



Castledown

 OPEN ACCESS

Technology in Language Teaching & Learning

ISSN 2652-1687

<https://www.castledown.com/journals/tltl/>

Technology in Language Teaching & Learning, 8, 104879 (2026)

<https://doi.org/10.29140/tltl.2026.104879>

Differential Responsiveness to AI-Mediated Feedback in EFL Segmental Pronunciation Learning



VUONG THI HAI YEN^a  (Corresponding author)

PHAM VUONG HA ANH^b 

^a*Faculty of Foreign Languages, Hanoi Metropolitan
University, Vietnam*
vthyen@hnmnu.edu.vn

^b*Faculty of Development Economics, University of
Economics and Business, Vietnam National University,
Hanoi, Vietnam*
phamvuonghaanh@gmail.com

Abstract

This study examines changes in recurrent English segmental errors during an AI-mediated pronunciation feedback condition. Although recurrent segmental difficulties among Vietnamese EFL learners are well documented, less is known about how these errors change during AI-mediated practice or how learners work with the feedback they receive. The study used a quasi-experimental mixed-methods design with 105 first-year English majors at a Vietnamese university. The experimental group ($n = 53$) completed three weeks of AI-assisted pronunciation training with ELSA Speak as the feedback environment; the control group ($n = 52$) continued with conventional classroom-based practice. Overall segmental error rates fell by 13.3 percentage points in the experimental group and by 4.1 points in the control group. The post-test difference between groups was statistically significant ($p < .01$), with a large recorded effect size (Cohen's $d = 0.85$). Change also varied by error category. Final consonant omission showed the clearest descriptive reduction, whereas change in the /ʌ/-/æ/ and /e/-/ei/ distinctions was less even. In the interviews, learners described several ways in which the automated feedback shaped their practice. They reported noticing errors that had previously escaped their attention, trying the same item repeatedly, and practicing outside scheduled class time. Yet the feedback was not always easy to interpret. Several participants found it difficult to determine what the system was identifying when the pronunciation problem was more complex. These results suggest that AI-mediated pronunciation feedback is best understood selectively rather than as uniformly effective

Copyright: © 2026 Vuong Thi Hai Yen & Pham Vuong Ha Anh. This is an open access article distributed under the terms of the Creative Commons Attribution Non-Commercial 4.0 International License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

across targets or learners. In this study, the condition was associated with larger overall gains, but category-level change remained uneven, and the practical value of the feedback depended partly on learners' ability to understand and act on it.

Keywords: AI-mediated feedback; segmental pronunciation; automatic speech recognition; EFL learners; feedback interpretability; Vietnamese EFL context

Introduction

Pronunciation is central to spoken intelligibility, yet it often receives limited instructional time in EFL classrooms (Derwing & Munro, 2005; Neri et al., 2002). Recent university-based evidence likewise shows that pronunciation can remain constrained by limited curricular time and competing instructional demands (Almusharraf, 2024). This constraint becomes more consequential when teachers have little scope for frequent individualized feedback and learners have few opportunities for sustained spoken rehearsal (Neri et al., 2002). The present study is situated in a Vietnamese university context. In this sample, recurrent targets included final consonant omission, medial consonant omission, /ð/ substitution, and vowel confusions involving /ʌ/-/æ/ and /e/-/ei/. Naming these targets at the outset is important because the analysis concerns change in particular segmental problems, not pronunciation accuracy as an undifferentiated outcome.

L2 speech learning research offers a useful basis for interpreting these difficulties. The Speech Learning Model and its revised formulation propose that L2 sound learning is shaped by perceived relationships between L1 and L2 phonetic categories (Flege, 1995; Flege & Bohn, 2021). The Perceptual Assimilation Model likewise describes how non-native contrasts may be assimilated to existing L1 categories, with consequences for discrimination and later production (Best, 1995; Best & Tyler, 2007). These perspectives are used here to frame recurrent segmental difficulty; perceptual categorization itself was not measured. Previous work on Vietnamese EFL learners has reported final consonant deletion, vowel-related difficulty, and consonant substitutions associated with L1 phonotactic constraints (Hung, 2024; Tran, 2024; Tran & Nguyen, 2022). In the present dataset, this broader pattern is represented by final and medial consonant omissions, /ð/ substitutions, and the /ʌ/-/æ/ and /e/-/ei/ contrasts. Appendix A gives the transcriptions used in the pronunciation test. The retained study materials do not identify them with a British, American, Australian, or other national target variety, and no such variety is assigned retrospectively.

ASR-based pronunciation tools are relevant here because they can change when feedback is available, how often learners receive it, and how individually they can practice (McCrocklin, 2016; Ngo et al., 2024; Sun, 2023). Technology-supported pronunciation practice has been shown to create additional opportunities for rehearsal and corrective feedback, although learners respond to ASR output in different ways (Inceoglu et al., 2024; Liakin et al., 2017; Sun, 2023). At the same time, difficult L2 speech contrasts have long been trained without AI. Lively et al. (1993) used high-variability input to train Japanese listeners on English /r/-/l/, and Bradlow et al. (1997) later reported production change following perceptual training. Iverson et al. (2005) compared auditory training approaches for the same contrast, while Nishi and Kewley-Port (2007) examined perceptual training with American English vowels. In a broader synthesis, Sakai and Moorman (2018) found larger gains in perception and smaller, though positive, gains in production. The issue addressed here is therefore not whether pronunciation intervention exists, but how recurrent segmental categories changed within an AI-mediated feedback environment in a Vietnamese university EFL setting.

The literature leaves three connected issues unresolved. First, although pronunciation problems among Vietnamese EFL learners have been documented in prior work, there is limited empirical research examining whether those errors can be systematically reduced through AI-assisted intervention rather than through conventional instruction alone. Second, studies of technology-supported pronunciation learning have frequently reported aggregate gains without tracing those gains to specific error categories, which limits what can be said about what actually improves and under what conditions (Fouz-González, 2020). Third, how AI-assisted tools might shape pronunciation development at the process level remains largely unexamined in the Vietnamese university context. As McCrocklin (2016) has argued, observing post-test improvement is analytically insufficient unless accompanied by an account of how learners interpret feedback, how they act upon it, and why certain forms of pronunciation change emerge more readily than others.

Against this background, the study asks not only whether learners improved, but where improvement was concentrated and how learners used the feedback. Two concepts organize this inquiry: differential responsiveness, referring to the extent to which different error categories show different degrees of reduction under the same feedback condition; and feedback interpretability, referring to the extent to which learners can understand an automated evaluation and translate it into a concrete adjustment in pronunciation. The three research questions are as follows:

- RQ1. To what extent is AI-assisted pronunciation training associated with reductions in segmental pronunciation errors among EFL learners?
- RQ2. Which categories of segmental error show the greatest responsiveness to AI-mediated feedback?
- RQ3. How do learners perceive and engage with AI-assisted pronunciation feedback during the learning process?

The study therefore examines improvement at the level of individual error categories rather than assuming that pronunciation gains are uniform. It also considers how learners reported using AI-mediated feedback during practice, allowing the performance patterns to be read alongside their accounts of engagement.

Literature Review

Pronunciation Difficulties in EFL Contexts

Research on pronunciation in EFL settings places segmental accuracy in relation to intelligibility rather than approximation to a native-speaker norm (Derwing & Munro, 2005). L2 speech learning models suggest that some difficulties may persist when L2 sounds or contrasts are perceived through established L1 categories (Best, 1995; Best & Tyler, 2007; Flege, 1995; Flege & Bohn, 2021). There is already a substantial body of intervention research. High-variability phonetic training has produced perceptual learning for difficult contrasts such as English /r/-/l/ and American English vowels (Iverson et al., 2005; Lively et al., 1993; Nishi & Kewley-Port, 2007), and some training studies have also reported production gains (Bradlow et al., 1997; Sakai & Moorman, 2018). The present study asks a narrower question within that literature: how recurrent segmental errors change during an AI-mediated, ASR-supported feedback condition in a Vietnamese university context.

