



From Output to Input: Using Echo-STT Rehearsal to Enhance L2 Listening Comprehension

Ching-Jung Yang

cjyang@stust.edu.tw, Center for Bilingual Education, Southern Taiwan University of Science and Technology

APA Citation: Yang, C. (2026). From output to input: Using Echo-STT rehearsal to enhance L2 listening comprehension. <i>LEARN Journal: Language Education and Acquisition Research Network</i>, 19(2), 239-264. https://doi.org/10.70730/YPHE8665	
Received 14/01/2026 Received in revised form 11/03/2026 Accepted 27/03/2026	ABSTRACT This mixed-methods study investigated the integration of the Echo Method with speech-to-text (STT) technology to enhance TOEIC listening instruction. The intervention addresses the washback effects of high-stakes testing—the influence of TOEIC on teaching and learning—by transforming test-focused drills into authentic, learner-centered listening practice. Forty-five Applied English majors at a Taiwanese technological university participated in an 18-week intervention featuring reflective echoing, STT-based self-monitoring, and collaborative transcript analysis. Quantitative results from four diagnostic tests (normalized to a 0–100 scale) showed significant listening gains, with weighted overall means rising from 73.98 (pretest) to 87.87 (posttest) ($d = 1.12$). Improvements were greatest in sections with lower cognitive load: Part 1 (Picture Descriptions, $d = 1.48$) and Part 2 (Question-Response, $d = 1.52$). Regression analysis revealed that STT engagement frequency, rather than class attendance, predicted listening gains ($\beta = .62, p < .001$). Qualitative data from journals, interviews, and surveys indicated that STT feedback enhanced metacognitive awareness (73%) and reduced listening anxiety (60%) by externalizing errors in a non-judgmental format. Learners also reported perceived improvements in speaking naturalness, suggesting receptive-

	to-productive transfer effects. Overall, the findings demonstrate that technology-enhanced articulatory rehearsal can bridge the gap between test-focused drills and communicative competence, transforming high-stakes preparation into a learner-centered experience. Keywords: listening comprehension, cognitive load, Echo Method, speech-to-text (STT), washback
--	---

Introduction

In East Asia's competitive labor market, English proficiency—often measured through the Test of English for International Communication (TOEIC)—functions as a major gatekeeping mechanism for university graduation and employment (Hsieh, 2017). In Taiwan specifically, the test remains widely used by universities and employers despite ongoing concerns about its broader educational consequences and validity as a measure of communicative competence.

High-stakes reliance on TOEIC heightens the relevance of washback, defined as the influence of testing on teaching and learning practices. Research consistently shows that intensive test preparation commonly narrows instruction to repetitive drills and strategic test-taking at the expense of communicative development (Alderson & Wall, 1993; Hsiao et al., 2018). TOEIC-oriented preparation has been associated with limited transferable skills, increased learner anxiety, and teaching practices that prioritize pattern recognition over genuine language use (Ahn, 2022). High-pressure environments further constrain learners' cognitive and affective resources, creating conditions that paradoxically undermine the very skills the test purports to measure.

Authentic listening is particularly vulnerable to such negative washback. Although TOEIC includes workplace-relevant listening tasks involving conversations, announcements, and short talks, test preparation often reduces these to pattern-recognition routines rather than developing the phonological processing, working memory management, and metacognitive strategies that underpin genuine listening comprehension. Learners memorize question types and response patterns but struggle to decode natural speech in real-world contexts.

To address this misalignment between test preparation and authentic listening development, the present study introduces an intervention combining the Echo Method with speech-to-text (STT) technology. The Echo Method engages the phonological loop through a reflective delay that separates comprehension from articulation, reducing cognitive load and

strengthening phonological processing (Baddeley & Hitch, 1974; Baddeley et al., 1998). Unlike traditional shadowing, which requires simultaneous listening and speaking, the Echo Method's reflective pause allows learners to attend fully to acoustic input before reproducing it, thereby promoting deeper encoding of segmental and prosodic features (Chung, 2011).

STT technology adds scalable, timestamped feedback that promotes noticing of pronunciation errors and supports iterative cycles of listening, silent processing, echoing, and transcript reflection. Meta-analytic evidence shows that automatic speech recognition (ASR)-based feedback produces medium-to-large improvements in pronunciation and speaking and reduces anxiety through private, repeated practice opportunities (Liu et al., 2025; Ngo et al., 2024; Sun, 2023). STT technology also addresses a fundamental challenge in listening instruction: learners cannot easily observe their own phonological processing. When learners echo spoken input, STT converts their articulation into visible text. This externalization makes decoding errors—which normally remain hidden in learners' minds—objectively visible and therefore correctable. Without such feedback, learners often cannot reliably assess whether they accurately heard and reproduced the target sounds.

Positioned within sociocultural accounts of test washback, the Echo-STT intervention is conceptualized as a mediational tool that can redirect TOEIC preparation from decontextualized rote drills toward guided, learner-centered engagement with authentic listening materials (Booth, 2018). This study examines whether the intervention enhances TOEIC-aligned listening performance, reduces cognitive load, supports metacognitive development, and sustains learner motivation within Taiwan's high-stakes assessment environment.

Research Questions

Three central questions guide this study:

- 1) RQ1 (Performance): Does the Echo-STT intervention improve TOEIC listening scores, and how do effects vary across test sections with different cognitive load demands?
- 2) RQ2 (Perception): How do learners perceive the intervention's impact on their listening-related anxiety, motivation, and metacognitive awareness, and what challenges do they report?
- 3) RQ3 (Practice): How do quantitative performance gains and qualitative learner perceptions converge to explain the effectiveness of the Echo-STT intervention?

Literature Review

This study draws on two complementary frameworks—Cognitive Load Theory (CLT) and sociocognitive principles—to explain how the Echo-STT intervention supports second language (L2) listening development. CLT provides the cognitive basis for managing learners' limited working memory during listening tasks through strategies such as graduated complexity and transcript preview, while sociocognitive theory adds a social dimension, highlighting how technological tools (STT transcripts) and peer interaction (collaborative analysis) mediate learning. Additional constructs, including phonological loop theory (Baddeley, 2003), distributed cognition (Grinschgl & Neubauer, 2022), and test washback research (Booth, 2018), further illuminate specific design choices. Together, these theoretical frameworks explain and support the Echo-STT intervention's effectiveness: CLT accounts for how the method manages working memory constraints during phonological processing; phonological loop theory specifies the cognitive mechanisms underlying articulatory rehearsal; and sociocognitive principles clarify how STT transcripts and peer interaction mediate collaborative learning. These complementary perspectives demonstrate that Echo-STT is both cognitively efficient and socially mediated, transforming high-stakes test preparation into authentic listening practice.

