 threshold. To achieve 80% power for an effect of this magnitude, approximately 52 participants per group ($N \approx 104$) would be required; detecting a small effect ($d = 0.20$) at the same threshold would require roughly 310 participants per group. These estimates clarify that the present study was appropriately sized for initial exploration but was not adequately powered to detect small-to-medium effects with confidence. The null findings on competence, relatedness, expectancy, and cost should accordingly be interpreted as inconclusive rather than as evidence of equivalence between conditions.

Quantitative analyses compared the AI and teacher groups on competence, relatedness, expectancy, and cost. Students with missing values on any of the four focal scales were excluded, yielding the analytic sample of 47 participants. For each construct and condition, means, standard deviations, and sample sizes were computed. Group differences were examined using Welch independent-samples *t* tests. We report *p* values alongside Cohen's *d* to describe the magnitude of observed differences. Given the modest sample size and posttest-only design, interpretation prioritised descriptive patterns and effect size estimates rather than dichotomous significance decisions, and null results are not treated as evidence of equivalence. We also examined zero-order correlations among the four focal scales to confirm that constructs were associated in expected directions but not redundant (Barron & Hulleman, 2015; Flake et al., 2015).

3.5.2 Qualitative analysis

Qualitative analysis followed a theoretically guided thematic coding approach (Braun & Clarke, 2006), using the four focal constructs—relatedness, competence, expectancy, and cost—as an analytic lens. The analysis was primarily deductive, while remaining open to emergent patterns within each construct. Because the dataset consisted of brief written survey comments rather than interviews, the aim was not deep interpretive theme development but careful categorisation of recurrent patterns that could help explain the quantitative results.

To support credibility, both authors independently coded a subset of responses (approximately 25%) and discussed discrepancies until agreement was reached on the coding framework and category boundaries. An analytic log was maintained throughout the process to document coding decisions and emerging interpretations, supporting an audit trail for dependability. Given the brevity of the responses and the explanatory role of the qualitative strand, we relied on negotiated agreement rather than formal coefficient reporting; this remains a limitation of the study. Qualitative findings were then compared against the quantitative pattern during integration, providing triangulation across strands (Lincoln & Guba, 1985).

3.6 Integration of quantitative and qualitative strands

Integration followed the convergent mixed-methods logic in which both strands are analysed separately and then merged during interpretation (Creswell & Plano Clark, 2018). Integration occurred through two mechanisms: (a) narrative integration in the Discussion, where quantitative patterns are presented alongside qualitative themes that explain or nuance those patterns; and (b) a joint display (Table 3) summarising, for each construct, the quantitative finding alongside the main qualitative themes and representative excerpts. This integration design supports the complementarity purpose of the study.

4 Results

4.1 Quantitative results

Table 1 summarises descriptive statistics and group comparisons for competence, relatedness, expectancy, and cost. Welch independent-samples *t* tests indicated no statistically significant differences between groups on any of the four constructs (all *ps* > .05). Competence beliefs were descriptively higher in the AI group (*M* = 5.58, *SD* = 1.14, *n* = 27) than in the teacher group (*M* = 5.03, *SD* = 1.02, *n* = 20), $t(43.30) = 1.73$, *p* = .091, with a medium effect size (*d* = 0.50). Expectancy showed a similar descriptive pattern, with the AI group reporting somewhat higher expectancy (*M* = 5.58, *SD* = 1.23) than the teacher group (*M* = 5.00, *SD* = 1.28), $t(40.22) = 1.57$, *p* = .125, *d* = 0.46. These results are best read as suggestive descriptive differences rather than as firm group effects, and the null findings should not be interpreted as evidence of equivalence.

Relatedness scores were closely comparable across conditions (AI: *M* = 5.47, *SD* = 1.39; teacher: *M* = 5.38, *SD* = 1.29), yielding a negligible effect size, $t(42.79) = 0.22$, *p* = .828, *d* = 0.06. Cost scores were lower in the AI group (*M* = 3.36, *SD* = 1.40) than in the teacher group (*M* = 4.10, *SD* = 1.32), $t(42.36) = -1.86$, *p* = .070, *d* = -0.54, at trend level with a medium effect size. Because higher cost scores indicate greater perceived burden, this pattern suggests that the teacher group may have experienced somewhat higher overall cost, but the estimate remains uncertain in a sample of this size.