Vietnamese learners of English have been reported to experience recurring difficulties involving consonant codas, consonant clusters, substitutions, and vowel distinctions (Honey, 1987; Hung, 2024; Tran, 2024; Tran & Nguyen, 2022; Tuan, 2021). The present study focuses specifically on final consonant omission, medial consonant omission, /ð/ substitution, and the /ʌ/-/æ/ and /e/-/ei/ contrasts. These

categories provide a concrete phonetic and phonological basis for examining differential responsiveness to feedback. Final consonant production is especially relevant because English codas may carry lexical or grammatical information that can be weakened or lost when required final segments are omitted. The consonant and vowel categories are treated as test-defined targets in the present dataset. Because the retained materials do not document an intended national target variety, the analysis is tied to the transcribed forms in Appendix A rather than generalized to a British, American, or Australian pronunciation model.

This focus also aligns with an intelligibility-oriented view of pronunciation instruction. Segmental errors are pedagogically important when they interfere with listeners' processing of meaning, not when they merely depart from an idealized native-speaker model. For Vietnamese learners, final consonant production is especially relevant because English codas can carry lexical and grammatical information that may be lost when the coda is omitted or weakened.

Technology-Enhanced Pronunciation Learning

Computer-assisted pronunciation training (CAPT) partly addresses a familiar classroom constraint: in many EFL settings, teachers cannot provide sustained individual pronunciation feedback to every learner. Digital platforms can create additional opportunities for individual rehearsal, particularly where classroom feedback is limited (Liakin et al., 2017; Neri et al., 2002). Previous research has also linked mobile-assisted language learning with oral practice beyond the classroom and gains in speaking accuracy (Moayeri & Khodareza, 2020). More recent evidence from Vietnamese university learners also reports perceived benefits of AI-based pronunciation practice alongside uneven support across pronunciation features and concerns about feedback accuracy (Nguyen Huu Lam & Tran Thanh, 2026). Studies of ASR-supported practice have reported gains in pronunciation and speaking, but the outcomes vary with feedback design and with learners' responses to the feedback (Inceoglu et al., 2024; Sun, 2023). A recent systematic review reports broadly positive outcomes for ASR-supported EFL pronunciation learning while also identifying limited theoretical grounding in much of the literature and a need for closer attention to classroom integration (Liu et al., 2025). Accordingly, this study focuses on how automated feedback was used with a defined set of segmental targets, rather than treating access to AI as an intervention in itself.

When learners can repeat an utterance multiple times, compare their production against a target model, and receive evaluative feedback without teacher mediation, the instructional conditions governing pronunciation learning shift in pedagogically significant ways. Research on pronunciation instruction and corrective feedback suggests that feedback becomes more useful when learners can notice, respond to, and revise their own production (Saito & Lyster, 2012). Less settled is how these affordances translate into change at the level of specific pronunciation problems. Aggregate improvements are frequently documented, yet existing research seldom asks whether gains are distributed evenly across error categories or concentrated in those more amenable to automated feedback (Fouz-González, 2020). Broader reviews of language-learning technologies show that technology can support language learning when it is tied to pedagogical design rather than added as an isolated resource (Golonka et al., 2014). Corpus-based pronunciation work likewise suggests that structured technological exposure can support pronunciation learning, although such work differs from ASR-based feedback in both design and mechanism (Ma et al., 2026). These studies point to the same limitation: access to a tool does not explain how learners interpret feedback or change their production. The pedagogical value of AI-mediated feedback depends on feedback design, the target error, and the learner's ability to understand and use the feedback provided.

A further issue is feedback interpretability. Automated feedback may identify a mismatch, but learners still need to understand what that mismatch means and how to revise their production. AI-mediated

practice remains pedagogically incomplete when it is separated from teacher explanation, peer discussion, or explicit phonetic guidance. Recent empirical work also shows why learner behavior matters. Inceoglu et al. (2024) found that learner responses to ASR output are far from uniform: some repair a failed attempt, others ignore the feedback entirely, and others continue practicing without clear improvement. Sun (2023) makes a related point, showing that ASR tends to support pronunciation and speaking development most effectively when it is embedded in a learning design that pairs automated feedback with meaningful practice opportunities. Taken together, these findings support the present study's decision to look beyond aggregate error reduction and examine how learners actually engaged with the feedback itself.

The AI-Mediated Pronunciation Environment

The intervention used ELSA Speak as the AI-mediated pronunciation environment. Because the exact software build was not recorded, the study identifies the tool as the ELSA Speak app environment available during the 2025 data-collection period. Official ELSA documentation describes the app as using AI and speech-recognition technology to provide feedback on sound, stress, and intonation (ELSA Speak, n.d.). For the present study, the pedagogically relevant functions recorded in the research materials were phoneme-level scoring, immediate evaluative feedback, and opportunities for repeated recording attempts. The analysis is restricted to the segmental targets elicited in the pronunciation tasks, including final and medial consonant omissions, /ð/ substitutions, and the /ʌ/-/æ/ and /e/-/eɪ/ contrasts. The retained records do not specify which visual cues, corrective prompts, or other interface elements appeared for each practice item, so those features are not reconstructed.

Figure 1 summarizes the feedback-practice sequence documented in the retained study materials. It represents the recorded cycle of producing a target item, receiving immediate automated evaluation, noticing the flagged mismatch, and repeating the item to revise production.

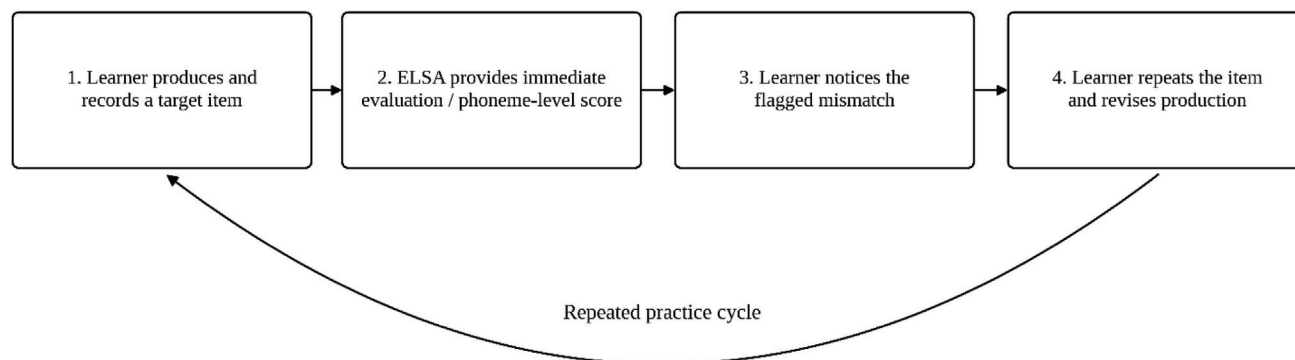


Figure 1. *Recorded AI-Mediated Pronunciation Feedback Cycle used to Interpret the Intervention*

Note. The figure represents only the feedback functions documented in the study materials. It is a workflow representation, not a reproduction of a specific ELSA Speak screen or software build.

The present study treats ELSA Speak narrowly: as a pronunciation feedback environment, not as a full speaking curriculum or a general AI communication platform. The focus is on the pedagogical function most directly connected to the research questions, namely, immediate, repeatable feedback on recurrent English segmental errors. Recent ELSA-focused studies support a cautious reading of its potential. A recent study of Vietnamese secondary-school EFL teachers likewise reported the use of ELSA Speak for pronunciation practice and identified perceived benefits including individualized corrective feedback and increased opportunities for practice, alongside implementation challenges (Chung, 2025). Among English majors in Vietnam, learners have generally reported satisfaction with ELSA Speak, with perceived usefulness, ease of use, and confirmation emerging as factors associated

with that satisfaction (Pham & Pham, 2025). A separate cluster of EFL studies points in a broadly similar direction, suggesting the application may support motivation, autonomy, and pronunciation practice. At the same time, these studies are not uncritical. Several flag the need for teacher mediation, contextualized use, and careful thinking about what automated feedback actually measures in practice (Nguyen et al., 2025; Sholekhah & Fakhrurriana, 2023; Winduastuti et al., 2025).

