Research Article

Evaluating the comparative effectiveness of CAPT-based vs. native speaker-led pronunciation training

Dana Ondrušková [1*]

🔗 0000-0001-9591-5673

[1] Palacký University, Olomouc, CZECH REPUBLIC
[*] Corresponding author: dana.ondruskova01@upol.cz

ARTICLE INFO	ABSTRACT
	This study examines the effectiveness of pronunciation training by comparing two instructional methods: native speaker-led instruction and artificial intelligence (AI)-driven pronunciation tools. The research aims to determine whether AI-assisted pronunciation training can serve as an effective alternative or complement to traditional instruction. This is particularly relevant for language programs where access to native speakers is limited. The study employs a quantitative research design, analyzing pronunciation accuracy, intelligibility, and listening comprehension among A1-level French learners. A statistical comparison of learners' performance in both groups was conducted, including Chi-square tests and standard deviation analysis. The findings suggest that AI-based pronunciation training is at least as effective as native speaker-led instruction, with students in the CAPT group performing even better, particularly in intelligibility and intonation. The reduced performance variability among AI-trained learners suggests that these tools provide a structured and uniform learning experience. These insights contribute to optimizing pronunciation teaching methods, highlighting the potential of AI as a scalable and accessible pronunciation training solution in foreign language education. **Keywords:** pronunciation training, AI-assisted learning, language education, listening comprehension

INTRODUCTION

Correct pronunciation is crucial for communicative competence, both in terms of perception and production. It plays a fundamental role in effective communication, as it directly impacts the message's clarity and understanding. Pronunciation not only ensures that speech is intelligible but also influences how speakers are perceived by others, affecting their credibility and engagement. When learners articulate phonemes and prosodic patterns correctly, they are better equipped to internalize acoustic cues and segment continuous speech, which directly supports listening comprehension (Dao et al., 2020). Mispronunciation often reflects underlying perceptual issues, improving pronunciation can concurrently sharpen auditory discrimination of sounds. Empirical findings suggest that targeted pronunciation training reduces cognitive load during listening, allowing learners to focus more on meaning rather than form (Verbeke & Simon, 2023). Consequently, integrating pronunciation practice into language curricula fosters not just speaking accuracy but also more effective and efficient listening skills (Ekşi & Yeşilçınar, 2016). Nevertheless, despite its recognized importance, pronunciation practice continues to receive limited attention in classroom practice, creating a gap that this study seeks to address.

Traditionally, pronunciation practice has been facilitated by native speakers, who provide real-time feedback, corrective intervention, and nuanced phonetic modeling. However, recent advancements in artificial intelligence (AI) have introduced new tools suitable for analyzing and correcting pronunciation with increasing precision. These AI-driven solutions offer learners the possibility of autonomous, immediate, and repetitive practice, potentially reducing their dependency on human instructors. While AI-powered software can provide detailed phonetic analysis, highlight mispronunciations, and suggest corrections, it lacks the

social and contextual elements that human instructors naturally incorporate. Engaging with native speakers is widely recognized as a beneficial method for enhancing language proficiency (O'Brien, 2020). Interactions with native speakers are supposed to be beneficial, but access to qualified speakers is not always feasible, and not all native speakers are trained to provide structured feedback. Pronunciation training with a native speaker has traditionally been regarded as the most effective approach in foreign language acquisition. Direct interaction further exposes students to real-life speech variations, including rhythm, stress patterns, and connected speech, enhancing their overall communicative competence (Chacón, 2022). While these constraints characterize native speaker-led instruction, recent advances in AI have introduced alternative solutions that warrant systematic examination (Huang et al., 2023). While these platforms offer potential benefits, they also present certain limitations, similar to those encountered in native speaker-led instruction. Firstly, many AI-based learning tools lack a structured framework, making it challenging for learners to follow a clear progression in improving their pronunciation and overall language proficiency, potentially leading to learning gaps. Secondly , native speakers available on such platforms are often not trained to provide systematic feedback, which may inadvertently reinforce incorrect pronunciation patterns (Ngo et al., 2024). These challenges underscore the importance of rigorous empirical research comparing AI-based training with traditional native speaker instruction. Therefore, it is necessary to assess whether AI-based pronunciation training can achieve results comparable to instruction with native speakers through a comparative analysis of both approaches, thereby ensuring the effective use of limited class time.

This study aims to address a gap in research on the use of text-to-speech tools in comparison with native-speaker instructors for pronunciation training. Through qualitative research and data collection, it compares and evaluates the pronunciation outcomes of two learning groups practicing French sentences at the A1 level. Two key research questions (RQs) guide our investigation:

RQ1: How do CAPT tools and native speaker-led pronunciation training compare in their effectiveness in improving pronunciation accuracy and listening comprehension among A1-level French learners?

RQ2: How does the variability in learners' performance differ between CAPT-based training and native speaker-assisted instruction, and what implications does this have for the adoption of CAPT tools in foreign language education?

LITERATURE REVIEW

Using Information and Communication Technology tools in language learning

While existing literature describes the potential benefits and limitations of both teaching methods, focuses on the potential of computational linguistics and speech technologies, and automated language evaluation but lacks a comparative analysis based on real data from training groups and their achieved outcomes. Previous studies have concluded that the use of information and communication technology (ICT) tools in pronunciation training fosters positive attitudes toward technology-mediated instruction among both learners and instructors. With advancements in digital technologies and the internet, they are becoming increasingly valuable in education (Sullivan & Czigler, 2002). ICT tools may foster constructive learning more effectively than traditional teaching methods (Bester & Brand, 2013). These new technology-mediated learning activities make lessons more engaging and interactive while encouraging learner autonomy and serving as powerful sources of motivation for language learners (Cardozo & Orta, 2012). Moreover, learning activities supported by technology can help reduce anxiety and create a more comfortable environment for learners as they practice pronunciation and intonation without feeling exposed or becoming the center of attention in the classroom (Chacón, 2022). Given these challenges, language learners can turn to alternative resources such as language exchange platforms, online tutoring services, and language learning applications, which offer structured lessons and feedback. These tools can supplement interactions with native speakers and provide a more comprehensive learning experience (Amrate & Tsai, 2025). Recent studies have found that pronunciation practice in the classroom is often met with reluctance from students, who may feel embarrassed or find traditional exercises monotonous (Tsunemoto et al., 2020). Therefore, it is essential to explore methods that make pronunciation training more engaging, efficient, and an integral part of language learning, with a focus on intelligible pronunciation.

