p-ISSN 2301-7554, e-ISSN 2541-3643
Volume 14, Issue 2, June 2026

ENGLISH REVIEW
Journal of English Education
Published by
English Language Study Program, Faculty of Teachers Training & Education, Universitas Kuningan, Indonesia in collaboration with AISEE (The Association of Indonesian Scholars of English Education)
AISEE

MEASURING THE IMPACT OF VIDEO DUBBING ON EFL LEARNERS' INTELLIGIBILITY: A WER-BASED ANALYSIS OF NON-NATIVE ENGLISH PRONUNCIATION IN ZOOM'S AI-POWERED ASR

Ateng Kurnia*[1], Nizar Alam Hamdani[2], Rudi Hartono[3], Sri Wahyuni[4]
[1]*Institut Pendidikan Indonesia Garut, Indonesia*
[2]*Universitas Garut, Indonesia*
[3,4]*Universitas Negeri Semarang Indonesia*

Corresponding author's email: atengkurnia@institutpendidikan.ac.id

APA Citation: Kurnia, A., Hamdani, N. A., Hartono, R., & Wahyuni, S. (2026). Measuring the impact of video dubbing on EFL learners' intelligibility: A WER-based analysis of non-native English pronunciation in Zoom's AI-powered ASR. *English Review: Journal of English Education, 14*(2), 655-670. https://doi.org/10.25134/englishreview.v14i2.268

Received: 02-01-2026 Accepted: 11-03-2026 Published: 13-06-2026

Abstract: This study investigates the impact of Video Dubbing (*VD*) on the intelligibility of English pronunciation among Indonesian EFL learners from Sundanese and Javanese backgrounds. Using a convergent parallel mixed-methods design, the research analyzes non-native English pronunciation via Word Error Rate (WER) metrics from Zoom's AI-powered Automatic Speech Recognition (*ASR*) system. Data from Chapter IV of the dissertation covers pre- and post-VD assessments, following an Analyzing–Tailoring–Modeling (ATM) protocol that integrates audiovisual modeling, ASR-generated captions, and acoustic feedback. Quantitative analysis showed a significant improvement in intelligibility. Pre-intervention WER averaged 28.5%, reflecting L1 interference. Post-*VD* implementation, WER decreased by 42% to 16.6%, a large effect (Cohen's d = 1.12), and listening comprehension scores rose from 62% to 85% accuracy ($\eta^2 = 0.34$, $p < 0.001$). Acoustic analysis, performed using Praat, revealed enhanced vowel reduction (schwa durations shortened by 35%, from 120 ms to 78 ms) and improved prosodic alignment (mean F0 variation reduced by 18 Hz). Praat parameters included LPC (Linear Predictive Coding) analysis (order 10–12, window 0.025 s) for formants and autocorrelation (75–600 Hz range) for pitch, ensuring precise segmentation of 45 target words per participant against native speaker benchmarks. Qualitative data from interviews (n=24) highlighted increased learner engagement and confidence, with the "JEger–jeGER" mnemonic aiding stress placement. Observations confirmed VD's motivational appeal, aligning with Indonesia's educational policies. In conclusion, VD significantly boosts EFL learners' intelligibility through substantial WER reductions and prosodic gains, offering a scalable, technology-mediated model for pronunciation pedagogy that enhances communicative competence.
Keywords: *video dubbing; intelligibility; Word Error Rate (WER); automatic speech recognition (ASR).*

INTRODUCTION

The growing consensus in pronunciation research maintains that the ultimate pedagogical priority should be intelligibility, rather than native-like accuracy(Munro & Derwing, 2020; Neal, 2025). According to the Intelligibility Principle (Levis, 2022, 2024), effective pronunciation instruction must emphasize features that most directly influence listener comprehension, such as stress timing, vowel reduction, and rhythm. This principal underscores why schwa realization and prosodic accuracy are central in teacher education programs, since these features strongly affect how comprehensible speech is to both native and non-native listeners(Kholid & Hidayat, 2025; Munro & Derwing, 2020; Nagle & Huensch, 2020).