Theoretical Framework: Speech Learning, Noticing, and Self-Regulated Practice

The theoretical framework combines a speech-learning account of why particular segmental contrasts may remain difficult with process-oriented accounts of how feedback may support change. The Speech Learning Model and Perceptual Assimilation Model locate persistent L2 difficulty partly in the relationship between established L1 categories and incoming L2 phonetic information (Best, 1995; Best & Tyler, 2007; Flege, 1995; Flege & Bohn, 2021). The present study does not test either model directly. They provide a principled explanation for why L1-related segmental targets may differ in difficulty. Schmidt's (1990) noticing account and Zimmerman's (2002) work on self-regulated learning are then used to interpret what learners reported doing with automated feedback: noticing discrepancies, repeating items, and extending practice beyond scheduled class time. This layered framework links the phonetic problem to the technology-mediated learning process without treating the app itself as a causal theory.

The study starts from the premise that AI feedback is unlikely to work uniformly across pronunciation problems. It treats pronunciation change as dependent on the mechanisms through which learners notice, repeat, and regulate practice. Three interrelated processes appear particularly relevant: phonological noticing, iterative practice, and self-regulated learning (Schmidt, 1990; Zimmerman, 2002). Within this framework, AI-assisted pronunciation learning becomes pedagogically meaningful when feedback makes phonological discrepancies perceptually salient, supports repeated re-attempts under evaluative conditions, and extends learner regulation beyond the classroom. These processes are analytically distinct, but they may reinforce one another during practice.

These processes may not operate in the same way for every target. An omission error at the end of a word, for example, creates a relatively discrete presence-versus-absence contrast that may be easier for a learner to notice after repeated scoring than a vowel contrast requiring a subtler articulatory adjustment. The /ð/ substitutions and the /ʌ/-/æ/ and /e/-/ei/ contrasts may therefore place different interpretive demands on the learner even under the same feedback condition. This is a theoretical expectation used to organize the analysis, not a mechanism directly tested by the present design. The empirical question is whether the recorded category-level patterns and learner accounts are consistent with that expectation.

Positioning of the Present Study

The study connects recurrent segmental error patterns, technology-mediated pronunciation feedback, and learner-process accounts of noticing and self-regulated practice. The empirical focus is concrete: final consonant omission, medial consonant omission, /ð/ substitution, and the /ʌ/-/æ/ and /e/-/ei/ contrasts. The contribution is therefore pedagogical as well as phonological. It asks how a mobile AI feedback environment reorganizes opportunities to notice, repeat, and revise specific language forms, and where that feedback remains difficult to interpret. This focus is directly relevant to technology-integrated language pedagogy because the central issue is not technological novelty but the instructional use of automated feedback within pronunciation learning.

The resulting question is therefore narrower and more closely tied to classroom use. When learners with known L1-related segmental difficulties receive access to AI-mediated pronunciation feedback,

what forms of change become visible, which errors prove most responsive, and what learner processes appear to underlie those changes? Framed this way, the study moves beyond descriptive error analysis toward a theoretically informed investigation of intervention, while remaining attentive to the practical realities of TESOL classrooms, where teacher feedback alone cannot provide sufficient density of pronunciation support.

Analytical Framework and Operational Definitions

Error analysis is used here to locate change rather than simply to catalog inaccuracies. Recurrent segmental errors are treated as patterned production problems that are stable enough to warrant pedagogical attention, while the phonological rationale for the selected targets is informed by descriptions of Vietnamese-English mismatch (Honey, 1987; Hung, 2024; Tran, 2024; Tran & Nguyen, 2022; Tuan, 2021). Differential responsiveness refers to descriptive variation in the degree of change across error categories under the same AI-mediated feedback condition. Feedback interpretability refers to whether learners report being able to understand an automated evaluation and use it to guide a concrete adjustment in practice. This construct is interpretive rather than a separately validated causal variable, consistent with the study's use of interview data to contextualize performance patterns (McCrocklin, 2016; Schmidt, 1990; Zimmerman, 2002). The specific error categories examined include final consonant omission, medial consonant omission, /ð/ substitution, and selected vowel confusions such as /ʌ/-/æ/ and /e/-/ei/. These categories were retained because they were empirically recurrent in the sample and consistent with documented Vietnamese-English phonological mismatch (Honey, 1987; Hung, 2024; Tran, 2024; Tran & Nguyen, 2022; Tuan, 2021). Particular attention is given to final consonant errors because English codas may carry lexical and grammatical information, while Vietnamese permits a more restricted range of coda structures. Noticing, iterative practice, and self-regulated engagement are used only as interpretive constructs for reading the quantitative and qualitative findings together (Saito & Lyster, 2012; Schmidt, 1990; Zimmerman, 2002). They were not measured as independent causal variables.

The analysis moves in two steps. At the descriptive level, it tracks changes in recurrent segmental error categories over time and across conditions. At the interpretive level, it uses learner perceptions and interview data to consider how feedback salience, repeated practice, and self-regulated engagement may help account for the observed patterns. Operationally, improvement in pronunciation accuracy is defined as a reduction in the frequency of identified segmental errors between pre-test and post-test. Error reduction is examined both overall and by category because a single aggregate score may conceal differences across target types.

Methodology

Research Design

A quasi-experimental mixed-methods design was used to examine both pronunciation outcomes and learners' reported engagement with AI-mediated feedback (Creswell & Plano Clark, 2018). The quantitative component measured changes in segmental error rates from pre-test to post-test. The qualitative component examined how learners interpreted and responded to feedback during practice. Together, these two strands were organized around a pre-test, intervention, and post-test structure involving an experimental group and a control group. The experimental group received AI-assisted pronunciation training, while the control group participated in conventional classroom-based pronunciation practice. Because participants were drawn from intact classes rather than assigned randomly, the design is quasi-experimental rather than fully experimental. Accordingly, the design permits conclusions about pedagogical association rather than strict causality. In this manuscript, "conventional classroom-based pronunciation practice" refers only to the non-AI comparison condition recorded in

the study documentation. A session-by-session inventory of the control group's pronunciation materials, feedback episodes, and practice dosage was not retained. The control condition is therefore used as a pragmatic classroom comparison rather than as a fully decomposed instructional treatment.

Participants

The participants were 105 first-year English majors enrolled in an English language program at a Vietnamese university. They were between 18 and 19 years old and were assessed as having intermediate English proficiency, approximately CEFR B1-B2. All participants had studied English for at least seven years yet continued to report difficulty with pronunciation at the segmental level. This participant profile situates the study not among beginners with limited formal exposure but among learners who had already received substantial instruction and for whom persistent pronunciation problems could reasonably be treated as stable and pedagogically significant. The students were divided into two groups: an experimental group comprising 53 learners who received AI-assisted pronunciation training, and a control group comprising 52 learners who participated in conventional pronunciation practice. The two groups had similar overall pre-test segmental error rates, with 31.5% in the experimental group and 30.8% in the control group, supporting the comparability of the two groups at the level of overall pre-test error rate. The possibility that other learner variables, such as prior exposure to pronunciation apps, motivation, or out-of-class English use, may have influenced the results cannot be ruled out.