Limitations of CAPT in Pronunciation Training

In educational contexts, the integration of technology often evokes hesitation and skepticism, as many scholars remain unconvinced about the extent to which CAPT systems can effectively foster accurate pronunciation and contribute to learners' overall comprehension (Amrate & Tsai, 2024; Revell-Rogerson, 2021). While modern language teaching approaches emphasize more achievable goals such as intelligibility and comprehensibility, many CAPT systems still rely on comparing second language learners' pronunciation to that of native adult speakers (O'Brien, 2020). This often results in automatic speech recognition (ASR) failures, particularly with accented second language speech (e.g., Henrichsen, 2021; Martin & Wright, 2023) or children's speech (e.g., Gelin et al., 2021). Such challenges have led to doubts about the effectiveness of ASR in assessing second language speech and sparked interest in non-ASR alternatives (Fontan et al., 2022). Additionally, CAPT feedback is frequently regarded as a technological feature rather than a pedagogically grounded tool, limiting its reliability in identifying second language pronunciation issues and adapting to individual learners' needs (Fontan et al., 2018). These criticisms and concerns highlight a significant research gap, underscoring the need for further studies that compare the outcomes of both CAPT learning and pronunciation practice with native speakers to evaluate their relative effectiveness. Building on these critical observations, this study designs an assessment procedure that systematically compares both approaches under controlled classroom conditions.

Assessment and Evaluation of Pronunciation Progress

Empirical research indicates that optimal pronunciation learning outcomes are achieved through a combination of implicit feedback and explicit instruction (Saito & Lyster, 2012). In accordance with contemporary pedagogical frameworks, pronunciation feedback should be strategically directed toward errors that most profoundly affect intelligibility and comprehensibility. Assessment constitutes a fundamental dimension of pronunciation instruction, encompassing a range of theoretical constructs, methodological approaches, and evaluative measures. A meta-analysis of second teaching language pronunciation studies conducted by Saito and Plonsky (2019) delineates a comprehensive model of pronunciation assessment, incorporating key dimensions such as the nature of assessed constructs, the mode of speech elicitation, and the evaluation method employed.

Aligned with pedagogical objectives, contemporary pronunciation assessment frameworks prioritize communicative relevance, favoring the evaluation of overarching pronunciation characteristics such as intelligibility and comprehensibility or the specific phonetic features that underpin these dimensions (Derwing & Munro, 2015). The selection of appropriate assessment tasks is contingent upon the particular pronunciation features under scrutiny, necessitating a methodological alignment between assessment design and instructional focus. Amid the ongoing discourse on the effectiveness of native speaker-assisted pronunciation training versus technology-driven approaches, a systematic examination of pronunciation teaching methodologies through empirical research is imperative. While existing reviews acknowledge the potential advantages of CAPT (e.g., Mahdi & Al Khateeb, 2019; Ngo et al., 2024), they fall short of providing a comprehensive comparative analysis that elucidates how learning outcomes, particularly in terms of learners' comprehension, differ between CAPT-based instruction and native speaker-led training.

METHODS

The present research focuses on beginner-level French learners studying at secondary schools. The acquisition of French as a second foreign language presents specific pronunciation challenges due to the phonetic complexities of the language (Chinedu & Anthonia, 2015). The connection between words in spoken French (*liaison*), the significant difference between written and spoken forms, and the presence of numerous silent letters create difficulties for learners. These challenges often discourage students in the early stages of learning, making them feel frustrated and less motivated to continue their studies. Without perceivable progress, students may disengage from the learning process entirely.

The research aims to compare two approaches to pronunciation training CAPT-based tools and native speaker-assisted practice to see how they differ in their impact on students' ability to understand spoken language and how understandable their own spoken language becomes. By integrating quantitative methods,

this study aims to present a comprehensive understanding of the effectiveness of CAPT tools in comparison to native speaker-led pronunciation training. To address these questions, this study evaluates and compares the effectiveness of AI-driven pronunciation training with traditional instruction from native speakers.

Participants

The study targeted high school students learning French as a second foreign language at the A1 level according to the common European framework of reference for languages. Participants were selected from five different schools across various regions of the Czech Republic, where native French-speaking instructors were available. The goal was to ensure a diverse sample while maintaining a consistent language proficiency level among students (Olsen et al., 2013) . At the beginning of the school year in September 2023, first-year students with no prior knowledge of French were identified. Each school provided a group of 10 to 16 students who participated in the study, with a total of 70 students involved. These students, aged 15-16, had already attained at least an A2 level in English, ensuring a foundational understanding of foreign language learning strategies (Cengiz, 2023). The participants were evenly distributed into two groups: 35 students in the native speaker group, receiving traditional pronunciation instruction from a native French speaker, and 35 students in the technology-based group, practicing pronunciation using CAPT tools. Given the limited number of schools that offer French in sufficiently large groups, have access to native speakers, and were willing to participate in this research, the sample size of 70 participants may seem small. However, this number is constrained by the factors mentioned earlier, and even a moderate sample can provide valuable preliminary findings that highlight the strengths and weaknesses of both AI-based pronunciation tools and native speaker-led instruction. The inclusion of schools from five different regions ensures that findings are not limited to a single institution's teaching methods, allowing for broader insights into the potential scalability of AI pronunciation training. This participant selection process ensures that the study captures meaningful data on the effectiveness of AI tools in pronunciation training while considering the practical realities of high school foreign language education. The lessons took place three times per week, with each class lasting 45 minutes. Twice a week learners dedicate 2 × 20 minutes to intensive pronunciation practice, either with a native speaker or with CAPT based tools following each school's standard curriculum.

Teaching Procedures

Throughout this research, students in the native speaker group (group A) followed their lessons with their French teacher, where they learned vocabulary, grammar, and phrases. In addition to this standard instruction, they participated in specialized pronunciation training led by a native French-speaking instructor. These dedicated sessions focused on refining phonetic accuracy, articulation, and prosody, providing students with direct exposure to native pronunciation models and immediate, tailored feedback. The instructor adjusted their teaching strategies based on individual students' needs, helping them overcome pronunciation challenges specific to the French language as a second foreign language. The study was conducted in actual classroom environments rather than making the results more applicable to real-world language learning.

Students in the non-native speaker group (group B) received instruction from their French teacher in the same manner as in group A. However, for specialized pronunciation training, they utilized CAPT tools in dedicated sessions conducted twice a week for 20 minutes. During these sessions, each student worked individually at a computer, engaging with pre-assigned words, sentences, and model dialogues. They listened to these recordings, repeated them, and received automated feedback from the CAPT system, which analyzed their pronunciation and provided corrective responses. This approach enabled learners to independently refine their pronunciation by comparing their speech to the target model and making necessary adjustments based on the system's guidance. The methods applied to both groups are summarized in **Table 1**. In addition to these parallel instructional designs, special attention is devoted to the CAPT tool used in this study which is described in detail below.