Table 1

Descriptive Statistics and Group Comparisons for Competence, Relatedness, Expectancy, and Cost by Feedback Condition (AI vs. Teacher)

Scale	Group	<i>n</i>	<i>M</i>	<i>SD</i>	<i>t</i>	<i>df</i>	<i>p</i>	<i>d</i>
Competence	AI	27	5.58	1.14	1.73	43.30	.091	0.50
	Teacher	20	5.03	1.02				
Relatedness	AI	27	5.47	1.39	0.22	42.79	.828	0.06
	Teacher	20	5.38	1.29				
Expectancy	AI	27	5.58	1.23	1.57	40.22	.125	0.46
	Teacher	20	5.00	1.28				
Cost	AI	27	3.36	1.40	-1.86	42.36	.070	-0.54
	Teacher	20	4.10	1.32				

Note. All scales ranged from 1 to 7, with higher scores indicating higher levels of the construct. Tests are Welch independent-samples *t*-tests. Cohen's *d* is based on the pooled standard deviation (positive values indicate higher scores in the AI group).

We also examined inter-scale correlations among the four constructs (Table 2). Competence and expectancy were strongly related (*r* = .77), as expected, but correlations with relatedness and cost were smaller, indicating the scales capture related but non-identical aspects of learners' motivational experience.

Table 2

Inter-scale Correlations Among Competence, Relatedness, Expectancy, and Cost (n = 47)

	Competence	Relatedness	Expectancy	Cost
Competence	—	0.57***	0.77***	-0.20
Relatedness	0.57***	—	0.62***	-0.24
Expectancy	0.77***	0.62***	—	-0.22
Cost	-0.20	-0.24	-0.22	—

Note. Pearson correlations. * $p < .05$, ** $p < .01$, *** $p < .001$. Cost is coded so that higher scores indicate higher perceived cost.

4.2 Qualitative themes

Thematic coding of the two open-ended questions identified recurring patterns aligned with the four focal constructs. Because the comments were brief, the findings below should be read as concise explanatory patterns rather than as a full-scale qualitative account. Illustrative excerpts are included to show how learners framed the two feedback environments.

4.2.1 Social connection and support (relatedness)

Across both conditions, students positioned teachers and peers as the primary sources of social connection. Teacher-group learners emphasised the value of face-to-face interaction and explanation (e.g., “Teacher. Face to face feedback is more understandable,” TF-008). Peer interaction was described as motivating and reducing pressure.

In the AI group, FlowSpeak was rarely described in relational terms. Instead, students framed it as a practice space that coexisted with classroom interaction rather than replacing it: “Both the teacher and the FlowSpeak were the most helpful part for me” (FS-020). No respondent explicitly described AI as weakening relationships with teachers or peers, although the brevity of the comments limits strong inference. Overall, relatedness in this dataset appeared to remain anchored in human interaction—teacher guidance, peer collaboration, and the comfort of practising with classmates—regardless of feedback modality.

4.2.2 Competence beliefs and perceived improvement

Learners in both groups described improvements in speaking, but emphasised different mechanisms. In the teacher group, competence was linked to teacher explanation and clarification. One learner noted that teacher feedback was “very helpful for my speaking ability” (TF-001), and others described how classroom opportunities to speak with classmates supported development through practice in authentic interaction. Teacher feedback was valued for its capacity to provide personalised, contextualised explanations of errors and strategies for improvement.

In the AI group, competence was strongly associated with repetition, specific feedback, and error noticing: “I can see my mistakes and fix them. It's good to learn from mistakes” (FS-024). Others emphasised the scoring system as supporting vocabulary, grammar, and pronunciation improvement: “The score system was good to help improve my vocabulary, grammar and pronunciation” (FS-014). Several learners linked these features to improved fluency and confidence, including the perception that the system did not judge them. The combination of private practice, immediate numerical feedback, and the opportunity for

multiple attempts appeared to create a learning environment in which competence could develop through iterative self-correction rather than through social interaction.

4.2.3 Expectancy beliefs and confidence

Across conditions, expectancy was expressed through comments about increased confidence and readiness. In the AI group, learners connected FlowSpeak to confidence building: “FlowSpeak increased my self-confidence when speaking” (FS-028). Others described practising without fear of negative evaluation (FS-004). In the teacher group, expectancy was grounded in authentic interaction, with one learner noting that speaking with “actual friends” increased confidence, while also observing that speaking directly to the teacher could feel “a little bit hard” (TF-020).

4.2.4 Emotional burden and perceived cost

Cost-related themes revealed different burden profiles across conditions. In the teacher group, some students described the social-evaluative difficulty of speaking in front of the teacher (TF-020) and expressed a desire for more speaking time.