Cognitive Load Theory and L2 Listening

L2 listening is a rapid, cognitively intensive process that requires simultaneous phonological decoding, lexical retrieval, syntactic parsing, and semantic integration (Alnajjar et al., 2023). CLT offers a useful framework for explaining why this task is so demanding: working memory capacity is limited, and auditory input is subject to the “transient information effect,” disappearing before learners can revisit it (Sweller et al., 2019). This momentary nature of spoken language often overloads less proficient listeners, who struggle to keep up with the rapid pace of natural speech.

CLT distinguishes three types of cognitive load that interact in real time during L2 listening. Intrinsic load refers to the inherent complexity of the content—longer utterances, unfamiliar vocabulary, complex syntax, and rapid delivery all increase intrinsic load. Extraneous load stems from inefficient instructional design, distractions, or anxiety that disrupts attentional control. Germane load represents the productive cognitive effort devoted to schema construction and learning—processing that leads to long-term retention and skill development (Sweller et al., 2019). Recent findings by Alnajjar et al. (2023) show that anxiety contributes substantially to extraneous load by disrupting attentional control; their structural equation model

indicates that anxious learners struggle to maintain focus during listening tasks, directly impairing working memory performance and comprehension accuracy. This result aligns with broader cognitive-energetic theories linking working memory capacity to attentional regulation systems (Paas et al., 2010).

These insights suggest that effective listening instruction must address multiple dimensions simultaneously: managing linguistic complexity (intrinsic load), reducing anxiety and distractions (extraneous load), and supporting productive processing strategies (germane load). The Echo-STT intervention targets all three dimensions through structured rehearsal, externalized feedback, and collaborative analysis.

The Echo Method: Rehearsal as Processing

The intervention is grounded in the Echo Method, a rehearsal-based approach originating in Chung’s (2008) work on effective language learning strategies and later validated in classroom studies (Chung, 2011). Unlike traditional shadowing, which requires learners to repeat speech simultaneously with a model, the Echo Method incorporates a purposeful “reflective delay”: learners first listen to an utterance, attend to the brief echoic trace in sensory memory, and only then reproduce it. This delay promotes deeper phonological processing rather than spontaneous surface-level mimicry.

Cognitively, the method separates perception from production, thereby reducing the attentional overload often associated with simultaneous shadowing. Chung’s (2011) empirical findings demonstrate that learners using the Echo Method improved significantly in phonological awareness, segmental accuracy, and listening discrimination—likely because their repetitions were based on fully encoded acoustic input rather than first language (L1)-filtered approximations. By attending to the entire utterance before speaking, learners engage in more complete phonological analysis, noticing features such as reduced forms, linking, assimilation, and intonation patterns that are often missed during simultaneous repetition.

Neurobiological research further corroborates the value of delayed rehearsal. Magnetoencephalography (MEG) evidence indicates that original speech and its delayed echo engage partly distinct neural processing pathways, suggesting that the echo functions as a separate cognitive object that can heighten noticing of prosodic or segmental details (Gao et al., 2024). This neural differentiation may explain why echoing—as opposed to passive listening—strengthens phonological representations more effectively. Classroom research aligns with this interpretation: Hu (2018) found significant listening gains among remedial seventh-grade English as a Foreign Language (EFL) learners in Taiwan who practiced the Echo Method,

accompanied by increased self-efficacy and improved sound-based representations. Learners reported that the reflective delay helped them “hear more clearly” and “catch small sounds” they had previously missed, suggesting that the delay functions as a cognitive scaffold by providing time to notice and encode phonological details before engaging articulatory systems.

Speech-to-Text Technology: Externalizing Feedback

Although the Echo Method enhances learners’ internal auditory representations, self-monitoring during individual practice is often inaccurate. L1-conditioned deviations may go unnoticed even when learners believe they are producing target-like forms (Hirai & Kovalyova, 2024). STT technology compensates for this limitation by converting spoken output into real-time textual feedback, effectively externalizing what learners cannot reliably assess internally—for example, whether they accurately distinguished minimal pairs such as /r/ and /l/ in “right” versus “light,” whether they heard the final /t/ in past tense forms, or whether they recognized weak forms such as “can” versus “can’t” in rapid speech. Such phonological details are often misperceived by learners yet remain unnoticed without external feedback.

Empirical evidence shows that widely available automatic speech recognition (ASR) tools—such as Google Docs Voice Typing, Apple Dictation, and Windows Dictation—are consistently sensitive to segmental errors and L1-influenced pronunciation patterns, despite differences in overall accuracy across platforms (Hirai & Kovalyova, 2024). When an STT system mistranscribes a word (e.g., right → light), the resulting discrepancy provides immediate, visible evidence of a phonological mismatch. This “error visibility” reduces extraneous cognitive load by shifting accuracy monitoring to the technology and enhances germane processing by directing attention to specific gaps in learners’ interlanguage phonology.

A growing body of synthesis research supports the pedagogical value of this feedback mechanism. A recent meta-analysis by Ngo et al. (2024) reports medium-to-large effect sizes ($d = 0.65\text{--}0.82$) for ASR-supported pronunciation training, particularly when visualized feedback is included. Similarly, Liu et al. (2025) found that ASR-mediated practice significantly improved L2 pronunciation accuracy, comprehensibility, and fluency across diverse learner populations and proficiency levels. Importantly, ASR-mediated practice has been found to lower speaking anxiety and create a low-stakes environment conducive to repeated, autonomous rehearsal (Sun, 2023). Because learners practice with a non-judgmental machine rather than a human interlocutor, they report greater willingness to take risks, make errors, and persist through multiple attempts. Together, these findings indicate that

STT complements the Echo Method by reinforcing perceptual training with objective, visually accessible feedback, thereby supporting both cognitive efficiency and affective readiness for sustained pronunciation development.

Sociocognitive Framework: Collaborative Learning and Distributed Cognition

The intervention situates the Echo Method and STT technology within a sociocognitive perspective in which cognition is distributed across learners, tools, and environments. From this viewpoint, learning is understood not as an isolated mental activity but as a dynamically coordinated system of individuals interacting with mediating artifacts (Grinschgl & Neubauer, 2022). Within such systems, STT transcripts function as externalized cognitive resources—shared representations that learners can jointly inspect, interpret, and manipulate.