A Tool for CAPT Learning

For the present study, Speechling was selected as an advanced educational platform designed to enhance pronunciation, listening comprehension, and fluency in multiple languages, including French. It integrates cutting-edge speech recognition technology with human-assisted feedback to provide a structured and

Table 1. Learning methods

Group A	Group B
Targeted pronunciation training with a native speaker (20 minutes twice a week in regular lessons)	Targeted pronunciation training with CAPT method (20 minutes twice a week in regular lessons)
Theoretical explanation	
Explanation of key pronunciation concepts: Stress patterns, intonation, rhytm, articulation	Theoretical aspects of pronunciation: Including phonetic symbols, sound patterns, the importance of clarity in spoken language
Practical activities	
Repeat words, phrases, practice sounds that are difficult for non-native speakers, the native speaker says sentences or short passages, and learners respond or summarize what they've understood	Interactive listening, pronunciation exercises, and CAPT speech recognition
Correction method	
Immediate correction from the native speaker, feedback on pronunciation errors and models correct pronunciation Example: *"Careful with the nasal vowel in* **vingt**. *Don't pronounce the final 't' and let the sound come through your nose: /vɛ̃/.In* **huit**, *you need to pronounce both the /ɥ/ glide and the /i/ vowel. Try rounding your lips, then glide into /i/: /ɥit/.Put it all together: elle a vingt-huit ans. Listen and repeat after me.*	*Automatic feedback on pronunciation and intonation Example: "Pronunciation error detected: vingt. Correct pronunciation: /vɛ̃/. Repeat after me: elle a vingt-huit ans.* The app plays the correct recording). No detailed explanation of how to form the nasal vowel — just identifies the error and supplies the target model.

interactive learning experience. The platform's pedagogical approach focuses on immersion, repetition, and corrective feedback, essential elements for mastering phonetics and pronunciation. One of Speechling's core features is its pronunciation training tool, which allows learners to record their speech and compare it to native speaker models. By utilizing AI-powered speech recognition, the platform analyzes user recordings and provides instant visual and auditory feedback, highlighting pronunciation discrepancies. This feature aids in developing accurate articulation by reinforcing proper phonetic structures. In addition to pronunciation refinement, Speechling enhances listening comprehension through its extensive library of professionally recorded sentences. These recordings enable learners to familiarize themselves with native intonation, rhythm, and phoneme articulation, which are crucial for improving auditory skills. The spaced repetition system ensures that learners are consistently exposed to phrases at optimal intervals, reinforcing memory retention. **Figure 1** presents records of pronunciation training sessions and feedback.

An important aspect of Speechling is its human tutor feedback system. Users on the free plan can submit recordings for evaluation, and certified language coaches provide constructive feedback, pinpointing specific pronunciation issues and offering tailored correction strategies. This personalized approach facilitates accelerated learning by directly addressing individual challenges. In this study, the feature was implemented by having participants record the target sentences within the Speechling platform after each practice session. These recordings were automatically forwarded to Speechling coaches, who then supplied written and audio comments within 24 hours. The feedback was integrated into the learners' dashboards, allowing them to replay the corrected model, compare it with their own attempt, and repeat the sentence until satisfactory performance was achieved.

Assessments Procedures

After the five-month study, the effectiveness of both native speaker-led instruction and AI-assisted training was assessed using a combination of quantitative and qualitative analyze. Each student's performance was evaluated anonymously to ensure objective results.

To evaluate the effectiveness of pronunciation training with a native speaker compared to a CAPT tool, all students were tested similarly after five months of instruction. The native speaker was an invited French language lecturer (a native speaker not affiliated with the participating schools) who had not practiced pronunciation with the students and served exclusively as an external evaluator of their pronunciation.

The assessment focused on two key categories:

1. Pronunciation accuracy
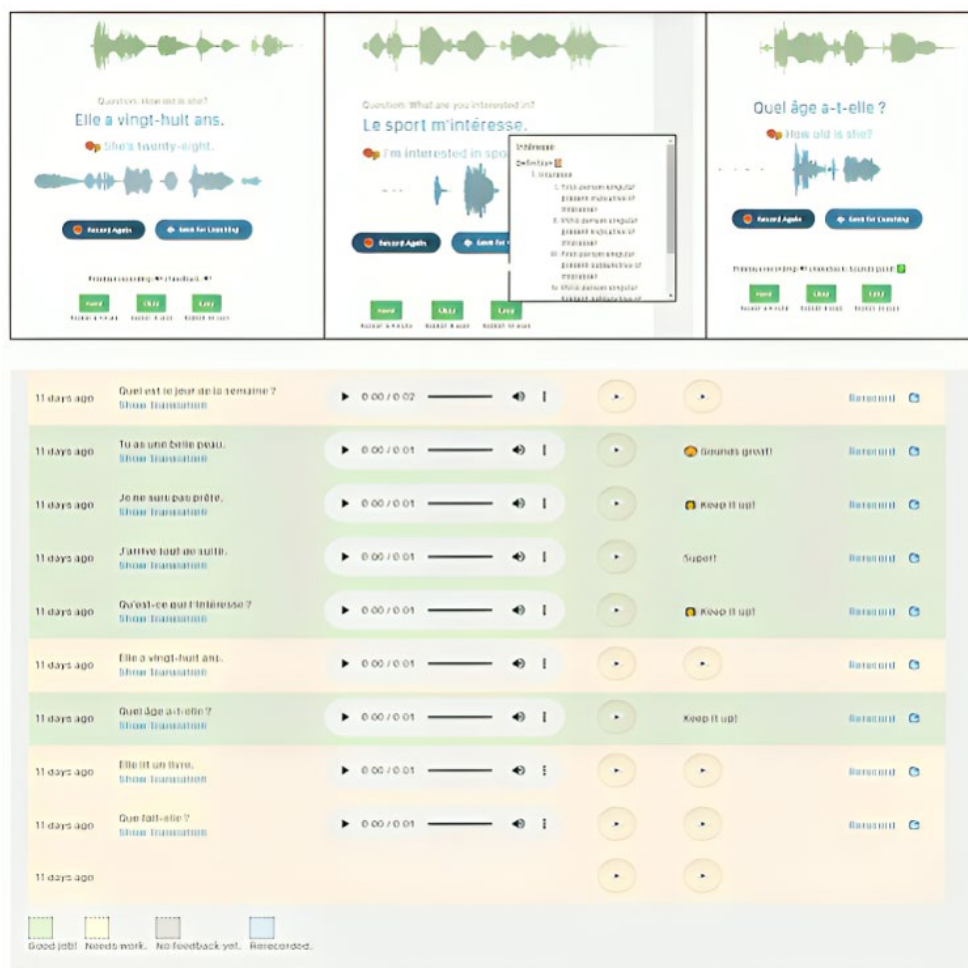2. Listening comprehension

Figure 1. The CAPT tool environment Speechling (Source: Results of the present research)