In the AI group, several learners described FlowSpeak as reducing social anxiety by providing private, low-judgement practice. Students referred to becoming “less tense” through repeated practice (FS-004). At the same time, learners also described burdens associated with AI practice, especially effort, time constraints, and technical friction. One student noted that recognition problems required extra repetition and requested improvements in detection accuracy (FS-014). Others asked for more time per task or changes in prompt difficulty (FS-015). These narratives indicate that cost is best interpreted as multi-dimensional: AI appears to reduce social-evaluative pressure for some learners while introducing different burdens linked to task design and technical limitations.

Table 3

Joint Display Integrating Quantitative Pattern and Qualitative Themes by Feedback Condition

Construct	Quantitative pattern (AI vs. teacher)	AI-mediated feedback (FlowSpeak) themes and excerpts	Teacher-led feedback themes and excerpts
Relatedness	Nearly identical (AI M = 5.47; Teacher M = 5.38; d = 0.06).	AI described as coexisting with classroom support rather than replacing it. “Both the teacher and the FlowSpeak were the most helpful part for me.” (FS-020, helpful)	Teacher and peer interaction framed as the clearest support. “Teacher. Face to face feedback is more understandable.” (TF-008, helpful)
Competence	AI descriptively higher (AI M = 5.58; Teacher M = 5.03; d = 0.50).	Improvement tied to repeatability and visible error feedback. “I can see my mistakes and fix them.” (FS-024, helpful) “The score system was good to help improve my vocabulary, grammar and	Improvement tied to explanation and interaction. “Teacher's feedback is very helpful for my speaking ability.” (TF-001, helpful)

Construct	Quantitative pattern (AI vs. teacher)	AI-mediated feedback (FlowSpeak) themes and excerpts	Teacher-led feedback themes and excerpts
		pronunciation.” (FS-014, helpful)	
Expectancy	AI descriptively higher (AI M = 5.58; Teacher M = 5.00; d = 0.46).	Confidence linked to low-judgement rehearsal. “FlowSpeak increased my self-confidence when speaking.” (FS-028, helpful)	Confidence linked to authentic interaction, with some social-evaluative challenge. “When I speak English with my actual friends, I feel more confident... speaking in front of the teacher is a little bit hard.” (TF-020, helpful)
Cost (higher scores = higher cost)	Teacher group reported higher cost than the AI group (trend-level; medium effect).	Lower social-evaluative pressure for some, but burden from time/tech/task design for others. “Sometimes the app doesn't understand what I say... it should detect my voice better.” (FS-014, change, translated) “I want more time for speaking.” (FS-015, change) “1 minute is too short.” (AI-group excerpt; FS-020, change)	

Note. Cost is coded so that higher scores indicate higher perceived cost / greater perceived burden. “Helpful” and “change” refer to the two open-ended prompts.

5 Discussion

This study examined learners' end-of-course perceptions of AI-mediated and teacher-led feedback in an L2 speaking course, focusing on competence, relatedness, expectancy, and cost. Quantitative analyses showed no statistically significant differences between conditions for competence, relatedness, expectancy, or cost, although the descriptive pattern suggested somewhat higher competence and expectancy and lower cost in the AI group. Given the modest sample and posttest-only design, these patterns should be interpreted cautiously. Thematic coding of brief open-ended comments helps clarify how learners described the two feedback environments in practice. Below, findings are discussed in relation to previous research and the two theoretical frameworks, with careful attention to what the present design can and cannot support.

5.1 Relatedness: similar end-of-course ratings across conditions

A recurring concern is that increased reliance on AI will weaken teacher–student relationships. The present dataset does not provide strong support for that concern, but neither does it justify

a definitive contrary claim. Relatedness scores were very similar across conditions ($d = 0.06$), and learners in both groups described teachers and peers as their primary sources of social support. FlowSpeak was usually positioned as a supplementary practice space rather than a relational substitute. The most defensible interpretation is therefore modest: in this implementation, students did not describe the AI tool as displacing classroom relationships, and post-course relatedness ratings were similar across the two intact classes.

This pattern is consistent with SDT-informed work showing that relatedness is primarily satisfied through human interaction (Ryan & Deci, 2017) and with studies in which learners portray GenAI as useful while still distinguishing it from the relational support they expect from teachers (Teng, 2024). It also aligns with broader arguments that AI may function differently depending on whether it supplements or replaces teacher interaction (Bao et al., 2025; Huang et al., 2023). At the same time, the present evidence is limited to end-of-course self-report and brief comments, so it should be understood as perception-based rather than as a direct measure of relational change.