These transcripts serve as collaborative artifacts that support socially shared regulation and evidence-based peer feedback. When learners examine STT outputs together, the visible transcription anchors their metalinguistic discussion, shifting evaluative talk from subjective impressions (e.g., “I think you said it wrong”) to jointly verifiable phonological evidence (e.g., “The STT wrote ‘light,’ but the text says ‘right,’ so we need the /r/ sound”—examples drawn from classroom observations during the intervention). This collaborative error analysis aligns with Vygotskian accounts of the Zone of Proximal Development (ZPD), in which learners scaffold one another’s understanding through dialogic interaction when they identify complementary knowledge gaps (Vygotsky, 1978). However, ZPD activation is not automatic; when peers share the same error or misunderstanding, collaborative analysis may not yield new developmental insights, and effective peer scaffolding depends on asymmetric knowledge that allows more knowledgeable peers to support those with greater learning needs.

It is important to clarify the broader theoretical positioning of this sociocognitive framing. While Bandura’s (1986) social cognitive theory emphasizes reciprocal determinism—how personal factors, behavior, and environment mutually shape learning—Vygotsky’s sociocultural theory foregrounds mediation and ZPD as mechanisms through which higher psychological functions develop. The present study adopts an overarching cognitive-social framework that integrates these perspectives with CLT and phonological loop theory: individual cognitive processes (e.g., managing working memory and articulatory rehearsal) are continuously shaped by social interaction, technological tools (STT transcripts), and classroom norms. Within this framework, STT transcripts and group echo practice operate as mediating artifacts that both enable self-regulation in Bandura’s sense (e.g.,

monitoring one's own performance against visible transcripts) and create ZPD opportunities in Vygotsky's sense (e.g., peers jointly working within each other's developmental range through transcript-based discussion).

Findings from multimodal peer-feedback research support this dynamic: technology-mediated shared artifacts enhance collaborative problem-solving, leading to higher-quality linguistic revisions and deeper metalinguistic engagement (Goh & Vandergrift, 2021). Furthermore, the sociocognitive framing extends to affective benefits. Practicing with non-judgmental AI tools or ASR systems can reduce fear of negative evaluation and increase willingness to communicate (WTC) by offering a psychologically safe rehearsal space before high-stakes interaction (Fathi et al., 2024). By validating their intelligibility through STT first, learners often gain greater confidence for subsequent communicative tasks, illustrating how cognitive, social, and affective dimensions are tightly interwoven in this cognitive-social model of Echo-STT-mediated listening development.

Research Gap and Study Rationale

Although substantial evidence supports the Echo Method for enhancing L2 pronunciation (Chung, 2011; Hu, 2018), research on STT and ASR technologies has largely targeted speaking-related outcomes rather than listening development. Recent studies show that ASR-mediated practice can significantly improve L2 pronunciation, comprehensibility, and global speaking proficiency (Sun, 2023). However, the combined potential of the Echo Method's "biological precision" and the multimodal noticing affordances provided by STT remains underexplored, particularly with respect to L2 listening comprehension.

A critical gap persists: no studies have investigated whether STT-mediated echo practice—designed to strengthen phonological decoding through visible, data-based feedback on articulatory accuracy—transfers to improved listening comprehension. Although both the Echo Method and STT individually support pronunciation development, it remains unclear whether the heightened phonological awareness and segmental accuracy they promote extend to receptive listening skills in high-stakes testing contexts.

A parallel gap exists in the literature on test washback. Standardized assessments such as the TOEIC are frequently associated with negative washback, narrowing instruction toward repetitive drills that increase extraneous cognitive load without fostering deeper competence (Hsiao et al., 2018). Even as mobile technologies and digital tools have become more accessible, few pedagogical interventions offer a principled, learner-centered alternative capable of transforming typically negative washback into cognitively aligned, autonomy-supportive practice.

To address these gaps, the present study empirically evaluates an integrated pedagogical model grounded in three complementary theoretical frameworks: (1) Cognitive Load Theory, which explains how the intervention manages working memory demands; (2) phonological loop theory, which accounts for articulatory rehearsal mechanisms; and (3) sociocognitive principles, which clarify how technological tools and peer interaction jointly mediate learning. By merging the biological precision of the Echo Method with the technological affordances of STT and the interpersonal benefits of collaborative transcript analysis, this study seeks to establish a scalable, evidence-based approach to L2 listening development within a high-stakes test preparation context.

Methodology

Research Design

This study adopted a convergent parallel mixed-methods design (Creswell & Plano Clark, 2018), integrating quantitative and qualitative data to capture both the measurable outcomes and the lived experiences associated with the intervention. The quantitative strand—comprising TOEIC listening scores and STT engagement logs—addressed RQ1 by providing objective indicators of performance change and usage patterns. In parallel, the qualitative strand—drawing from reflective journals, focus group interviews, and surveys—addressed RQ2, offering insight into learners' cognitive processes, affective responses, and perceptions of pedagogical effectiveness. Both strands inform RQ3 through systematic convergence of performance data and learner-reported experiences.

Data for both strands were collected concurrently across the 18-week semester. This synchronicity enabled triangulation, allowing the study to compare observed performance gains with learners' self-reported experiences. The convergence of multiple data sources strengthened the interpretive validity of the findings and provided a more nuanced understanding of how the Echo-STT integration shaped learning outcomes.

Participants and Ethical Considerations

Although formal Institutional Review Board (IRB) approval was not required, the study adhered to institutional ethical guidelines for Scholarship of Teaching and Learning (SoTL) research. All participants provided informed consent, with clear assurance that their participation in research-related activities—such as journal analysis and interviews—was voluntary and unrelated to course grades. To protect privacy, all data were anonymized

before analysis.

Participants were 45 Applied English majors (32 female, 13 male; M age = 19.5) enrolled in a required first-year course, Fundamental English Speaking and Listening Practicum, at a technological university in southern Taiwan. Most participants (95%) had graduated from vocational high schools, and among these, 80% had majored in English during their secondary education. Despite this English-focused background, a pre-intervention mock TOEIC listening assessment established a low-intermediate baseline proficiency ($M = 70.81$, $SD \approx 11$ on a 0–100 normalized scale), based on raw correct responses out of 100 items. Scores are reported on a normalized scale for clarity and are not official TOEIC scaled scores. All 45 students completed the course, and no attrition occurred. However, because several instruments (e.g., journals, surveys) were voluntary or low-stakes, not all participants contributed to every dataset; each analysis used all available data for its respective source.

The Echo-STT Intervention

The 18-week intervention took place in weekly 100-minute sessions conducted in a computer lab equipped with desktop computers and headsets, which facilitated independent STT practice alongside small-group collaboration. Task complexity increased gradually, with linguistic scaffolding progressing from segmental accuracy in the early weeks to discourse-level processing near the end of the semester. This progression managed intrinsic cognitive load by beginning with short, high-frequency utterances (1–3 seconds) and advancing to longer stretches of connected speech (30 seconds), allowing learners to build foundational skills before tackling more complex listening demands.