Pronunciation accuracy assessment–Native speaker evaluation

Students were asked to read aloud 5 familiar and 5 unfamiliar sentences, all corresponding to the A1 proficiency level. All test participants read the same sentences. Pronunciation accuracy was evaluated based on the following three criteria:

1. Intelligibility–the extent to which the pronunciation was understandable to a native speaker
2. Phonetic accuracy–correct articulation of individual sounds and appropriate linking between words
3. Intonation–appropriate use of pitch and rhythm in sentences.

Each sentence was scored on a three-point scale: 3 points–all three criteria met, 2 points–two criteria met, 1 point–one criterion met, 0 points–none of the criteria met.

Each student's pronunciation was also assessed using an AI-based pronunciation analysis tool. The program provided feedback in two categories: "good job!" for accurate pronunciation and "needs work" for areas requiring improvement.

Listening comprehension assessment

For the listening comprehension assessment, students were provided with a text containing missing words. The text was read aloud twice by the teacher (whom the students are accustomed to), and students were required to fill in the 10 missing, familiar words. Each missing word was worth 1 point, awarded for correct completion. The student's total score was calculated based on the number of correct words they inserted into the text, with a maximum possible score of 10 points. To complement this evaluation, statistical analyses were conducted to examine central tendencies, variability, and group differences in greater depth.

Table 2. A table on top

	Group A		Group B	
	Assessment by NS	Assessment by CAPT	Assessment by NS	Assessment by CAPT
M	2.0	2.6	2.1	2.8
SD	0.939336436627	0.694516336148	0.95089519962	0.83816526318

Statistical Analysis of Results

Firstly, the descriptive statistics summarize the data the mean (M), standard deviation (SD), and range of pronunciation accuracy scores across three evaluation conditions. It provides insights into central tendencies and variability. Through a systematic analysis of pronunciation training outcomes, the study will evaluate the effectiveness of both instructional methods of native speaker-led training and CAPT-based practice. Given the nature of this study, results will be generalized based on an inductive approach, deriving conclusions from the collected empirical data rather than testing a predefined hypothesis. The analyses will allow for a comprehensive assessment of the strengths and limitations of each approach, facilitating the formulation of generalized conclusions regarding their impact on learners' pronunciation development. Based on these findings, appropriate recommendations will be proposed to guide future instructional strategies, ensuring the selection of the most effective pronunciation training methods in foreign language education. These methods focused on:

(1) quantitative analysis of pronunciation accuracy and listening comprehension scores, including M comparisons and Chi-square tests addressing RQ1 and

(2) statistical evaluation of performance variability through SD analysis and comparative assessments between CAPT-based training and native speaker-led instruction addressing RQ2.

RESULTS

To evaluate the overall effectiveness of both instructional approaches, the study first examined learners' pronunciation outcomes. Since pronunciation constitutes a central component of oral proficiency, accuracy in segmental and suprasegmental features provides a meaningful indicator of progress. Particular attention was devoted to intelligibility, phonetic accuracy, and intonation, as these elements directly influence how learners are perceived and understood by interlocutors. The following section presents the comparative results of both groups in terms of pronunciation accuracy, before turning to listening comprehension performance.

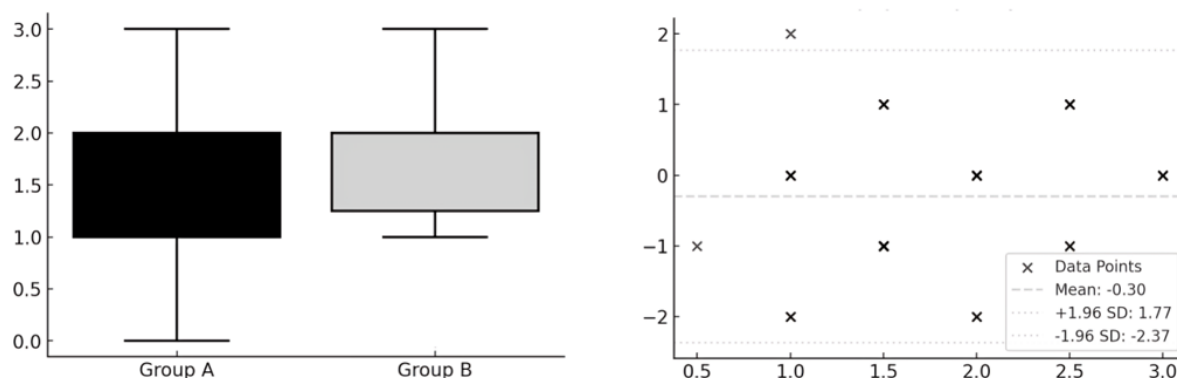### Pronunciation Accuracy Assessment

The following results compare the pronunciation accuracy scores of both groups A and B. Each group's pronunciation was evaluated by both native speakers and CAPT models, allowing for an in-depth comparison of scoring discrepancies.

The initial analysis serves to verify the consistency and reliability of the assessments conducted by the native speaker and AI tools. The findings of the descriptive statistics are presented in **Table 2**.

The CAPT tool gives slightly higher scores than the native speaker assessment for group A. This suggests that CAPT tools may be more lenient or focus on different criteria compared to human assessment. The lower SD in the CAPT assessment indicates that CAPT tools evaluate learners more consistently, whereas NS evaluations have more variation. The CAPT tool rates group B higher than NS assessment. This indicates some discrepancy between human and automated evaluation. The gap suggests that CAPT tools may overestimate learners' listening comprehension skills when they are trained within the same system. In some cases, the CAPT tools and native speakers assess listening comprehension differently. CAPT tools seem to assign higher scores, especially for learners trained within the system. Native speakers may evaluate with stricter, more nuanced criteria that CAPT does not fully capture. CAPT training may create artificially high scores within its own system, but native speakers perceive their comprehension as significantly lower. CAPT tools tend to give higher and more consistent scores, especially for CAPT-trained learners. Subsequent analyses focus on the comparative evaluation of the findings from both group A and group B. The results of the comparison in terms

Table 3. Comparison of pronunciation accuracy between group A and group B

	Group A	Group B
M	1.63	1.93
SD	0.93	0.78

Figure 2. Box and Bland-Altman plots-1 (Source: Results of the present research)

of pronunciation accuracy, focusing on intelligibility, phonetic accuracy, and intonation, as assessed by a native speaker for group A and group B, are displayed in **Table 3**.