The pattern is noteworthy because the AI tool occupied a distinct functional role—private, repeatable practice—while classroom interaction remained the main locus of relational experience. A cautious interpretation is that the course design preserved a clear relational anchor in teacher and peer interaction. Future studies should test this more directly with baseline measures, larger samples, and richer qualitative data, especially in implementations where AI plays a more central instructional role.

5.2 Competence and expectancy: suggestive differences, not firm group effects

Although competence and expectancy were not significantly different between groups, both showed medium-sized descriptive trends favouring the AI group. These patterns are consistent with previous research suggesting that AI-supported practice may strengthen capability beliefs through repetition, visible progress indicators, and opportunities for low-pressure rehearsal (Huang, 2025). However, in the present study they should be treated as suggestive rather than conclusive, because the sample was modest and the constructs were measured only at the end of the course.

The qualitative comments help explain these descriptive trends. AI-group learners often emphasised repeatability, specific feedback, error noticing, and scoring as useful for improvement, whereas teacher-group learners valued explanation, modelling, and authentic interaction. This aligns with Wu and Liu's (2025) cognitive-motivational model and with Zou et al.'s (2025) observation that learners respond differently to feedback depending on perceived usefulness and source credibility. What the present data support most clearly is not a firm advantage of one modality over the other, but the possibility that the two modalities supported capability beliefs through different routes.

From an expectancy–value–cost perspective (Wigfield & Eccles, 2000; Barron & Hulleman, 2015), the AI setting may have provided frequent, concrete cues of progress that some learners interpreted as confidence-building. The teacher setting, by contrast, may have supported confidence through interactional guidance and teacher validation. Because no baseline expectancy or competence data were collected, these possibilities remain interpretive rather than demonstrative.

These dual pathways to competence and expectancy echo a broader pattern in the literature: different feedback modalities can support capability beliefs through different mechanisms, and learners may value both private rehearsal and social interaction as complementary sources of confidence (Klimova et al., 2024; Huang, 2025). For that reason, the present findings are better read as supporting complementarity than as establishing superiority.

5.3 Cost as a multi-dimensional construct

The cost results require careful interpretation. The teacher group reported descriptively higher cost ($M = 4.10$) than the AI group ($M = 3.36$), with a medium effect size ($d = -0.54$) at trend level. This suggests that teacher-led practice was associated with somewhat greater overall perceived burden in this sample. However, the qualitative data reveal that cost was not experienced as a single uniform construct.

AI-group learners described FlowSpeak as reducing social-evaluative pressure by providing a private, low-judgement space for practice. At the same time, they described effort and time burdens related to task repetition, time constraints, and speech-recognition failures. In contrast, teacher-group learners generated fewer technology-related complaints but described the interpersonal challenge of speaking directly to the teacher.

This pattern aligns with the theoretical proposition that cost is multi-faceted and can include emotional cost, effort cost, and opportunity cost (Barron & Hulleman, 2015; Flake et al., 2015). The findings extend this framework by showing that AI-mediated practice does not uniformly reduce cost but rather redistributes it — lowering social-evaluative burden while potentially increasing effort-related and technical burden. This aligns with Biju et al.'s (2024) finding that technology-mediated assessment formats can influence affective variables in complex ways, and it echoes Wu and Wang's (2025) observation that learners' GenAI experiences vary depending on perceived fit with goals and context.

The redistribution pattern has implications for how cost is conceptualised in technology-enhanced learning research. Flake et al. (2015) argued that cost is a neglected but theoretically essential component of motivation, and the present findings illustrate why a composite measure of cost may obscure important within-construct variation. In this study, a learner who experiences reduced social anxiety but increased time pressure might report a moderate overall cost score that masks a qualitatively different experience from a learner with uniformly moderate cost across all dimensions. This points to the value of measuring cost subtypes — emotional, effort, and opportunity cost — separately, as proposed in more recent conceptualisations of the expectancy–value–cost framework (Kosovich et al., 2015).

Taken together, the two strands yield a coherent but appropriately modest account. Quantitatively, the two classes looked broadly similar on relatedness, with suggestive descriptive differences on competence, expectancy, and cost. Qualitatively, learners described the two feedback environments as serving partly different functions: teachers and peers as relational anchors, AI as a private and repeatable practice space. The value of the mixed-methods design here is explanatory rather than confirmatory. The qualitative strand helps interpret why end-of-course ratings may have looked similar in some areas while still reflecting different learner experiences.