Intervention Timeline

Weeks 1–3 focused on orientation, pre-testing, and segmental accuracy, with learners echoing 1–3 second phrases to build decoding automaticity for individual phonemes and minimal pairs. Weeks 4–9 targeted TOEIC Parts 1 and 2 (Picture Descriptions and Question-Response), introducing suprasegmental features such as sentence stress and intonation contours with full sentences (5–7 seconds). Weeks 10–12 addressed Part 3 (Short Conversations), emphasizing reduced forms and weak vowels in conversational speech through 15-second excerpts. Weeks 13–15 focused on Part 4 (Short Talks), teaching discourse markers and top-down processing strategies with 30-second excerpts. Weeks 16–18 integrated all skills through mixed-task review, peer teaching, and post-testing.

Weekly Class Procedure

Each session followed a structured four-phase cycle designed to balance cognitive load and support the transition from receptive processing to productive rehearsal.

- 1) Phase 1: Pre-class preview (schema construction). Before class, students listened to the target audio while reading the script via the LMS (FlipClass). Following Chung's (2011) principle that comprehension precedes reproduction, this preview stage built semantic schemas and reduced intrinsic load, allowing learners to focus on phonological features during echoing.
- 2) Phase 2: Echo Method with STT (40 minutes). In class, students practiced reflective echoing without scripts, following a strict protocol: (1) Listen, (2) Reflective delay, (3) Repeat into the Whisper-based STT tool (with technical support), and (4) Verify accuracy against the original text. From Week 10, echo tasks were standardized to 15 seconds based on Baddeley's (2003) phonological loop model to avoid overload.
- 3) Phase 3: Collaborative transcript analysis (30 minutes). Groups of 3–4 analyzed STT transcripts as collaborative artifacts, identifying shared errors (e.g., deleted /t/, mislinked words) and creating vocabulary lists for biweekly quizzes. This distributed cognition reduced error-detection load and enhanced metalinguistic awareness.
- 4) Phase 4: Consolidation (15 minutes). The instructor synthesized recurring errors and provided explicit instruction on phonological rules (e.g., linking, weak forms), reinforcing noticing and stabilizing representations through metalinguistic feedback.

Transforming Washback: From Decontextualized Drills to Authentic Engagement

Traditional TOEIC preparation often exemplifies negative washback, reducing authentic listening materials to pattern-recognition exercises focused on answer selection rather than comprehension development (Alderson & Wall, 1993; Hsiao et al., 2018). The Echo-STT intervention was designed to transform this dynamic by recontextualizing TOEIC materials as authentic listening practice rather than mere test items. Rather than drilling students on answer choices, the intervention shifted focus to the linguistic content of TOEIC audio: learners attended to phonological features (reduced forms, linking, intonation), lexical items (vocabulary in workplace contexts), and discourse patterns (conversation structure, cohesion) embedded in the

test materials. The four-phase cycle—preview, echo, collaborative analysis, and consolidation—transformed what would typically be passive listening-and-answer drills into active, metacognitive engagement. Learners became investigators of their own phonological processing rather than test-takers executing strategies. STT feedback provided objective evidence of comprehension accuracy, while collaborative transcript analysis fostered peer-mediated problem-solving. This design intentionally leveraged the authentic workplace scenarios, natural speech, and discourse variety already present in TOEIC materials, redirecting attention from scores to communicative competence. In this way, the intervention harnessed positive washback by aligning test preparation with genuine listening development and learner-centered pedagogy.

Homework and Engagement

To ensure sufficient rehearsal beyond classroom time, students completed five STT-based echoing tasks per week using parallel TOEIC listening materials. All recordings were submitted via FlipClass, which automatically generated timestamped logs for each attempt. These logs documented students' real-time practice patterns across the semester, including submission frequency, timing, and number of attempts per task. Engagement frequency was operationalized as the total number of weekly submissions per student, providing an index of both consistency and intensity of participation in the Echo-STT intervention.

Data Sources and Instruments

TOEIC Listening Diagnostic Tests

Listening proficiency was tracked using four parallel forms of TOEIC Listening diagnostic tests administered in Week 1 (pretest), Week 8 (midterm), Week 18 (posttest), and six weeks later (delayed posttest). Each form contained 100 items distributed across the standard four sections: Part 1 (Picture Descriptions, 6 items), Part 2 (Question-Response, 25 items), Part 3 (Short Conversations, 39 items), and Part 4 (Short Talks, 30 items). Raw scores (number of correct responses out of 100) were converted to a 0–100 normalized scale for reporting purposes, rather than the official TOEIC scaled score, to ensure clarity and facilitate meaningful comparison across administrations and sections. This longitudinal design—spanning four assessment points—enabled tracking of sustained improvement and helped mitigate concerns about short-term practice effects by observing growth patterns across the entire semester and beyond.

STT Engagement Logs

Student engagement was documented through STT Engagement Logs generated via FlipClass in conjunction with a locally hosted Whisper STT engine. Each week, students submitted echoing recordings as audio files. FlipClass captured submission timestamps, file metadata, and the total number of attempts for each task. All audio files were later batch-processed using Whisper, producing full text transcripts and audio-duration data. These logs served as the primary behavioral measure of engagement intensity and later functioned as predictor variables in regression analyses examining how STT-supported echoing practice contributed to listening gains.

Surveys

A Likert-scale survey (1 = strongly disagree, 5 = strongly agree) was administered in Week 17 to capture learners' perceptions of the intervention. The survey assessed listening anxiety, motivation, metacognitive awareness, and self-efficacy (see Appendix A). The 16 items were adapted from established instruments, including the Foreign Language Listening Anxiety Scale (Horwitz et al., 1986) and the Metacognitive Awareness Listening Questionnaire (Vandergrift et al., 2006), ensuring content validity and alignment with theoretical constructs.

Reflective Journals

Students submitted weekly reflective journals (3–5 sentences) describing their learning experiences, challenges, perceived progress, and emerging strategies. Prompts were intentionally open-ended to encourage authentic reflection (e.g., “What did you learn from your STT practice this week? What challenges did you face?”). Across 16 weeks of journaling (Weeks 3–18), the study collected 600 entries out of a possible 720, with missing entries due to occasional absences or incomplete assignments.