Group B achieved a higher mean pronunciation accuracy score (1.93) than group A (1.63), with a lower SD (0.78 vs. 0.93). This indicates that CAPT training not only produced slightly better outcomes but also ensured greater consistency among learners. The structured, repetitive, and standardized input provided by CAPT systems such as Speechling likely contributed to this uniformity, minimizing variability in learner performance. In contrast, native speaker-led instruction, though authentic and interactive, introduced greater variation, possibly due to differences in teaching style, spontaneous corrections, and individual learner responses. The statistical analysis further confirmed that CAPT learners outperformed their peers in intelligibility and intonation (p < 0.001), suggesting that automated feedback and repetitive practice may be particularly effective in these domains.

Nevertheless, these results should be interpreted cautiously. CAPT assessments may favor learners trained within the same system, potentially inflating performance scores relative to human evaluation. Moreover, native speakers provide valuable contextual and communicative input that CAPT tools cannot replicate. While CAPT appears advantageous in terms of accuracy, consistency, and scalability, further research is needed to evaluate whether these benefits translate into improved performance in spontaneous, real-world communication.

The graphical representation of the results is shown in **Figure 2**.

The boxplot shows the distribution of pronunciation accuracy scores for group A. Both groups have a similar range, but group B's scores appear more concentrated, indicating less variability. Group A has a wider spread, meaning that some individuals performed significantly less or higher compared to the rest of the group. Group B has a lower outlier, suggesting that while most participants performed consistently, some had difficulties. The Bland-Altman plot visualizes the agreement between group A and group B. The mean difference (-0.30) is close to zero, suggesting that both groups performed similarly overall. The limits of agreement (±1.96 SD) range from -2.37 to 1.77, meaning that most differences between the groups fall within this range. These results suggest that both pronunciation training methods yielded comparable overall performance, though the consistency of improvement may vary between groups.

The paired sample T-test was conducted to compare the M scores of group A and group B across the three pronunciation categories. In that case, a significant p-value (p < 0.05) suggests that one group significantly outperforms the other. Intelligibility and intonation show significant differences; group B is likely to exhibit stronger skills in these areas. The phonetic accuracy scores are identical for both groups, which confirms no significant difference. Again, the focus is on intelligibility, phonetic accuracy, and intonation. Furthermore, a Chi-square test was conducted to compare the frequency distributions between the groups (**Figure 3**).
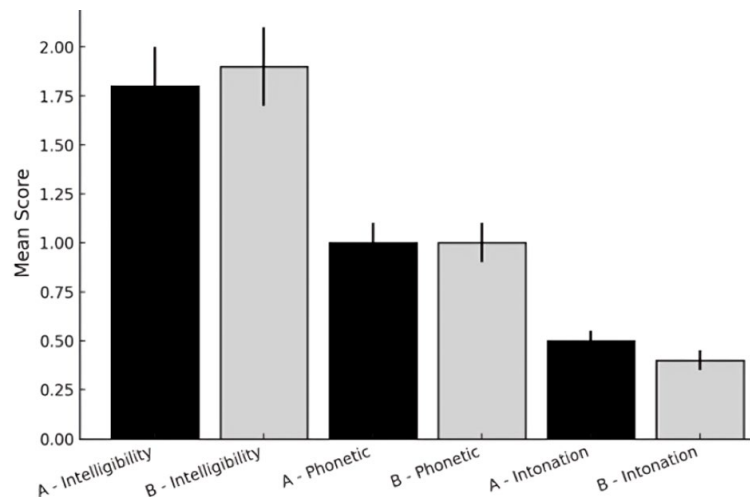
Figure 3. Mean scores with confidence intervals (Source: Results of the present research)

Table 4. Intelligibility, phonetic accuracy, and intonation

	p	Explanation
Intelligibility	< 0.001	Statistically significant difference between groups.
Phonetic accuracy	1.00)	No significant difference, confirming both groups performed equally.
Intonation	< 0.001	Statistically significant difference, suggesting one group demonstrated stronger intonation.

Table 5. Comparison of listening comprehension performance between group A and group B

	Group A	Group B
M	5.609756097560975	6.609756097560975
Median	5	7
SD	2.73749930393130	2.2458634061368
Min & max	1 & 10	2 & 10

Group B outperformed group A in intelligibility and Intonation, as indicated by the significant Chi-square results (**Table 4**).

Listening Comprehension Assessment

Further attention in the evaluation of data analyses is directed toward the second assessment, which focuses on listening comprehension (**Table 5**).

Slightly higher M score group B (6.61) suggests that students using CAPT tools performed slightly better in listening comprehension than those trained with native speakers. Group A shows higher SD (2.74), indicating greater variability among learners. Lower SD (2.25) of group B indicates their scores are more consistent and less widespread. This suggests that while some learners in group A improved significantly, others struggled, possibly due to individual differences in interaction with native speakers. Specifically, the native speaker group (group A) showed a higher SD in both pronunciation accuracy (SD = 0.93) and listening comprehension (SD = 2.74) compared to the CAPT group (SD = 0.78 and SD = 2.25, respectively). This broader spread of scores suggests that while some learners benefited substantially from native speaker-led training, others improved less, leading to greater inconsistency.