Instruments

Three principal types of instruments were used: pronunciation production tasks, a learner perception questionnaire, and semi-structured interviews. The pronunciation assessment comprised a word-list reading task and a sentence-reading task. The retained test materials show that both task types were constructed to elicit the recurrent targets represented in Appendix A, especially final consonant omission, medial consonant omission, /ð/ substitution, /ʌ/-/æ/ confusion, and /e/-/ei/ confusion. Using isolated words and sentence reading allowed the same focal segments to be observed in more than one production context. The study record does not preserve a separate instrument-development protocol, item-validation study, or written rationale for each individual word and sentence, so no such procedure is added retrospectively. The earlier draft referred to twelve recurrent segmental error types; however, the surviving test and results materials provide a complete item-level record for the five named categories and only an aggregated reference to other categories. The full twelve-category inventory cannot therefore be reconstructed. Recordings were evaluated by two trained raters, and inter-rater agreement was reported as Cohen's $\kappa = .82$. The retained record establishes that the raters evaluated the recorded productions, but it does not document their first-language backgrounds, English variety, detailed training protocol, or the precise sequence of rating decisions. These details are reported as unavailable rather than inferred.

To complement the performance data, the study employed a 20-item Likert-scale questionnaire designed to capture learner perceptions of usefulness, ease of use, and perceived learning impact. The questionnaire provided contextual evidence of learners' experience of the intervention. Semi-structured interviews were also conducted with 15 students from the experimental group. These interviews focused on how learners interpreted feedback, how they practiced with it, and what difficulties they encountered.

Procedure

The overall study spanned four weeks. Week 1 was devoted to baseline testing; the intervention was implemented from Weeks 2 through 4, and post-test, questionnaire, and interview data were collected

in Week 4. In the experimental condition, learners were assigned 20 minutes of daily practice with the AI tool. The retained materials identify the assessment targets listed in Appendix A, but they do not preserve a session-by-session log of the exact in-app activities each learner completed. They also do not establish whether all learners completed an identical sequence of app activities, whether a standardized ELSA familiarization session was delivered before the intervention, or how the assigned daily practice time was independently verified. These details are not inferred. The intervention is therefore interpreted as a supplementary AI-mediated feedback-practice condition, and the incomplete activity and compliance record is treated as a limitation.

Data Analysis

Quantitative data were analyzed using IBM SPSS Statistics. Paired-samples and independent-samples t-tests were used to compare changes within and across groups. The analysis focused not only on overall improvement but also on category-level responsiveness across recurrent segmental errors, since a single global gain score would not capture the study's central analytical concern. The qualitative data were analyzed using thematic analysis following Braun and Clarke (2006). Interview data were read repeatedly, coded for recurring patterns in learners' engagement with AI feedback, and organized into themes related to error noticing, repeated practice, learner independence, and feedback interpretation.

Ethical Considerations

The study received institutional ethics approval in April 2025. Participation was voluntary, informed consent was obtained, and responses were reported in aggregate form. Participants could decline or withdraw without penalty.

Results

Baseline Profile of Segmental Errors

At pre-test, the retained baseline summary states that twelve segmental error types were identified across the full participant sample. Five categories are individually documented in the surviving results table: final consonant omission (15.1%), confusion between /ʌ/ and /æ/ (12.2%), medial consonant omission (11.3%), substitution of /ð/ with /d/ or /z/ (9.3%), and confusion between /e/ and /ei/ (8.1%). The remaining categories are not separately itemized in the retained summary and are therefore not reconstructed. Table 1 reports the baseline profile at the level supported by the available record.

Table 1 *Baseline Segmental Error Profile (Pre-Test, Full Sample, N = 105)*

Error category	Error rate (%)
Final consonant omission	15.1
/ʌ/-/æ/ confusion	12.2
Medial consonant omission	11.3
/ð/ substitution (/d/ or /z/)	9.3
/e/-/ei/ confusion	8.1
Other error categories	Remaining distribution

Note. Error rates represent the percentage of target items on which each error type was observed. The retained baseline summary states that twelve error types were identified, but only the five named categories are individually preserved; the remaining categories are aggregated in the source record.

The baseline profile matters because it shows that the study targeted patterned errors rather than occasional mispronunciations. It also provides the basis for interpreting later differences in responsiveness across error categories.

Overall Reduction in Segmental Error Rates

Table 2 presents the recorded pre-test and post-test error rates for both groups. The experimental group decreased from 31.5% to 18.2%, a reduction of 13.3 percentage points, whereas the control group decreased from 30.8% to 26.7%, a reduction of 4.1 percentage points. According to the retained statistical summary, the between-group post-test difference was statistically significant, $p < .01$, and the experimental group showed a statistically significant change from pre-test to post-test, $p < .001$. The reported standardized effect size was Cohen's $d = 0.85$. The retained analysis record does not include the corresponding t statistics, degrees of freedom, confidence intervals, or underlying output. These values therefore cannot be reported in the complete APA test-statistic format without reconstructing unavailable results. The inferential claims are limited to the statistics preserved in the study record.

Table 2 *Pre-Test and Post-Test Overall Segmental Error Rates by Group*

Group	Pre-test error rate (%)	Post-test error rate (%)	Reduction (pp)
Experimental ($n = 53$)	31.5	18.2	13.3
Control ($n = 52$)	30.8	26.7	4.1

Note. pp = percentage points. The retained statistical summary reports a statistically significant between-group post-test difference, $p < .01$, and a statistically significant experimental-group change from pre-test to post-test, $p < .001$. The reported standardized effect size was Cohen's $d = 0.85$. Exact t statistics and degrees of freedom were not retained.

These figures indicate that the AI-mediated condition was associated with a larger reduction in overall error rates than the conventional classroom-based condition. The experimental condition included documented opportunities for immediate feedback and repeated revision. Because detailed records of feedback episodes, practice dosage, materials, and teacher input in the comparison condition were not retained, the observed between-group difference cannot be attributed to any single instructional feature. The result should therefore be interpreted as an intervention-condition association rather than as evidence for one verified mechanism.

Descriptive Error-Specific Responsiveness

Table 3 presents category-level patterns for the targeted errors after AI-assisted practice. The baseline percentages are drawn from the full-sample pre-test profile in Table 1, while the post-test percentages report the observed rates in the experimental condition. Because group-specific pre-test rates for each error category were not separately recorded, these figures should be interpreted as descriptive indicators of category-level responsiveness rather than as inferential within-group reduction statistics. Within that constraint, final consonant omission showed the clearest descriptive decline, followed by /ʌ/-/æ/ confusion and /ð/ substitution.

Descriptively, change under AI-assisted practice was not evenly distributed across error categories. Final consonant omission showed the largest change, followed by /ʌ/-/æ/ confusion and /ð/ substitution. The smaller descriptive change for /e/-/ei/ confusion indicates that the two vowel contrasts did not follow the same pattern.

Table 3 *Descriptive Error-Specific Responsiveness*

Error category	Baseline profile (%)	Post-test in AI condition (%)	Descriptive change (pp)
Final consonant omission	15.1	7.2	7.9
/ʌ/-/æ/ confusion	12.2	6.8	5.4
Medial consonant omission	11.3	7.5	3.8
/ð/ substitution	9.3	5.1	4.2
/e/-/ei/ confusion	8.1	5.4	2.7

Note. pp = percentage points. Baseline percentages are drawn from the full-sample pre-test profile. Post-test percentages are reported for the experimental condition. Because group-specific pre-test rates by error category were not separately recorded, the table should be interpreted descriptively rather than as inferential within-group reduction evidence.

Learner Perceptions of AI-Assisted Pronunciation Feedback

The questionnaire data indicate which aspects of the AI-supported environment learners found most helpful. The highest-rated item was “helps me notice errors” ($M = 4.6$), followed by “AI feedback is helpful” ($M = 4.5$), “easy to use” ($M = 4.3$), and “increases motivation” ($M = 4.2$). The prominence of error noticing connects directly to the theoretical framework developed earlier in the study. These positive perceptions should not be read as self-sufficient evidence of effectiveness: learner satisfaction and learning gain are not equivalent outcomes. The perception data help explain why the feedback condition may have supported practice: learners reported that it made errors more noticeable and practice more manageable.

Learner Interaction with the AI Feedback: Qualitative Insights

Thematic analysis generated four themes that describe how learners engaged with the AI-mediated feedback. These themes are presented separately to make the qualitative evidence easier to follow.