The joint display (Table 3) summarises this integration. For relatedness, both strands point to the continued centrality of classroom relationships. For competence and expectancy, the descriptive quantitative pattern and the qualitative comments both suggest that some learners experienced AI practice as confidence-supporting, though the design does not justify a strong group-effect claim. For cost, the integration helps explain how overall burden could look somewhat lower in the AI group while still including specific complaints about time pressure and technical friction.

This integrative reading also clarifies the limits of the evidence. The quantitative strand alone cannot establish that AI preserved relatedness or improved capability beliefs, and the qualitative strand alone cannot support broad generalisation from brief survey comments. Read together, however, the two strands support a cautious interpretation of AI as a complementary practice space in this particular implementation.

5.4 Implications for practice

The findings suggest several implications. First, the stability of relatedness across conditions underscores the importance of integrating AI as an addition to relational teaching rather than a replacement. Second, learners' accounts suggest that AI functions especially well as a private rehearsal space, and instructors can leverage this by sequencing AI practice as preparation for in-class tasks and framing it as low-stakes experimentation. Third, qualitative evidence highlights design levers that shape perceived burden: calibrating task difficulty to proficiency, providing clear instructions, reducing unnecessary retries, and addressing technical issues directly. Finally, learners' descriptions suggest a productive division of labour: AI as a practice partner for repeatable, low-judgement rehearsal, and the teacher as the relational anchor and provider of nuanced feedback.

Several limitations should be acknowledged. The study was conducted in a single course context with a modest sample size, which limits generalisability and statistical power. Medium effect sizes should therefore be interpreted as suggestive rather than definitive. The design relied on end-of-course self-report for the four focal motivational constructs, so no baseline-adjusted comparison was possible. Because the classes were intact rather than randomly assigned, unmeasured pre-existing differences between groups cannot be ruled out. The qualitative dataset consisted of short open-ended survey comments rather than interviews or focus groups, which limited depth, and a formal intercoder reliability coefficient was not reported. The cost measure was an overall composite with modest internal consistency ($\alpha = .63$), which may obscure shifts in different cost subtypes. Future work would benefit from baseline measures, larger samples, richer qualitative evidence, separate measurement of emotional, effort, and opportunity costs, and, where possible, longitudinal tracking of how perceptions evolve with continued AI use.

6 Conclusion

This study examined learners' end-of-course perceptions of AI-mediated and teacher-led feedback in an L2 speaking course, focusing on competence, relatedness, expectancy, and cost. Using a convergent mixed-methods design, it combined post-course survey ratings with brief open-ended learner comments. Quantitative comparisons showed no statistically significant between-group differences on the four focal constructs, although the descriptive pattern suggested somewhat higher competence and expectancy and lower cost in the AI group. Qualitative comments indicated that learners continued to locate social support primarily in teachers and peers, while often describing AI as a private, low-judgement space for repeated practice. The most defensible conclusion is therefore a cautious one: in this implementation, AI-mediated speaking feedback was generally experienced as complementary to, rather than a replacement for, classroom relationships. At the same time, the study does not establish equivalence between feedback conditions, and stronger claims about relational or motivational effects will require larger, baseline-informed, and methodologically richer designs.

Declarations

Competing interests. The authors declare that they have no competing interests.

Use of Generative AI

Generative AI tools were used for language editing and to assist with the organisation of the manuscript. Statistical analyses and calculations were conducted on the study dataset by the author, and all reported descriptive statistics and effect sizes were checked against the

underlying data prior to inclusion. OpenAI ChatGPT was used during manuscript preparation to support language refinement, structural revision, and wording of selected sections. The tool was not used to generate or analyse the study data, perform statistical analysis, or make final interpretive decisions. All outputs were critically reviewed, revised, and verified by the authors, who take full responsibility for the accuracy, integrity, and originality of the manuscript.