Focus Group Interviews

In Week 18, two semi-structured focus group interviews were held with 5-6 students each, representing high- and low-gain learners (top and bottom quartiles). Sessions (30-40 minutes) examined experiences with echo practice, STT feedback, peer collaboration, and transfer to speaking. Based on survey findings of an unexpected transfer effect, the interview guide was refined to include this topic for deeper exploration and triangulation. Interviews were audio-recorded, transcribed, and translated into English by a

bilingual assistant.

Data Analysis

Quantitative Analysis

To address RQ1, changes in normalized TOEIC listening scores across the four testing points were examined using repeated-measures ANOVA, with test section (Parts 1-4) included as a within-subjects factor. All underlying assumptions were met: normality (Shapiro-Wilk, $p > .05$), sphericity (Mauchly's $\chi^2(5) = 8.34$, $p = .31$), and homogeneity of variance (Levene's test, $p > .05$). To examine predictors of individual listening gains, multiple linear regression was conducted using STT engagement frequency and class attendance as independent variables. Listening gain was operationalized as the difference between post-test and pre-test scores. Multicollinearity diagnostics indicated no concerns ($VIF < 2$).

Qualitative Analysis

Qualitative data—600 reflective journal entries, two focus group transcripts, and open-ended survey responses—were analyzed using Braun and Clarke's (2006) six-phase thematic analysis. Inductive coding across the full corpus yielded 87 preliminary codes, later consolidated into broader themes (see Appendix B). To ensure credibility, a graduate assistant double-coded 20% of journals and both transcripts, achieving substantial agreement (Cohen's $\kappa = .82$). Peer debriefing with two applied linguistics colleagues further validated the thematic map and interpretations.

Findings

RQ1: Improvement in TOEIC Listening Scores

Overall Listening Gains (0–100 normalized scale)

A repeated-measures ANOVA showed a significant main effect of time on normalized TOEIC listening scores, $F(3, 132) = 58.43$, $p < .001$, $\eta^2 = .57$. Weighted overall means (item-weighted across Parts 1-4) increased from Week 1 ($M = 73.98$) to Week 8 ($M = 80.88$), Week 18 ($M = 87.87$), and Delayed Posttest ($M = 84.61$), indicating a large overall effect size (Cohen's $d = 1.12$). This pattern suggests that learners made substantial short-term gains during the intervention and retained much of this improvement six weeks later, indicating durable changes in listening competence rather than short-

lived practice effects.

Section-Specific Gains: Differential Effects by Cognitive Load

Separate analyses showed significant gains across all TOEIC sections ($p < .001$). Effect sizes were largest for Parts 1 and 2—tasks involving short, low-load utterances—while Parts 3 and 4, which require processing extended discourse, showed slightly smaller but still large effects:

- 1) Part 1 (Picture Descriptions): 62.1 $\rightarrow$ 78.3 ($d = 1.48$, very large)
- 2) Part 2 (Question-Response): 62.1 $\rightarrow$ 92.6 ($d = 1.52$, very large)
- 3) Part 3 (Short Conversations): 77.2 $\rightarrow$ 88.6 ($d = 0.87$, large)
- 4) Part 4 (Short Talks): 71.5 $\rightarrow$ 84.9 ($d = 1.05$, large)

All scores are reported on a 0–100 normalized scale (raw correct out of 100 items), not official TOEIC scaled scores. This sectional profile is consistent with CLT: when intrinsic load is lower, as in Parts 1-2, learners can devote more resources to germane processing of phonological detail, enabling Echo-STT practice to exert its strongest impact. By contrast, the higher intrinsic demands of extended discourse in Parts 3-4 appear to constrain the additional benefits of phonological training, even though gains in these sections remain educationally meaningful.

Predictors of Listening Gains

A multiple regression analysis predicting listening gains (posttest minus pretest scores) was significant, $F(2, 42) = 18.67$, $p < .001$, $R^2 = .47$. STT engagement frequency emerged as a strong predictor ($\beta = .62$, $p < .001$), accounting for 47% of the variance in listening improvement, whereas class attendance was not a significant predictor ($\beta = .14$, $p = .18$). This divergence indicates that sustained, hands-on engagement with Echo-STT practice drove learning: students who repeatedly submitted echoing tasks and experimented with multiple attempts showed markedly greater gains than peers who attended regularly but engaged minimally with the STT system. Pedagogically, this underscores that Echo-STT functions as an active training regimen whose benefits depend on cumulative practice volume, not on passive exposure. Taken together, these results suggest that when high-stakes TOEIC preparation time is redirected from decontextualized drills to sustained Echo-STT engagement, increased practice intensity yields measurable gains in listening performance and exemplifies a shift from negative to positive washback.

RQ2: Learner Perceptions and Challenges

Analysis of 600 journals, two focus group transcripts, and open-ended survey responses revealed five major themes reflecting both perceived benefits and challenges. First, 73% (33/45) of respondents agreed that STT enhanced metacognitive awareness by helping them “see” errors, shifting from passive hearing to active noticing; as one student noted, “*STT showed me I was confusing ‘can’ and ‘can’t’ because I wasn’t paying attention to the /t/ sound, and I also realized that I was not producing the correct stress or tone*” (Focus Group, High-Gain). However, occasional ASR inaccuracies undermined trust for some learners: “*I said ‘right’ five times, and it kept typing ‘light’*” (Focus Group, Low-Gain).

Second, 60% (27/45) reported reduced listening anxiety due to STT’s private format—“*When I echo, it’s just me and the computer. I don’t feel embarrassed*” (Focus Group, High-Gain)—though peer transcript sharing continued to provoke nervousness. Third, visible progress motivated high-gain learners—“*My Week 3 transcript was full of errors. By Week 15 it was almost perfect*” (Focus Group, High-Gain)—while others noted cognitive fatigue: “*After 20 minutes, my mind wanders and I just repeat*” (Focus Group, Low-Gain).

Fourth, 81% valued transcript previewing for reducing cognitive load, although a minority perceived it as redundant: “*Previewing took too much time; I already understood the meaning*” (Journal, Week 9). Finally, an emergent theme concerned perceived transfer to speaking: 71% (29/41) believed echoing improved speaking naturalness—“*Echoing makes phrases stick in my mouth. Later they just come out naturally*” (Focus Group, High-Gain)—whereas others dismissed it as mechanical mimicry: “*It felt like I was just mimicking a robot*” (Survey). Overall, these findings suggest that while STT and echoing strategies generally enhanced metacognitive engagement and confidence, their effectiveness varied according to learner expectations, technological reliability, and task design.