Error noticing

Learners described the feedback as making pronunciation differences more visible. One participant reported not recognizing a persistent mispronunciation until the application identified its location. This account is treated as learner-reported awareness of an error, not as direct evidence that a cognitive noticing mechanism was measured.

Iterative practice and self-correction

Learners also described repeating the same word after negative feedback. One participant stated, “I repeated the word several times until I got the correct score.” The account shows how immediate evaluation could organize practice into repeated production, evaluation, and revision.

Learner independence beyond class

A third theme concerned the flexibility of practice beyond scheduled class time. As one participant put it, “I can practice anytime, not only in class.” The interview evidence shows that the app extended opportunities for self-directed practice, but the study did not independently measure autonomy or the quality of each out-of-class practice episode.

Feedback interpretability

The final theme qualifies the positive accounts. Some learners found it difficult to understand why a production had been judged inaccurate. One participant explained, “Sometimes I don’t understand why it marks me wrong.” The theme identifies a gap between receiving an evaluation and knowing what to change in the next attempt.

Taken together, the four themes show a conditional pattern of engagement. Feedback was most usable when learners could connect an evaluation to a recognizable pronunciation problem and then make another attempt. When that connection was unclear, the feedback was available but less actionable. Out-of-class access therefore expanded opportunities to practice, yet the interviews do not show that additional practice was equally productive for every learner or target. This distinction between feedback availability and feedback usability provides a qualitative basis for interpreting the uneven category-level changes reported above.

Discussion

AI-Mediated Feedback and Overall Pronunciation Gains

This study investigated how recurrent English segmental errors changed under an AI-mediated pronunciation feedback condition, which error categories showed the clearest descriptive responsiveness, and how learners reported using and interpreting automated feedback. Overall segmental error rates fell more in the experimental group than in the control group, and the retained record reports Cohen’s $d = 0.85$ for the post-test difference. Immediate evaluation and repeated recording were documented features of the AI condition. The comparison condition, however, is not described in enough detail to separate the possible effects of feedback density, practice dosage, materials, or teacher input. The between-group difference should therefore be read at the level of the intervention condition as a whole. This cautious interpretation is consistent with recent ASR research, which reports pronunciation gains alongside substantial variation in learner response and feedback use (Inceoglu et al., 2024; Ngo et al., 2024; Sun, 2023). Recent studies in Vietnamese contexts also identify both pedagogical affordances and limits. Teachers have reported value in individualized corrective feedback and additional practice opportunities (Chung, 2025), while university learners have reported benefits from AI pronunciation practice together with technical inaccuracies and uneven support across pronunciation features (Nguyen Huu Lam & Tran Thanh, 2026).

What matters pedagogically is not whether the application replaced instruction. In this intervention, automated evaluation and repeated attempts were incorporated into a pronunciation-learning sequence. The contribution lies in examining the organization of technology-mediated feedback, not in making a general claim about the effectiveness of a particular commercial application.

Error-Specific Responsiveness

Descriptively, the clearest category-level change was found for final consonant omission; /ʌ/-/æ/ confusion also declined, while /e/-/ei/ confusion showed a smaller change within the same intervention window. Because Table 3 uses full-sample baseline percentages rather than group-specific pre-test rates, these differences do not establish category-specific treatment effects. The pattern nevertheless matters because it shows that segmental targets did not move uniformly under the same feedback condition. Recent research has similarly shown that learners do not experience AI pronunciation support as equally useful across pronunciation features (Nguyen Huu Lam & Tran Thanh, 2026). The present analysis narrows that issue to variation within segmental targets. One possible explanation is that a

discrete omission error may be easier to identify and revise after repeated evaluation than a vowel contrast that requires a finer phonetic adjustment. The current design cannot test that mechanism. The smaller change for /e/-/ei/ therefore raises a focused question for future research about when automated scoring needs to be supplemented by explicit phonetic explanation or teacher guidance.

Learner Processes and Feedback Interpretability

The qualitative findings provide learner-reported context for the quantitative pattern rather than direct tests of learning mechanisms. Learners described becoming aware of pronunciation problems, repeating items after feedback, and practicing beyond scheduled class time. These accounts are compatible with noticing and self-regulated learning perspectives (Schmidt, 1990; Zimmerman, 2002), but those constructs were not measured independently. The interpretability theme places an important limit on that reading. Some learners received a negative evaluation without understanding what needed to change. Recent monitoring of autonomous ASR practice has likewise shown that learner responses to system output vary considerably, including repair, continued practice, and non-response (Inceoglu et al., 2024). In the present study, this makes feedback interpretability more than a usability issue. It marks the point at which an automated judgment may or may not become instructionally actionable. A bounded pedagogical implication follows: difficult targets may require phonetic explanation or teacher clarification when the system score alone does not support a concrete revision.

Theoretical Contribution

The study's theoretical contribution lies in linking three levels of analysis while keeping their functions distinct. Speech-learning accounts explain why particular L1-L2 segmental relationships may remain difficult (Best & Tyler, 2007; Flege & Bohn, 2021). Noticing and self-regulation perspectives help explain how learners may respond once automated evaluation becomes available (Schmidt, 1990; Zimmerman, 2002). The findings indicate that these levels should not be collapsed into a general claim about AI effectiveness. Immediate feedback can create occasions for noticing and repeated practice, but such occasions become instructionally useful only when learners can interpret the evaluation and identify a phonetic adjustment. In this sense, feedback interpretability connects feedback availability with learner action. Differential responsiveness is therefore treated not as a fixed property of the technology, but as an observed pattern that may emerge from the interaction among target characteristics, feedback salience, and learner interpretation. The present design does not test this relation causally, so it should be examined directly in future technology-mediated pronunciation research.

Pedagogical Implications

The pedagogical issue is not simply whether app-based practice should be added, but where automated feedback belongs in a teacher-designed pronunciation sequence. The findings suggest a role for AI feedback as supplementary rehearsal on clearly defined targets, with classroom explanation remaining important when learners cannot interpret a score on their own. One possible instructional sequence would begin with teacher modeling or explanation, move to individual AI-supported rehearsal, and return to teacher clarification when difficulties persist. This sequence was not compared experimentally in the present study. It is a pedagogical implication drawn from the observed error patterns and from learners' accounts of feedback that they could not always interpret. On this reading, technology supplies additional evaluative opportunities, while phonological explanation and pedagogical judgment remain the teacher's responsibility.

For curriculum design, the findings support selective use rather than expansive adoption. In EFL contexts where face-to-face speaking time is constrained, AI-supported pronunciation work may extend

access to individualized practice, especially for recurrent and structurally discrete errors. Learner autonomy should not be treated as a direct consequence of app access. It becomes pedagogically productive when learners understand the feedback, when practice is sustained, and when the tool is aligned with explicit instructional goals. At the program level, AI-assisted pronunciation tools should be treated as components of a wider language-learning ecology. Implementation requires more than access to devices or licenses. It requires a decision about which parts of pronunciation learning can be delegated to repeated automated feedback and which parts still require explanation, modeling, and judgment. It requires task design, teacher preparation, and procedures for helping learners interpret automated scores. The value of technology, then, lies less in the tool itself than in how it reorganizes practice, feedback, and teacher mediation in language learning.

The study also speaks to two different contexts without making the same claim at each level. In the Vietnamese university setting examined here, recurrent segmental problems provided a practical setting for studying supplementary technology-mediated rehearsal. Beyond that setting, the findings add an error-sensitive perspective to ASR-supported pronunciation research: descriptive change differed across segmental targets, and the interviews show that immediate feedback was not always easy for learners to interpret. The quasi-experimental design and incomplete intervention logs place clear limits on both claims.