References

- Bao, W., Wang, T., Zhang, L., Yusop, F. D., & Ruan, X. (2025). A systematic review of AI in second language acquisition using the expanded SAMR model (2015–2024). *Discover Computing*, 28, 292. <https://doi.org/10.1007/s10791-025-09833-6>
- Barron, K. E., & Hulleman, C. S. (2015). Expectancy–value–cost model of motivation. In J. D. Wright (Ed.), *International encyclopedia of the social & behavioral sciences* (2nd ed., Vol. 8, pp. 503–509). Elsevier. <https://doi.org/10.1016/B978-0-08-097086-8.26099-6>
- Biju, N., Abdelrasheed, N. S. G., Bakiyeva, K., Prasad, K. D. V., & Jember, B. (2024). Which one? AI-assisted language assessment or paper format. *Language Testing in Asia*, 14(1), 45. <https://doi.org/10.1186/s40468-024-00322-z>
- Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2), 77–101. <https://doi.org/10.1191/1478088706qp063oa>
- Cortina, J. M. (1993). What is coefficient alpha? An examination of theory and applications. *Journal of Applied Psychology*, 78(1), 98. <https://doi.org/10.1037/0021-9010.78.1.98>
- Creswell, J. W., & Plano Clark, V. L. (2018). *Designing and conducting mixed methods research* (3rd ed.). SAGE.
- Deci, E. L., & Ryan, R. M. (2000). The “what” and “why” of goal pursuits: Human needs and the self-determination of behavior. *Psychological Inquiry*, 11(4), 227–268. https://doi.org/10.1207/S15327965PLI1104_01
- Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, 39(2), 175–191. <https://doi.org/10.3758/BF03193146>
- Flake, J. K., Barron, K. E., Hulleman, C. S., McCoach, D. B., & Welsh, M. E. (2015). Measuring cost: The forgotten component of expectancy-value theory. *Contemporary Educational Psychology*, 41, 232–244. <https://doi.org/10.1016/j.cedpsych.2015.03.002>
- Huang, M. (2025). Student engagement and speaking performance in AI-assisted learning environments. *Education and Information Technologies*, 30(6), 7143–7165. <https://doi.org/10.1007/s10639-024-12989-1>
- Huang, X., Zou, D., Cheng, G., Chen, X., & Xie, H. (2023). Trends, research issues and applications of artificial intelligence in language education. *Educational Technology & Society*, 26(1), 112–131. [https://doi.org/10.30191/ETS.202301_26\(1\).0009](https://doi.org/10.30191/ETS.202301_26(1).0009)
- Klimova, B., Pikhart, M., & Al-Obaydi, L. H. (2024). Exploring the potential of ChatGPT for foreign language education at the university level. *Frontiers in Psychology*, 15, 1269319. <https://doi.org/10.3389/fpsyg.2024.1269319>
- Kosovich, J. J., Hulleman, C. S., Barron, K. E., & Getty, S. (2015). A practical measure of student motivation: Establishing validity evidence for the expectancy-value-cost scale in

- middle school. *The Journal of Early Adolescence*, 35(5–6), 790–816.
<https://doi.org/10.1177/0272431614556890>
- Lincoln, Y. S., & Guba, E. G. (1985). *Naturalistic inquiry*. SAGE.
- Ryan, R. M., & Deci, E. L. (2017). *Self-determination theory: Basic psychological needs in motivation, development, and wellness*. Guilford Press. https://doi.org/10.1007/978-3-031-17299-1_2630
- Taber, K. S. (2018). The use of Cronbach's alpha when developing and reporting research instruments in science education. *Research in Science Education*, 48(6), 1273–1296.
<https://doi.org/10.1007/s11165-016-9602-2>
- Teng, M. F. (2024). “ChatGPT is the companion, not enemies”: EFL learners' perceptions and experiences in using ChatGPT for feedback in writing. *Computers and Education: Artificial Intelligence*, 7, 100270. <https://doi.org/10.1016/j.caeai.2024.100270>
- Wigfield, A., & Eccles, J. S. (2000). Expectancy–value theory of achievement motivation. *Contemporary Educational Psychology*, 25(1), 68–81.
<https://doi.org/10.1006/ceps.1999.1015>
- Wu, H., & Liu, W. (2025). Exploring mechanisms of effective informal GenAI-supported second language speaking practice. *Discover Computing*, 28, 119.
<https://doi.org/10.1007/s10791-025-09635-w>
- Wu, H., & Wang, Y. (2025). Disclosing Chinese college students' flow experience in GenAI-assisted informal digital learning of English. *Learning and Motivation*, 90, 102134.
<https://doi.org/10.1016/j.lmot.2025.102134>
- Zou, S., Guo, K., Wang, J., & Liu, Y. (2025). Investigating students' uptake of teacher- and ChatGPT-generated feedback in EFL writing. *Computer Assisted Language Learning*. Advance online publication. <https://doi.org/10.1080/09588221.2024.2447279>