RQ3: Convergence of Quantitative and Qualitative Data

To address RQ3, quantitative performance gains and qualitative learner perceptions were systematically triangulated. The convergence reveals five key patterns: (1) graduated task complexity aligned with larger gains in low-complexity sections and learner reports of confidence-building; (2) transcript preview reduced cognitive load as evidenced by both survey agreement (82%) and improved focus during echoing; (3) collaborative transcript analysis distributed cognitive effort while enhancing metalinguistic awareness, consistent with STT engagement predicting gains ($\beta = .62, p < .001$); (4) vocabulary consolidation contextualized lexical learning in

authentic workplace scenarios; and (5) adaptive scaffolding addressed attentional limitations, though perceptions varied between sustained focus and fatigue. Together, these patterns demonstrate that the intervention's effectiveness derives from synergistic integration of cognitive load management, technological mediation, and social collaboration rather than any single component.

Discussion

Effectiveness and Mechanisms of Echo-STT

The convergence of quantitative performance data and qualitative learner perceptions for RQ3 indicates that the effectiveness of the Echo-STT intervention arose from the coordinated interaction of several design features rather than from any single component. Graduated task complexity and transcript preview managed intrinsic and extraneous load, helping explain both the larger gains in low-complexity sections ($d = 1.48\text{--}1.52$) compared with high-complexity sections ($d = 0.87\text{--}1.05$) and learner comments such as *"Starting with short phrases built my confidence"* and *"Previewing helped me focus on sounds, not just meaning."* Collaborative STT transcript analysis and repeated homework echoing shifted learners from passive test takers to active problem solvers, consistent with the finding that STT engagement—but not class attendance—predicted listening gains ($\beta = .62, p < .001$). Vocabulary consolidation based on STT-identified errors and ongoing teacher scaffolding further supported this shift, even as some students reported cognitive fatigue during intensive echoing. The delayed posttest pattern, with scores declining slightly yet remaining substantially above baseline, suggests that Echo-STT fostered relatively durable changes in phonological representations and listening routines rather than short-term cramming effects.

At a cognitive level, the central mechanism appears to be the combination of reflective echoing and externalized feedback. The Echo Method's intentional delay between hearing and repetition prevents learners from defaulting to purely meaning-focused processing and instead directs attention to reduced forms, connected speech, and prosodic cues that are often missed in real-time listening. Through repeated articulatory rehearsal, the phonological loop is activated and reinforced, leading to stronger mental representations of spoken language and improved recognition accuracy, a process supported by neurocognitive evidence that original speech and its echo are processed along partly distinct neural pathways (Baddeley, 2003; Baddeley et al., 1998; Gao et al., 2024). STT technology amplifies this mechanism by turning internal phonological processing into visible, analyzable text: at the individual level it provides immediate, form-focused

feedback that reduces extraneous load and frees working memory for germane processing, while at the social level transcripts function as collaborative artifacts that anchor distributed cognition and dialogic interaction (Goh & Vandergrift, 2021; Grinschgl & Neubauer, 2022; Vygotsky, 1978). The strong predictive value of STT engagement, contrasted with the nonsignificant effect of class attendance, underscores that ASR tools contribute to listening development only when embedded in systematic, reflective practice routines rather than used as occasional add-ons (Fathi et al., 2024; Liu et al., 2025; Ngo et al., 2024; Sun, 2023).

Development of Specific Listening Skills

Within this CLT-sociocognitive framework, the data clarify which aspects of learners' listening actually improved and how Echo-STT contributed to those gains. First, phonological processing became more fine-grained: students increasingly noticed minimal pairs, final consonants, reduced forms, and prosodic cues, reporting that they could “hear more clearly” and “catch small sounds” that had previously gone unnoticed. This pattern corresponds closely to Hu's (2018) findings with middle school learners, suggesting that the sharpening of phonological awareness emerges consistently across age groups, even though the current study involved college students. Second, working-memory management improved as the reflective delay separated perception from production, giving learners time to encode and rehearse input before speaking; this mechanism aligns with the especially large gains in low-load sections where such encoding was most feasible. Third, metacognitive monitoring strengthened when learners used STT transcripts to compare intended and actual output, moving from strategy-driven guessing toward self-diagnosis of recurrent listening problems and clearer awareness of “what to listen for.” Finally, reduced listening anxiety and perceived gains in speaking naturalness indicate that Echo-STT supported affective readiness and receptive-to-productive transfer, extending its impact from test performance to broader communicative competence in a high-stakes assessment environment (Fathi et al., 2024; Goh & Vandergrift, 2021; Horwitz et al., 1986; Vandergrift et al., 2006).

Reframing Test Washback Through Echo-STT

A central motivation for this study was to address the negative washback commonly associated with high-stakes TOEIC preparation in Taiwanese higher education (Ahn, 2022; Alderson & Wall, 1993; Hsiao et al., 2018). The quantitative and qualitative findings together suggest that the Echo-STT intervention partially reframed this washback from a narrowing

influence toward a more productive, learning-oriented force. Instead of reducing TOEIC listening materials to repetitive answer-focused drills, the intervention repurposed them as sources of authentic phonological, lexical, and discourse practice, prompting learners to attend to connected speech, prosody, and meaning in workplace-related contexts (Booth, 2018). Learners' sustained gains in listening scores, increased metacognitive awareness, reduced anxiety, and perceived gains in speaking naturalness indicate that preparation time was used to build transferable communicative skills rather than only test-specific strategies. In this sense, Echo-STT illustrates how principled instructional design can convert typically negative washback into positive washback, aligning test preparation with broader goals of listening development and learner autonomy (Goh & Vandergrift, 2021; Grinschgl & Neubauer, 2022).

Pedagogical Implications

Findings suggest several strategies for designing listening instruction in test-oriented contexts that seek to harness positive washback. First, echo-based practice with STT feedback offers a principled alternative to rote drills by fostering metacognitive engagement and reducing anxiety; teachers can use short, carefully sequenced echoing tasks with transcript preview to manage intrinsic and extraneous load while directing attention to phonological detail (Goh & Vandergrift, 2021; Horwitz et al., 1986; Paas et al., 2010; Sweller et al., 2019). Second, STT tools are most effective when used as formative feedback rather than summative testing: learners should be encouraged to treat transcripts as diagnostic resources for noticing recurrent errors and experimenting with self-correction, rather than as accuracy scores to be maximized (Grinschgl & Neubauer, 2022; Sun, 2023). Third, structuring regular collaborative transcript analysis—brief, focused segments where small groups compare STT output with target scripts—can promote peer-supported problem solving and deepen metalinguistic awareness, consistent with sociocultural principles of mediation and ZPD activation (Bozorgian et al., 2025; Vygotsky, 1978). Finally, given reports of cognitive fatigue and time burden, instructors should differentiate homework requirements and adjust pacing, for example, by alternating high-intensity echoing days with lighter reflection or strategy-focused activities, to sustain motivation while maintaining sufficient practice volume. These strategies illustrate how technology-mediated echoing can reconcile test preparation with communicative competence and learner autonomy, offering a practical template for redesigning TOEIC-oriented listening courses.