Conclusion

This study examined how recurrent English segmental errors changed during an AI-mediated pronunciation feedback condition. Overall error rates fell more in the experimental group than in the control group, by 13.3 percentage points compared with 4.1 percentage points under the recorded comparison condition. The reported effect size indicates a substantial post-test difference during the study period, but the quasi-experimental design and incomplete intervention logs do not support a strict causal reading. The category-level evidence is descriptive. Final consonant omission showed the clearest reduction, whereas the two vowel contrasts changed less consistently. The interviews add another part of the picture: learners reported noticing errors, repeating attempts, practicing outside class, and at times being unsure why the app had marked an attempt as inaccurate. The results therefore point to variation both in how segmental targets changed and in how usable learners found the feedback. They do not show that a single mechanism produced the gains, nor do they establish that the same pattern would hold for other learner groups, target varieties, or feedback systems. A question left open by the study is how automated evaluation should be paired with explanation when a learner receives a score but cannot translate it into a specific phonetic adjustment.

Limitations and Future Research

Several limitations shape how these findings can be interpreted. The use of intact classes rather than random assignment limits the study to claims of pedagogical association. The three-week intervention also provides no basis for conclusions about retention. In addition, the retained records do not contain a complete session-by-session inventory of app activities, a detailed control-group practice log, independent verification of the assigned 20 minutes of daily practice, or documentation of a standardized ELSA familiarization session. The surviving rater documentation confirms two trained raters and Cohen's $\kappa = .82$, but not their first-language backgrounds, English variety, detailed training protocol, or full rating procedure. The statistical record is similarly incomplete: significance levels and Cohen's d were retained, whereas exact t statistics, degrees of freedom, confidence intervals, and the underlying output were not. These values have not been reconstructed. The materials also do not identify a national English target variety for the Appendix A transcriptions or preserve a separate validation record for the pronunciation test. No questionnaire reliability coefficient was recorded, so the perception data

are treated as contextual evidence. For the interview analysis, formal intercoder agreement, codebook validation, and member-checking records were not retained. The exact ELSA Speak build and SPSS version are also unavailable. Finally, Table 3 draws its baseline percentages from the full sample rather than from group-specific pre-test rates for each error category, which limits inferential interpretation of category-level change.

Longer-term studies are needed to examine retention and to determine whether the error-specific responsiveness observed here remains stable over time. This direction connects to a broader point in the literature: pronunciation development tends to unfold over extended periods, and short-term gains should not be taken as a reliable index of durable change (Thomson et al., 2024). Comparative work on different feedback designs would also be worth pursuing, particularly studies that distinguish between systems offering only evaluative scoring and those that provide more explicit phonetic explanation. Whether the patterns identified here hold across other learner populations, proficiency levels, and first language backgrounds remains an open question and a natural direction for future research. The next step is to examine how automated feedback and teacher explanation can be sequenced so that learners not only receive correction but also understand how to act on it. Future studies should also specify the exact app build, practice tasks, feedback screens, and teacher-guided reflection procedures so that the contribution of each component can be evaluated more precisely.

Data Availability Statement

The datasets generated and analyzed during the current study are not publicly available due to ethical protections for research participants and privacy commitments made during the informed consent process. Anonymized data may be made available from the corresponding author upon reasonable request, subject to appropriate ethical approval and data-sharing agreements. Survey instruments and interview protocols are available upon request.

Use of Generative AI

OpenAI's ChatGPT was used to assist with wording refinement, clarity, and language editing in selected sections of the manuscript. The tool was not used to generate research data or reconstruct unavailable statistical output. The authors retain responsibility for verifying the manuscript's accuracy, citations, interpretations, originality, and final scholarly content.

References

- Almusharraf, A. (2024). Pronunciation instruction in the context of world English: Exploring university EFL instructors' perceptions and practices. *Humanities and Social Sciences Communications*, 11, Article 847. <https://doi.org/10.1057/s41599-024-03365-y>
- Best, C. T. (1995). A direct realist view of cross-language speech perception. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 171–204). York Press.
- Best, C. T., & Tyler, M. D. (2007). Nonnative and second-language speech perception: Commonalities and complementarities. In O.-S. Bohn & M. J. Munro (Eds.), *Language experience in second language speech learning: In honor of James Emil Flege* (pp. 13–34). John Benjamins Publishing.
- Bradlow, A. R., Pisoni, D. B., Akahane-Yamada, R., & Tohkura, Y. (1997). Training Japanese listeners to identify English /r/ and /l/: IV. Some effects of perceptual learning on speech production. *The Journal of the Acoustical Society of America*, 101(4), 2299–2310. <https://doi.org/10.1121/1.418276>

- 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, A. Q. (2025). AI integration in pronunciation instruction: Insights from secondary school EFL teachers. *Technology in Language Teaching & Learning*, 7(4), Article 103014. <https://doi.org/10.29140/tltl.v7n4.103014>
- Creswell, J. W., & Plano Clark, V. L. (2018). *Designing and conducting mixed methods research* (3rd ed.). SAGE Publications.
- Derwing, T. M., & Munro, M. J. (2005). Second language accent and pronunciation teaching: A research-based approach. *TESOL Quarterly*, 39(3), 379–397. <https://doi.org/10.2307/3588486>
- ELSA Speak. (n.d.). *ELSA Speak: The world's best way to improve your English*. Retrieved June 1, 2026, from <https://elsaspeak.com/en>
- Flege, J. E. (1995). Second language speech learning: Theory, findings, and problems. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 233–277). York Press.
- Flege, J. E., & Bohn, O.-S. (2021). The revised Speech Learning Model (SLM-r). In R. Wayland (Ed.), *Second language speech learning: Theoretical and empirical progress* (pp. 3–83). Cambridge University Press.
- Fouz-González, J. (2020). Using apps for pronunciation training: An empirical evaluation of the English File Pronunciation app. *Language Learning & Technology*, 24(1), 62–85. <https://doi.org/10.64152/10125/44709>
- Golonka, E. M., Bowles, A. R., Frank, V. M., Richardson, D. L., & Freynik, S. (2014). Technologies for foreign language learning: A review of technology types and their effectiveness. *Computer Assisted Language Learning*, 27(1), 70–105. <https://doi.org/10.1080/09588221.2012.700315>
- Honey, P. J. (1987). Vietnamese speakers. In M. Swan & B. Smith (Eds.), *Learner English* (pp. 243–248). Cambridge University Press.
- Hung, D. M. (2024). Updating pronunciation deviations on sound-segments exhibited by Vietnamese students in English speaking. *Arab World English Journal*, 15(2), 376–389. <https://doi.org/10.24093/awej/vol15no2.23>
- Inceoglu, S., Chen, W.-H., & Lim, H. (2024). Monitoring student behavior in autonomous automatic speech recognition-based pronunciation practice. *System*, 124, Article 103387. <https://doi.org/10.1016/j.system.2024.103387>
- Iverson, P., Hazan, V., & Bannister, K. (2005). Phonetic training with acoustic cue manipulations: A comparison of methods for teaching English /r/-/l/ to Japanese adults. *The Journal of the Acoustical Society of America*, 118(5), 3267–3278. <https://doi.org/10.1121/1.2062307>
- Liakin, D., Cardoso, W., & Liakina, N. (2017). The pedagogical use of mobile speech synthesis (TTS): Focus on French liaison. *Computer Assisted Language Learning*, 30(3–4), 325–342. <https://doi.org/10.1080/09588221.2017.1312463>
- Liu, Y., Rahman, F. B. A., & Zain, F. B. M. (2025). A systematic literature review of research on automatic speech recognition in EFL pronunciation. *Cogent Education*, 12(1), Article 2466288. <https://doi.org/10.1080/2331186X.2025.2466288>
- Lively, S. E., Logan, J. S., & Pisoni, D. B. (1993). Training Japanese listeners to identify English /r/ and /l/: II. The role of phonetic environment and talker variability in learning new perceptual categories. *The Journal of the Acoustical Society of America*, 94(3), 1242–1255. <https://doi.org/10.1121/1.408177>
- Ma, Q., Mei, F., & Qian, B. (2026). Exploring EFL students' pronunciation learning supported by corpus-based language pedagogy. *Computer Assisted Language Learning*, 39(4), 915–941. <https://doi.org/10.1080/09588221.2024.2432965>
- McCrocklin, S. (2016). Pronunciation learner autonomy: The potential of automatic speech recognition. *System*, 57, 25–42. <https://doi.org/10.1016/j.system.2015.12.013>