The study interprets this variability as potentially linked to the interpersonal and pedagogical dynamics inherent in human-led instruction–factors such as how actively learners engaged with the instructor, their comfort with immediate correction, or the extent to which they could adapt to spontaneous interaction. While this explanation remains tentative, it aligns with prior literature noting that learner responses to corrective feedback and teacher interaction differ widely (Lyster et al., 2013). Thus, the claim reflects an interpretation grounded in the statistical evidence of variability, rather than a directly observed measure of individual differences.
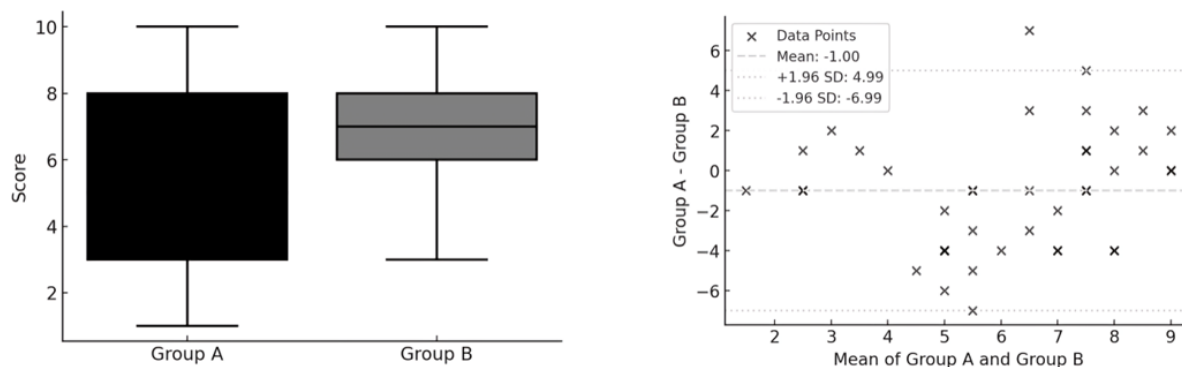
Figure 4. Box and Bland-Altman plots-2 (Source: Results of the present research)
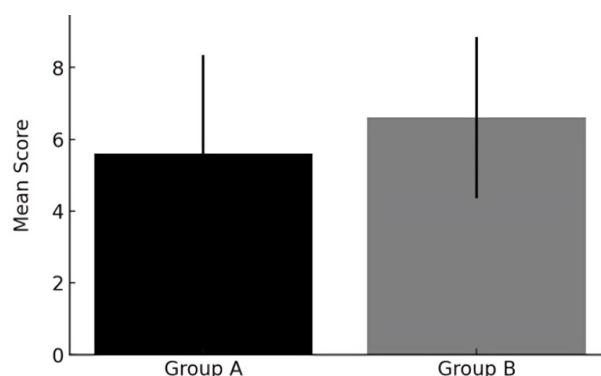

Figure 5. Mean listening comprehension scores for group A and group B (Source: Results of the present research)

In contrast, CAPT tools provide more uniform improvement across learners. Group A's median score (5.0) is lower than group B's (7.0), meaning most students in group B achieved better results. The graphical representation of the results is shown in **Figure 4**.

The plot indicates that for some values, differences between group A and group B are larger, particularly at higher Ms. The Bland-Altman plot visualizes the agreement between group A and group B. The mean difference (-0.49) is close to zero, suggesting that the two groups are similar. The limits of agreement (±1.96 SD) range from -6.64 to 5.66, which means that most differences fall within this interval.

Figure 5 shows the mean listening comprehension scores for group A and group B.

These empirical results provide the basis for a broader discussion of the comparative effectiveness of CAPT and native speaker-led instruction, as outlined in the following section.

DISCUSSION

Analysis of Results

The findings of this study provide critical insights into the comparative effectiveness of CAPT tools and native speaker-led pronunciation training, addressing both RQ1 and RQ2. Specifically, RQ1 explores the extent to which these instructional methods differ in enhancing pronunciation accuracy and listening comprehension among A1-level French learners, while RQ2 examines variability in learners' performance across the two training modalities and its implications for language education.

The quantitative analysis reveals that group B, which underwent CAPT-based training, exhibited marginally higher pronunciation accuracy (M = 1.93, SD = 0.78) compared to group A, which received instruction from a native speaker (M = 1.63, SD = 0.93). The SD values indicate that CAPT-based instruction resulted in a more consistent performance among learners, whereas the native speaker-led training group demonstrated a broader range of individual outcomes. This suggests that CAPT tools provide a more standardized approach to pronunciation training, ensuring uniform progression across learners with different linguistic backgrounds

and learning paces. Further statistical examination using the Chi-square test confirms significant differences between the two instructional approaches in terms of intelligibility ($p < 0.001$) and intonation ($p < 0.001$), with CAPT-based training yielding superior outcomes in both areas. These findings support the hypothesis that structured AI-driven pronunciation training offers enhanced control over phonetic input, leading to greater improvements in these aspects of pronunciation. However, phonetic accuracy did not exhibit any statistically significant difference between the two groups ($p = 1.00$), implying that both CAPT tools and native speaker-led instruction were equally effective in this domain. These results partially answer RQ1, as they suggest that while CAPT tools outperform native speaker-led instruction in some aspects of pronunciation training, both methods are equally effective in phonetic precision.

Listening comprehension results followed a similar pattern. CAPT-trained learners in group B achieved a higher mean score (M = 6.61) compared to learners in group A (M = 5.61). Furthermore, group B demonstrated a lower SD (SD = 2.25) than group A (SD = 2.74), further supporting the notion that CAPT-based training leads to more uniform improvements across learners. The reduced variability observed in group B indicates that CAPT tools provide a structured and predictable learning environment that mitigates disparities in learner performance. This directly addresses RQ2 by demonstrating that CAPT-based training results in lower inter-individual variability, which is a critical factor in instructional design for diverse student populations. In contrast, the broader range of scores in group A suggests that native speaker-led instruction may be influenced by additional factors such as instructor variability, individual differences in learner interaction, and the level of personalized feedback provided.

The implications of these findings extend beyond the immediate comparison of instructional effectiveness. The structured nature of CAPT-based training ensures consistency in pronunciation input and feedback, which is particularly beneficial for large and heterogeneous classrooms where learners may have varying degrees of phonetic awareness and auditory discrimination abilities. In contrast, while native speaker-led instruction provides authentic communicative experiences, the variability in learner outcomes highlights potential challenges in ensuring uniform pronunciation and listening comprehension improvements.

Overall, these findings indicate that CAPT tools can serve as a viable and in some respects advantageous alternative to native speaker-led pronunciation training, particularly with respect to consistency of outcomes; however, this inference should be regarded as provisional given the design and measurement constraints. However, while CAPT-based training fosters more consistent performance across learners, it is essential to recognize that human-led instruction provides valuable exposure to the variability of natural speech, which may play a crucial role in developing real-world communicative competence. Future research should explore whether CAPT-trained learners encounter greater difficulties in adapting to spontaneous and dynamic speech environments, thereby assessing the long-term effectiveness of these tools in fostering holistic language proficiency. As scoring modalities differed (human rating vs. system-internal scoring), cross-method comparability should be interpreted cautiously; future work should incorporate common rubrics, multiple blinded human raters, and cross-system validation to strengthen causal inferences.