Limitations and Future Research

This study has several limitations that qualify the findings and suggest directions for future work. The single-institution, relatively homogeneous sample limits generalizability to other proficiency levels, institutional types, and assessment regimes, and the absence of a control group prevents strong causal attribution to Echo-STT rather than to general course effects. ASR performance remained imperfect, occasionally undermining learner trust, and cognitive load and sustained attention were inferred from design features and self-report rather than measured directly (Hirai & Kovalyova, 2024). In addition, speaking outcomes were not directly assessed despite frequent learner reports of perceived transfer, and the 18-week duration captures only one semester of development. Future research should examine Echo-STT with more diverse learner groups, adopt experimental or quasi-experimental designs to isolate causal mechanisms, and systematically compare different ASR platforms in terms of pedagogical reliability (Liu et al., 2025; Ngo et al., 2024). Studies incorporating physiological or behavioral indicators of cognitive load (e.g., eye-tracking, pupil dilation), pre-post speaking measures, and longitudinal follow-up across multiple semesters would further clarify how Echo-STT shapes both listening and speaking over time and how its washback effects evolve as it becomes institutionalized.

Conclusion

This study demonstrates that technology-enhanced articulatory rehearsal can transform the washback effects of high-stakes testing from a pedagogical constraint into a developmental opportunity. By integrating the Echo Method with speech-to-text feedback, the intervention redirected TOEIC preparation from decontextualized drills toward authentic, learner-centered listening practice that meaningfully enhanced L2 comprehension. Grounded in a cognitive-social framework that integrates Cognitive Load Theory with sociocognitive and sociocultural principles (Bandura, 1986; Paas et al., 2010; Sweller et al., 2019; Vygotsky, 1978), the intervention managed working-memory constraints, increased metacognitive awareness (Goh & Vandergrift, 2021), reduced anxiety (Horwitz et al., 1986), and fostered collaborative learning mediated by STT transcripts. These findings challenge the assumption that test preparation must rely on rote drills that exemplify negative washback (Ahn, 2022; Alderson & Wall, 1993). With principled design—leveraging open-source STT tools and practices such as graduated complexity, transcript preview, collaborative analysis, and adaptive scaffolding—educators can meet high-stakes assessment demands while promoting communicative competence, learner autonomy, and positive

washback. Future research should explore how these design principles can be scaled across diverse contexts and integrated with speaking-focused interventions to capitalize on the emergent transfer effects observed in this study.

Acknowledgements

This research was supported by Southern Taiwan University of Science and Technology through a Pre-SoTL grant. I gratefully acknowledge the 45 students who participated in this study for their dedication and honest reflections. I also extend my sincere thanks to the two colleagues who contributed to the peer debriefing process, to Duke Du for technical assistance with the Whisper STT tool, and to my foreign colleague for generously helping edit this paper.

About the Author

Ching-Jung Yang: An Associate Professor at the Center for Bilingual Education, Southern Taiwan University of Science and Technology. Her research examines technology-enhanced language learning, cognitive approaches to L2 listening, and pedagogical innovation. She also investigates student well-being, intercultural communication, and gender equity in language education.

References

- Ahn, J. (2022). A study of the washback effects of a standardized test on the EFL context. *The European Conference on Language Learning 2022: Official Conference Proceedings*, 149–173.
<https://doi.org/10.22492/issn.2188-112X.2022.14>
- Alderson, J. C., & Wall, D. (1993). Does washback exist? *Applied Linguistics*, 14(2), 115–129. <https://doi.org/10.1093/applin/14.2.115>
- Alnajjar, K., Alkhutaba, M., Rasheed Al-Qeyam, F., Sanad, M., & Hikmat Freihat, R. (2023). The impact of working memory on EFL learners' listening comprehension. *International Journal of Early Childhood Special Education*, 14(8), Article 62.
<https://doi.org/10.9756/INTJECSE/V14I8.62>
- Baddeley, A. (2003). Working memory and language: An overview. *Journal of Communication Disorders*, 36(3), 189–208.
[https://doi.org/10.1016/S0021-9924\(03\)00019-4](https://doi.org/10.1016/S0021-9924(03)00019-4)
- Baddeley, A., Gathercole, S., & Papagno, C. (1998). The phonological loop as a language learning device. *Psychological Review*, 105(1), 158–173.

- <https://doi.org/10.1037/0033-295X.105.1.158>
- Baddeley, A., & Hitch, G. (1974). Working memory. In G. H. Bower (Ed.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 8, pp. 47–89). Academic Press.
[https://doi.org/10.1016/S0079-7421\(08\)60452-1](https://doi.org/10.1016/S0079-7421(08)60452-1)
- Bandura, A. (1986). *Social foundations of thought and action: A social cognitive theory*. Prentice-Hall.
- Booth, M. (2018). *Washback of standardized testing on students' attitudes and motivation in Japanese universities* (Doctoral dissertation, Durham University). Durham University Repository.
- Bozorgian, H., Muhammadpour, M., & Alehchali, A. B. (2025). The effects of metacognitive intervention and metacognitive intervention through dialogic interaction on L2 learners' listening comprehension and cognitive load. *Language Awareness*, 1–27.
<https://doi.org/10.1080/09658416.2025.2489937>
- Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2), 77–101.
<https://doi.org/10.1191/1478088706qp063oa>
- Chung, K. S. (2008). Pronunciation and the good language learner. In C. Griffiths (Ed.), *Lessons from good language learners* (pp. 60–73). Cambridge University Press.
- Chung, K. S. (2011). The effectiveness of the echo method. *English Teaching & Learning*, 35(3), 35–64. <https://link.springer.com/journal/42321>
- Creswell, J. W., & Plano Clark, V. L. (2018). *Designing and conducting mixed methods research* (3rd ed.). SAGE Publications.
- Fathi, J., Rahimi, M., & Derakhshan, A. (2024). Improving EFL learners' speaking skills and willingness to communicate via artificial intelligence-mediated interactions. *System*, 121, 103254.
<https://doi.org/10.1016/j.system.2024.103254>
- Gao, J., Chen, H., Fang, M., & Ding, N. (2024). Original speech and its echo are segregated and separately processed in the human brain. *PLOS Biology*, 22(2), e3002498.
<https://doi.org/10.1371/journal.pbio.3002498>
- Goh, C. C. M., & Vandergrift, L. (2021). *Teaching and learning second language listening: Metacognition in action* (2nd ed.). Routledge.
<https://doi.org/10.4324/9780429287749>
- Grinschgl, S., & Neubauer, A. C. (2022). Supporting cognition with modern technology: Distributed cognition today and in an AI-enhanced future. *Frontiers in Artificial Intelligence*, 5, 908261.
<https://doi.org/10.3389/frai.2022.908261>
- Hirai, A., & Kovalyova, A. (2024). Speech-to-text applications' accuracy in English language learners' speech transcription. *Language Learning &*