- Moayeri, M., & Khodareza, M. R. (2020). The effect of mobile-assisted language learning on speaking accuracy of EFL learners. *Technology in Language Teaching & Learning*, 2(1), 22–35. <https://doi.org/10.29140/tltl.v2n1.140>
- Neri, A., Cucchiaroni, C., Strik, H., & Boves, L. (2002). The pedagogy-technology interface in computer-assisted pronunciation training. *Computer Assisted Language Learning*, 15(5), 441–467. <https://doi.org/10.1076/call.15.5.441.13473>
- Ngo, T. T. N., Chen, H. H. J., & Lai, K. K. W. (2024). The effectiveness of automatic speech recognition in ESL/EFL pronunciation: A meta-analysis. *ReCALL*, 36(1), 4–21. <https://doi.org/10.1017/S0958344023000113>
- Nguyen, T. S., Nguyen, T. D. T., Hoang, N. Q. N., & Do, T. K. H. (2025). How AI-powered voice recognition has supported pronunciation competence among EFL university learners. *Computer-Assisted Language Learning Electronic Journal*, 26(3), 64–83. <https://doi.org/10.54855/callegj.252634>
- Nguyen Huu Lam, K., & Tran Thanh, T. (2026). Segmental and suprasegmental pronunciation training via AI: An exploration of university students' perceptions and attitudes. *Technology in Language Teaching & Learning*, 8, Article 103341. <https://doi.org/10.29140/tltl.2026.103341>
- Nishi, K., & Kewley-Port, D. (2007). Training Japanese listeners to perceive American English vowels: Influence of training sets. *Journal of Speech, Language, and Hearing Research*, 50(6), 1496–1509. [https://doi.org/10.1044/1092-4388\(2007/103\)](https://doi.org/10.1044/1092-4388(2007/103))
- Pham, V. T. T., & Pham, A. T. (2025). English major students' satisfaction with ELSA Speak in English pronunciation courses. *PLOS ONE*, 20(1), Article e0317378. <https://doi.org/10.1371/journal.pone.0317378>
- Saito, K., & Lyster, R. (2012). Effects of form-focused instruction and corrective feedback on L2 pronunciation development of /ɹ/ by Japanese learners of English. *Language Learning*, 62(2), 595–633. <https://doi.org/10.1111/j.1467-9922.2011.00639.x>
- Sakai, M., & Moorman, C. (2018). Can perception training improve the production of second language phonemes? A meta-analytic review of 25 years of perception training research. *Applied Psycholinguistics*, 39(1), 187–224. <https://doi.org/10.1017/S0142716417000418>
- Schmidt, R. (1990). The role of consciousness in second language learning. *Applied Linguistics*, 11(2), 129–158. <https://doi.org/10.1093/applin/11.2.129>
- Sholekhah, M. F., & Fakhurriana, R. (2023). The use of ELSA Speak as a mobile-assisted language learning (MALL) towards EFL students' pronunciation. *JELITA: Journal of Education, Language Innovation, and Applied Linguistics*, 2(2), 93–100. <https://doi.org/10.37058/jelita.v2i2.7596>
- 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, Article 1210187. <https://doi.org/10.3389/fpsyg.2023.1210187>
- Thomson, R. I., Derwing, T. M., & Munro, M. J. (2024). How long can naturalistic L2 pronunciation learning continue in adults? A 10-year study. *Language Awareness*, 33(2), 201–223. <https://doi.org/10.1080/09658416.2023.2227559>
- Tran, D. T. (2024). Common English pronunciation problems among Vietnamese learners: A case of non-English majors. *Multidisciplinary Reviews*, 7(9), Article 2024227. <https://doi.org/10.31893/multirev.2024227>
- Tran, T. K. L., & Nguyen, A. T. (2022). Common mistakes in pronouncing English consonant clusters: A case study of Vietnamese learners. *Can Tho University Journal of Science*, 14(3), 32–39. <https://doi.org/10.22144/ctu.jen.2022.040>
- Tuan, D. A. (2021). Intelligible pronunciation: Teaching English to Vietnamese learners. *VNU Journal of Foreign Studies*, 37(1), 1–15. <https://doi.org/10.25073/2525-2445/vnufs.4666>

- Winduastuti, N. K., Utami, I. G. A. L. P., & Artini, N. N. (2025). The integration of ELSA Speak on English pronunciation skills in EFL university students: A systematic literature review. *EDU Society: Jurnal Pendidikan, Ilmu Sosial dan Pengabdian Kepada Masyarakat*, 5(3), 120–135. <https://doi.org/10.56832/edu.v5i3.2534>
- Zimmerman, B. J. (2002). Becoming a self-regulated learner. *Theory Into Practice*, 41(2), 64–70. https://doi.org/10.1207/s15430421tip4102_2

Appendix A

Pronunciation Test

Task 1: Word-List Reading

Read each word aloud clearly, once only. Pause briefly between words. Aim to produce each word as accurately as you can. There is no time pressure.

Retained test instruction: Play a short audio model of the task format before recording begins if participants appear unfamiliar with the procedure. The retained records do not establish whether this optional instruction was used during data collection.

No.	Target word	Transcription	Error category	Focal phonological feature
1	asked	/ɑːskt/	Final Consonant Omission	<i>Final consonant cluster: /skt/</i>
2	loved	/lʌvd/	Final Consonant Omission	<i>Final consonant: /vd/ cluster</i>
3	speaks	/spiːks/	Final Consonant Omission	<i>Final consonant: /ks/</i>
4	stopped	/stɒpt/	Final Consonant Omission	<i>Final consonant: /pt/</i>
5	fast	/fɑːst/	Final Consonant Omission	<i>Final consonant: /st/</i>
6	walked	/wɔːkt/	Final Consonant Omission	<i>Final consonant: /kt/</i>
7	months	/mʌnθs/	Final Consonant Omission	<i>Final consonant cluster: /nθs/</i>
8	fixed	/fɪkst/	Final Consonant Omission	<i>Final consonant cluster: /kst/</i>
9	cut	/kʌt/	/ʌ-/æ/ Confusion	<i>Vowel /ʌ/ vs. /æ/</i>
10	cat	/kæt/	/ʌ-/æ/ Confusion	<i>Vowel /æ/ vs. /ʌ/</i>
11	fun	/fʌn/	/ʌ-/æ/ Confusion	<i>Vowel /ʌ/ vs. /æ/</i>
12	fan	/fæn/	/ʌ-/æ/ Confusion	<i>Vowel /æ/ vs. /ʌ/</i>
13	luck	/lʌk/	/ʌ-/æ/ Confusion	<i>Vowel /ʌ/ vs. /æ/</i>
14	exactly	/ɪgˈzæktli/	Medial Consonant Omission	<i>Medial consonant cluster</i>
15	empty	/ˈempti/	Medial Consonant Omission	<i>Medial consonant: /pt/</i>
16	winter	/ˈwɪntər/	Medial Consonant Omission	<i>Medial consonant: /nt/</i>
17	distance	/ˈdɪstəns/	Medial Consonant Omission	<i>Medial consonant cluster</i>
18	this	/ðɪs/	/ð/ Substitution	<i>/ð/ not /d/ or /z/</i>
19	though	/ðəʊ/	/ð/ Substitution	<i>/ð/ not /d/ or /z/</i>
20	breathe	/briːð/	/ð/ Substitution	<i>Final /ð/ not /d/</i>
21	mother	/ˈmʌðər/	/ð/ Substitution	<i>Medial /ð/ not /d/</i>
22	brother	/ˈbrʌðər/	/ð/ Substitution	<i>Medial /ð/ not /d/</i>
23	late	/leɪt/	/e/-/eɪ/ Confusion	<i>Diphthong /eɪ/ vs. /e/</i>
24	let	/let/	/e/-/eɪ/ Confusion	<i>Short /e/ vs. /eɪ/</i>
25	pain	/peɪn/	/e/-/eɪ/ Confusion	<i>Diphthong /eɪ/ vs. /e/</i>
26	pen	/pen/	/e/-/eɪ/ Confusion	<i>Short /e/ vs. /eɪ/</i>
27	name	/neɪm/	/e/-/eɪ/ Confusion	<i>Diphthong /eɪ/ vs. /e/</i>

Task 2: Sentence-Reading

Read each sentence aloud at a natural pace. Try not to stop mid-sentence. Read each sentence once only. A short practice sentence is provided below before the test begins.