The Potential of CAPT Tools as a Substitute for Native Speaker Instruction

Taken together, these findings highlight important implications for language pedagogy and educational policy, which are elaborated in the subsequent section. Taken together, our findings suggest that CAPT tools may function as a viable alternative or partial substitute for native speaker-led pronunciation training under the conditions of this study (A1 learners, five-month intervention, specified tasks). The comparable or slightly better results of group B in pronunciation accuracy and listening comprehension demonstrate that well-designed CAPT tools provide high-quality pronunciation models, corrective feedback, and structured repetition, which are essential for pronunciation development. One possible explanation for CAPT's effectiveness is the absence of disruptive variables that naturally occur in face-to-face instruction. CAPT tools provide consistent and unbiased feedback, allowing learners to repeatedly practice specific pronunciation patterns without external distractions. In contrast, native speaker instruction may involve more situational variability, spontaneous interactions, and real-time corrections, which, while beneficial for naturalistic learning, could lead to inconsistent training outcomes.

Additionally, CAPT tools eliminate the influence of accent variation, background noise, and conversational unpredictability, which are integral components of human communication. While group A learners benefited

from authentic conversational exchanges, group B's controlled exposure to native pronunciation models through CAPT tools led to more uniform improvements in pronunciation and comprehension. However, this raises the question of whether CAPT-trained learners would face greater challenges when transitioning to real-world communication, where speech is less structured and more dynamic. The explanation for why CAPT-trained learners scored higher and native speaker-led learners lower must be approached cautiously, as the evidence does not allow for definitive causal claims. The article itself shows that group B (CAPT) achieved a higher mean pronunciation accuracy (M = 1.93) than group A (M = 1.63) and that CAPT learners demonstrated significantly better results in intelligibility and intonation (p < 0.001). These quantitative findings support the conclusion that CAPT can be effective.

The study also highlights that CAPT evaluations tend to be more lenient and internally consistent, whereas native speaker judgments introduce greater variability. This raises the possibility that scoring differences–rather than inherent superiority of one method–contributed to the results. Furthermore, native speakers provide contextual and communicative input that CAPT tools cannot replicate, meaning their perceived "lower scores" may reflect stricter evaluative standards rather than weaker instructional impact. Therefore, while the data suggest that CAPT training produced more consistent learner gains, claims about why this occurred should remain provisional. Stronger evidence would require methodological controls such as multiple raters, standardized scoring rubrics across conditions, or longitudinal testing of communicative competence beyond structured tasks.

Limitations and Considerations

Beyond the methodological limitations already noted, several contextual constraints should be acknowledged. First, the study was conducted with a relatively small sample (N = 70) of secondary school learners in the Czech Republic, all at the A1 level of French. While this design ensured homogeneity of proficiency, it also restricted the generalizability of findings to other proficiency levels, age groups, or cultural contexts. Second, the intervention period lasted five months, which may not capture longer-term retention or the durability of CAPT-related gains in real-world communicative settings. Third, the assessment design relied primarily on structured reading and fill-in-the-blank tasks; these measures provide valuable insights into accuracy and comprehension but may not fully represent spontaneous spoken interaction. Finally, while both groups followed parallel instructional procedures, variables such as learner motivation, prior exposure to technology, or rapport with instructors were not systematically controlled, and these may have influenced outcomes.

By recognizing these constraints, the study underscores that its findings should be interpreted as exploratory rather than conclusive. Future research would benefit from larger and more diverse samples, longitudinal designs, and assessment methods that include authentic communicative tasks alongside controlled pronunciation measures. Despite the strong performance of CAPT-based training, it is important to recognize the limitations of AI-driven pronunciation tools in comparison to human instruction. While CAPT tools effectively assess pronunciation accuracy and provide corrective feedback, they lack the ability to offer personalized, targeted corrections based on real-time interaction. A native speaker instructor can identify specific pronunciation errors and provide immediate, detailed feedback tailored to individual students' needs, whereas AI-based tools typically evaluate pronunciation holistically without pinpointing specific issues.

Another critical limitation of CAPT tools is their inability to simulate real-life conversational dynamics. Human instructors naturally adapt their speech to students' proficiency levels, incorporate contextual variations, and facilitate interactive role-playing scenarios that mirror real-world communication. In contrast, CAPT systems lack spontaneous adaptability, focusing instead on predefined phonetic models and scripted exercises. This suggests that while CAPT tools are highly effective for structured pronunciation training, they may not fully prepare learners for the fluid and unpredictable nature of face-to-face interactions. Additionally, CAPT tools do not account for dialectal variations, speech hesitations, and contextual speech patterns, which are crucial components of authentic communication. In contrast, exposure to native speakers allows learners to experience regional accents, informal speech patterns, and conversational nuances, all of which are vital for developing natural fluency.

The validity of comparing native speaker-led instruction and CAPT-based training depends critically on the scoring methods employed. In this study, pronunciation outcomes were assessed both by a native speaker

evaluator and by the CAPT tool itself. As the results indicate, CAPT tools consistently assigned higher and more uniform scores, particularly for learners trained within the system, while native speakers evaluated more strictly and with greater variability. This discrepancy suggests that the two assessment modes may not rely on identical criteria: CAPT tends to emphasize alignment with its own phonetic models, whereas native speakers apply more nuanced judgments that incorporate intelligibility, intonation, and subtle prosodic features.

Implications and Significance of the Study

The significance of this research lies in its initial evidence that CAPT tools may constitute a highly viable alternative in classroom settings, while not supplanting the pedagogical value of interactions with native speakers**.** Given that the pronunciation and comprehension outcomes of CAPT trained learners were on par with those of students instructed by a native speaker, this study provides empirical support for integrating AI-based pronunciation tools into language curricula. From a practical standpoint, CAPT tools present an accessible and scalable solution for language programs that lack access to native speaking instructors. This is particularly relevant for educational institutions where finding qualified native speakers seems to be challenging due to logistical constraints. The consistency of CAPT training ensures that all learners receive uniform, high-quality pronunciation practice, reducing disparities caused by differences in teaching styles or instructor availability.

However, while CAPT tools are demonstrably effective for pronunciation practice, they should be viewed as a complementary tool rather than a complete replacement for native speaker interaction. To achieve well-rounded communicative competence, learners should have opportunities to engage in interactive speaking exercises, spontaneous dialogues, and real-world language exposure in addition to structured pronunciation practice.