- Technology*, 28(1), 1–21.
<https://doi.org/10.64152/10125/73555>
- Horwitz, E. K., Horwitz, M. B., & Cope, J. (1986). Foreign language classroom anxiety. *The Modern Language Journal*, 70(2), 125–132.
<https://doi.org/10.1111/j.1540-4781.1986.tb05256.x>
- Hsiao, I.-C. V., Lin, Y. S., & Hung, S.-T. A. (2018). An investigation of the washback effects of an intensive test-preparation program on EFL learning. *Electronic Journal of Foreign Language Teaching*, 15(1), 81–99.
<https://e-flt.nus.edu.sg/wpcontent/uploads/docroot/v15n12018/hsiao.pdf>
- Hsieh, C. N. (2017). The case of Taiwan: Perceptions of college students and teachers of the TOEIC tests. *ETS Research Report Series*, 2017(1), 1–15. <https://doi.org/10.1002/ets2.12179>
- Hu, S.-Y. (2018). *The study of echo method in remedial instruction to English listening and self-efficacy for 7th graders* (Master's thesis, Tamkang University). <https://doi.org/10.6846/TKU.2018.00242>
- Liu, M., Xu, W., & Chen, Y. (2025). A meta-analysis of automatic speech recognition in second language pronunciation learning. *Computer Assisted Language Learning*. Advance online publication.
<https://doi.org/10.1080/09588221.2024.2466288>
- Ngo, T.-N., Chen, H.-J., & Lai, K. K.-W. (2024). The effectiveness of automatic speech recognition in ESL/EFL pronunciation: A meta-analysis. *ReCALL*, 36(1), 4–21.
<https://doi.org/10.1017/S0958344023000113>
- Paas, F., van Gog, T., & Sweller, J. (2010). Cognitive load theory: New conceptualizations, specifications, and integrated research perspectives. *Educational Psychology Review*, 22(2), 115–121.
<https://doi.org/10.1007/s10648-010-9133-8>
- Sun, W. (2023). The impact of automatic speech recognition technology on second language pronunciation and speaking skills of EFL learners: A mixed methods investigation. *Frontiers in Psychology*, 14, 1210187.
<https://doi.org/10.3389/fpsyg.2023.1210187>
- Sweller, J., van Merriënboer, J. J. G., & Paas, F. (2019). Cognitive architecture and instructional design: 20 years later. *Educational Psychology Review*, 31(2), 261–292.
<https://doi.org/10.1007/s10648-019-09465-5>
- Vandergrift, L., Goh, C. C. M., Mareschal, C. J., & Tafaghodtari, M. H. (2006). The metacognitive awareness listening questionnaire: Development and validation. *Language Learning*, 56(3), 431–462.
<https://doi.org/10.1111/j.1467-9922.2006.00373.x>
- Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Harvard University Press.

Appendix A

End-of-Semester Survey Instrument

Part 1: Likert-Scale Items		Scale: 1 = Strongly Disagree, 2 = Disagree, 3 = Neutral, 4 = Agree, 5 = Strongly Agree
Item	Section	Statement
1	Anxiety & Confidence	I felt anxious about TOEIC listening at the start of the semester.
2		I feel less anxious about TOEIC listening now.
3		Using Otter.ai for private practice made me less nervous about making mistakes.
4		My confidence in my English listening ability has increased.
5	Metacognition (Noticing)	Comparing my STT transcript to the original script helped me “see” my listening errors.
6		I can now identify specific phonological features (e.g., linked sounds, -ed endings) that I used to miss.
7		The collaborative transcript analysis (group work) helped me find errors I missed on my own.
8	Cognitive Load & Attention	The Echo Method (listen, pause, repeat) was a clear and easy procedure to follow.
9		Starting with short sentences (Parts 1 & 2) before moving to longer talks (Parts 3 & 4) was a helpful progression.
10		I found the repetitive echo practice helped train my attention and focus.
11		At times, I found the repetitive echo practice to be mentally tiring or fatiguing.
12	Transcript Preview	Previewing transcripts before class made echoing easier and reduced my cognitive load.
13	Overall Value	The Echo Method was a useful technique for improving my listening skills.

14	The STT tool provided useful feedback for my practice.
15	The collaborative group analysis was a valuable part of the learning process.
16	I would recommend this “Echo + STT” method to other students.

Part 2: Open-Ended Questions

Item	Question
1	What was the most helpful part of this 18-week listening intervention? (e.g., the echo method, the STT feedback, group work, and vocabulary lists) Please explain why briefly.
2	What was the least helpful or most frustrating part of this intervention? Please explain why briefly.

Appendix B

Qualitative Thematic Analysis Coding Scheme for Learner Perceptions and Challenges

Theme	Definition	Sample Codes	Example Quote
Metacognitive Awareness and Technological Limits	Awareness of errors/strategies; frustrations with ASR inaccuracies	Noticing phonology; Error prediction; ASR mistrust	“STT showed me I was confusing ‘can’ and ‘can’t.’” / “I said ‘right’ five times and it typed ‘light.’”
Reduced Anxiety vs. Social Stress	Anxiety reduction via private practice; lingering embarrassment in peer sharing	Safe space; Normalizing errors; Social discomfort	“When I echo, it’s just me and the computer.” / “I still felt nervous when classmates saw mistakes.”
Motivation and Cognitive Cost	Motivation from visible progress; fatigue and tedium as barriers	Visible progress; Attention training; Fatigue	“My Week 3 transcript was full of errors; by Week 15 it was almost perfect.” / “After 20 minutes, my mind wanders.”
Value and Burden of Transcript Preview	Benefits of previewing for load reduction;	Dual-task difficulty; Time burden	“Previewing made echoing easier.” / “Previewing took

Emergent: Perceived Transfer to Speaking	complaints of time burden		too much time; I already understood.”
	Reports of improved speaking naturalness; skepticism about relevance	Phrases ‘sticking’; Lack of transfer	“Echoing makes phrases stick; later they come out naturally.” / “It felt like mimicking a robot.”