No.	Sentence	Target phonological features (rater annotation)
1	She asked her teacher to check the test.	<i>Final consonant cluster /skt/ in asked, /st/ in test</i>
2	The months passed quickly, but he loved the work.	<i>Final consonants in months /nθs/, passed /st/, loved /vd/</i>
3	My brother thinks that this is the best method.	<i>/ð/ in brother, this, the, that; /θ/ in thinks, method</i>
4	She walked exactly five kilometers in the winter.	<i>Final /kt/ in walked; medial clusters in exactly, winter</i>
5	His luck ran out when the fun came to an end.	<i>/ʌ/ in luck, fun vs. /æ/ contrast</i>
6	Cut the fabric carefully before the cat can reach it.	<i>/ʌ/ in cut vs. /æ/ in cat</i>
7	They plan to name the new baby late next spring.	<i>/eɪ/ in name, late vs. /e/ in plan, next</i>
8	She felt pain writing her pen pal's final name.	<i>/eɪ/ in pain, name vs. /e/ in felt, pen</i>
9	The empty room was fixed before the distance increased.	<i>Medial consonants in empty /pt/, fixed /kst/, distance /st/</i>
10	Though she was tired, she stopped to breathe slowly.	<i>/ð/ in though, breathe; final /pt/ in stopped</i>

Appendix B

Questionnaire

Please read each statement carefully and circle the number that best reflects your experience using ELSA Speak during the three-week pronunciation practice period. There are no right or wrong answers. Your responses are confidential and will not affect your course grade. Please respond to every item.

No.	Dimension	Statement (1 = Strongly disagree 2 = Disagree 3 = Neutral 4 = Agree 5 = Strongly agree)	1	2	3	4	5
1	Error Noticing	The AI feedback helped me notice specific errors in my pronunciation that I had not been aware of before.	☐	☐	☐	☐	☐
2	Error Noticing	After receiving feedback from ELSA Speak, I could identify the exact sound I had produced incorrectly.	☐	☐	☐	☐	☐
3	Error Noticing	The feedback made it easier to detect differences between my production and the target pronunciation.	☐	☐	☐	☐	☐
4	Error Noticing	I found the AI feedback useful for drawing my attention to my pronunciation errors.	☐	☐	☐	☐	☐
5	Iterative Practice	I repeated the same words or sentences several times after receiving negative feedback from the application.	☐	☐	☐	☐	☐
6	Iterative Practice	The immediate feedback encouraged me to keep trying until I produced the target sound more accurately.	☐	☐	☐	☐	☐
7	Iterative Practice	I practiced pronunciation items more times with the AI tool than I would have done in a regular classroom.	☐	☐	☐	☐	☐
8	Ease of Use	I found ELSA Speak easy to navigate and use for pronunciation practice.	☐	☐	☐	☐	☐
9	Ease of Use	The interface of the application was clear enough for me to focus on practicing pronunciation rather than operating the app.	☐	☐	☐	☐	☐
10	Ease of Use	The scoring displayed by the app was presented in a way I could follow.	☐	☐	☐	☐	☐
11	Motivation	Practicing with the AI tool increased my motivation to work on my English pronunciation.	☐	☐	☐	☐	☐
12	Motivation	I felt more confident about my pronunciation ability after using the AI feedback tool.	☐	☐	☐	☐	☐
13	Motivation	Receiving an improved score from the application encouraged me to continue practicing.	☐	☐	☐	☐	☐
14	Learner Autonomy	The AI tool allowed me to practice pronunciation outside of class at times that suited me.	☐	☐	☐	☐	☐
15	Learner Autonomy	I used the application for pronunciation practice beyond the time required by my teacher.	☐	☐	☐	☐	☐

(Continued)

No.	Dimension	Statement (1 = Strongly disagree 2 = Disagree 3 = Neutral 4 = Agree 5 = Strongly agree)	1	2	3	4	5
16	<i>Learner Autonomy</i>	I felt I could manage my own pronunciation practice sessions without needing constant teacher guidance.	☐	☐	☐	☐	☐
17	<i>Feedback Interpretability</i>	When the application marked my pronunciation as incorrect, I understood clearly what I needed to change.	☐	☐	☐	☐	☐
18	<i>Feedback Interpretability</i>	The feedback provided by the app was easy enough to interpret that I could act on it immediately.	☐	☐	☐	☐	☐
19	<i>Feedback Interpretability</i>	I sometimes found it difficult to understand why the application gave me a low score for a particular attempt.	☐	☐	☐	☐	☐
20	<i>Feedback Interpretability</i>	I would have benefited from additional explanation alongside the app's feedback to help me understand what to correct.	☐	☐	☐	☐	☐

Open Response Items

1. Which aspect of the AI feedback did you find most useful for improving your pronunciation? Please explain briefly.
Response: _____

2. Was there any type of feedback from the app that you found difficult to understand or act upon? Please describe.
Response: _____

Appendix C

Semi-Structured Interview

Thank you for agreeing to speak with me today. Everything you share in this conversation will remain strictly confidential and will have no bearing on your course grades. My interest is in your personal experience using the pronunciation application over the past three weeks. There are no right or wrong answers here. I simply want to understand what the experience was actually like for you, what worked, and what did not. Please feel free to respond in Vietnamese or English, whichever feels more natural.

Theme A: Error Noticing

- A1. When you received feedback from the app, what was the first thing you noticed or looked at?
- A2. Were there any pronunciation errors you only became aware of through using ELSA Speak? Could you walk me through one?
- A3. In what ways did the feedback help you see the gap between how you said something and how it should sound?

Theme B: Iterative Practice and Self-Correction

- B1. What did you typically do right after the app flagged your pronunciation as incorrect?
- B2. Did you ever record the same word or sentence more than three times in one session? What kept you going?
- B3. How did you know when your pronunciation of a particular item was good enough to move on?

Theme C: Learner Autonomy and Practice Beyond Class

- C1. Did you use the app outside of class time? If so, when and where did that usually happen?
- C2. How did you decide which sounds or words to work on during those independent sessions?
- C3. Did working with the app affect how you think about pronunciation practice more broadly?

Theme D: Feedback Interpretability and Difficulty

- D1. Were there times when you genuinely could not work out why the app gave you a low score? Can you describe a specific moment?
- D2. When you found the feedback hard to make sense of, what did you do next?
- D3. Looking across the different types of errors the app flagged, which were easiest to understand and act on? Which gave you the most difficulty?

Theme E: Category-Specific Perceptions

- E1. The app gave feedback on sounds at the ends of words, for instance the /t/ or /d/ in ‘walked’ or ‘loved’. How useful did you find that particular type of feedback?
- E2. Did you find vowel feedback harder to interpret and act on than feedback on consonants?
- E3. Were there sounds for which you felt the app’s judgment did not quite match what you thought you were actually producing?

Interview Closing

Before we finish, I have two final questions.

- F1. Is there anything about your experience with the app or with pronunciation practice during this period that feels important to mention but has not come up so far?
- F2. If you were advising a teacher on how to use this kind of AI pronunciation tool with students in a similar situation to yours, what would you want them to know?

Thank you very much for your time. What you have shared will contribute to a clearer understanding of how AI-assisted tools can support pronunciation development in university EFL settings.