CONCLUSION

This study provides a comparative analysis of CAPT tools and native speaker-led pronunciation training among A1-level French learners. Results indicate that CAPT training produced comparable–and in some respects superior–outcomes, particularly in intelligibility and intonation, while also yielding lower performance variability. These findings suggest that CAPT systems can deliver consistent, structured feedback that benefits diverse classrooms where equal access to native speakers may not be feasible. At the same time, the study highlights important limitations. Although CAPT appears effective for controlled measures of pronunciation and listening comprehension, it cannot replicate the variability and spontaneity of authentic interaction with native speakers. Consequently, CAPT should be seen not as a replacement but as a complementary resource within language curricula. When combined with communicative practice, these tools offer a scalable and accessible way to enhance pronunciation training while preserving the essential benefits of human interaction.

Ethics declaration: Since the study is not a human or animal study, ethics committee approval is not required. Author personally collected the data during the classes without including any private information from participants.

Declaration of interest: The author declared no competing interest.

Data availability: Data generated or analyzed during this study are available from the author on request.

REFERENCES

Amrate, M., & Tsai, P. (2025). Computer-assisted pronunciation training: A systematic review. *ReCALL, 37*(1), 22–42. https://doi.org/10.1017/S0958344024000181

Bester, G., & Brand, L. (2013). The effect of technology on learner attention and achievement in the classroom. *South African Journal of Education, 33*(2), 1–15. https://doi.org/10.15700/saje.v33n2a405

Cardozo, A. C., & Orta, G. M. D. (2012). Technology-mediated activities as a boost to motivation and autonomy in the pronunciation class. In L. Anglada, & D. L. Banegas (Eds.), *Views on motivation and autonomy in ELT selected papers from the XXXVII FAAPI Conference*.

Cengiz, B. C. (2023). Computer-assisted pronunciation teaching: An analysis of empirical research. *Participatory Educational Research, 10*(3), 72–88. https://doi.org/10.17275/per.23.45.10.3

Chacón, G. P. (2022). Students' perception of interacting with native speakers. *Letras, 72*, Article 7. https://doi.org/10.15359/rl.2-72.7

Chinedu, A. J., & Anthonia, A. (2015). Learning French as a second language: Challenges for a native English speaker. *Continental Journal of Arts and Humanities, 7*(1), 21–29.

Dao, P., Nguyen, M. X. N. C., & Nguyen, C. B. (2020). Effect of pronunciation instruction on L2 learners' listening comprehension. *Journal of Second Language Pronunciation, 7*(1), 10–37. https://doi.org/10.1075/jslp.19012.dao

Derwing, T. M., & Munro, M. J. (2015). *Pronunciation fundamentals: Evidence-based perspectives for L2 teaching and research*. John Benjamins. https://doi.org/10.1075/lllt.42

Ekşi, G. Y., & Yeşilçınar, S. (2016). An investigation of the effectiveness of online text-to-speech tools in improving EFL teacher trainees' pronunciation. *English Language Teaching, 9*(2), 205–215. https://doi.org/10.5539/elt.v9n2p205

Fontan, L., Kim, S., De Fino, V., & Detey, S. (2022). Predicting speech fluency in children using automatic acoustic features. In *Proceedings of the Asia-Pacific Signal and Information Processing Association Annual Summit and Conference*. https://doi.org/10.23919/APSIPAASC55919.2022.9979884

Fontan, L., Le Coz, M., & Detey, S. (2018). Automatically measuring L2 speech fluency without the need of ASR: A proof-of-concept study with Japanese learners of French. In *Proceedings of Interspeech 2018* (pp. 2544–2548). ISCA. https://doi.org/10.21437/Interspeech.2018-1336

Gelin, L., Pellegrini, T., Pinquier, J., & Daniel, M. (2021). Simulating reading mistakes for child speech transformer-based phone recognition. In *Proceedings of Interspeech 2021*. https://doi.org/10.21437/Interspeech.2021-2202

Henrichsen, L. E. (2021). An illustrated taxonomy of online CAPT resources. *RELC Journal, 52*(1), 179–188. https://doi.org/10.1177/0033688220954560

Huang, Y., Lee, A. H., & Ballinger, S. (2023). The characteristics and effects of peer feedback on second language pronunciation. *Journal of Second Language Pronunciation, 9*(1), 47–70. https://doi.org/10.1075/jslp.22034.hua

Lyster, R., Saito, K., & Sato, M. (2013). Oral corrective feedback in second language classrooms. *Language Teaching, 46*, 1–40. https://doi.org/10.1017/S0261444812000365

Mahdi, H. S., & Al Khateeb, A. A. (2019). Context and implications document for: The effectiveness of computer-assisted pronunciation training: A meta-analysis. *Review of Education, 7*(3), 502–505. https://doi.org/10.1002/rev3.3164

Martin, J. L., & Wright, K. E. (2023). Bias in automatic speech recognition: The case of African American language. *Applied Linguistics, 44*(4), 613–630. https://doi.org/10.1093/applin/amac066

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

O'Brien, M. G. (2020). Ease and difficulty in L2 pronunciation teaching: A mini-review. *Frontiers in Communication, 5*. https://doi.org/10.3389/fcomm.2020.626985

Olsen, R. B., Orr, L. L., Bell, S. H., & Stuart, E. A. (2013). External validity in policy evaluations that choose sites purposively. *Journal of Policy Analysis and Management, 32*(1), 107–121. https://doi.org/10.1002/pam.21660

Revell-Rogerson, P. M. (2021). Computer-assisted pronunciation training (CAPT): Current issues and future directions. *RELC Journal, 52*(1), 189–205. https://doi.org/10.1177/0033688220977406

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://10.1111/j.1467-9922.2011.00639

Saito, K., & Plonsky, L. (2019). Effects of second language pronunciation teaching revisited: A proposed measurement framework and meta-analysis. *Language Learning, 69*(3), 652–703. https://doi.org/10.1111/lang.12345

Sullivan, K., & Czigler, P. (2002). Maximising the educational affordances of a technology supported learning environment for introductory undergraduate phonetics. *British Journal of Educational Technology, 33*(3), 333–343. https://doi.org/10.1111/1467-8535.00268

Tsunemoto, A., Trofimovich, P., & Kennedy, S. (2020). Pre-service teachers' beliefs about second language pronunciation teaching, their experience, and speech assessments. *Language Teaching Research, 27*(1). https://doi.org/10.1177/1362168820937273

Verbeke, G., & Simon, E. (2023). *Listening to accents: Comprehensibility, accentedness and intelligibility of native and non-native English speech*. Lingua. https://doi.org/10.1016/j.lingua.2023.103572

◆❖◆