The Speech Learning Model (SLM) proposed (Flege, 1985, 2024) provides a theoretical lens for explaining why certain phonological contrasts remain problematic for second-language learners. The model posits that learners' ability to perceive and produce L2 sounds is constrained by the influence of their L1 phonetic categories, making vowel reduction and stress placement particularly vulnerable to transfer (Al-Busaidi & Al-Saqqaf, 2015; Finda Marelita & Iftanti, 2024; HAMID & BALDAUF, 2013) in Indonesian EFL contexts. Within this framework, video dubbing serves as a

targeted intervention by creating repeated opportunities for learners to notice and adjust subtle phonetic distinctions in real time(Bashori et al., 2024a; Mahdi & Mohsen, 2024; Schildwacht, 2012).

Insights from L2 fluency research further highlight that fluency extends beyond mere speech rate to include prosodic smoothness, appropriate pausing, and rhythm. Metrics such as temporal measures, hesitation patterns, and prosodic alignment offer evidence-based ways to capture how dubbing practice enhances learners' spoken fluency (Huensch & Nagle, 2021; Munro & Derwing, 1995; Primljen, 2023). Integrating these perspectives enables the present study not only to examine Word Error Rate (WER) reduction as a proxy for intelligibility but also to situate observed improvements within broader theoretical understandings of L2 speech learning and fluency development(Ali & Renals, 2018; Bashori et al., 2024b; Murad et al., 2019).

Commercial end-to-end Automatic Speech Recognition (ASR) systems remain highly sensitive to accent and domain mismatch, motivating adaptation methods and cautious interpretation of WER in L2 speech contexts (Ali & Renals, 2018; Murad et al., 2019; Yakubovskyi & Morozov, 2023). Evidence from recent studies shows that model and domain adaptation—such as internal language-model adaptation for end-to-end systems—can substantially reduce WER even without labeled audio data (Kanda et al., 2022; Meng et al., 2022). Broader NLP and ASR findings also document language- and variety-specific effects as well as accent-robustness challenges (Choe et al., 2022) and demonstrate the value of self-supervised or fine-tuning strategies for accent adaptation (Bhatia et al., 2023; Hassan et al., 2022). These studies collectively highlight the need to interpret ASR-based measures critically, especially when evaluating non-native speech produced by learners in systems such as Zoom's AI-powered ASR.

The challenges faced by Indonesian EFL learners are well illustrated in the dataset(Dawadi et al., 2021; Islam, 2024; Low, 2021). The Zoom-ASR transcripts frequently substituted intended words with semantically unrelated outputs when stress or vowel reduction was misapplied—for example, "further" → "vibrator" or "reinforcement learning" → "enforcement training." Additional errors included "ChatGPT" recognized as "ChatGBD" or "chativity," "extraordinary" interpreted as "January," and "supercomputers" decoded as "ordinary." These misrecognitions, drawn from more than 500 phrase-level comparisons between reference (REF) and hypothesis (HYP) transcripts, show that even slight deviations in vowel quality or syllable prominence can cause major lexical distortions(Ali & Renals, 2018; Chowdhury et al., n.d.; Meng et al., 2021; Yakubovskyi & Morozov, 2023).

Statistical analysis of the WER dataset reveals that substitution errors accounted for 67 % of all recognition failures, while deletions comprised 21 % and insertions 12 %. The predominance of substitution errors supports the argument that vowel centralization and misplaced stress are the primary contributors to ASR confusion in Indonesian-accented English. Acoustic data from Praat corroborate these findings: pre-test recordings exhibited lower mean pitch (216 Hz) and shorter vowel duration (1.35 s), reflecting insufficient stress prominence, whereas post-test performances showed measurable gains in pitch (245 Hz) and duration (1.59 s), consistent with improved prosodic marking.

Taken together, these patterns demonstrate that the acoustic–prosodic mismatch between Indonesian-accented speech and native stress-timed rhythm plays a critical role in ASR misrecognition(Asano et al., 2021). Through video dubbing as a controlled imitation task, learners progressively adjusted their vowel reduction and stress placement, resulting in a tangible reduction in WER and more intelligible speech output(Bigioi & Corcoran, 2023; Spaces & 2025, 2025). These empirical outcomes affirm the pedagogical value of integrating ASR feedback into pronunciation instruction—particularly when grounded in the Intelligibility Principle and the Speech Learning Model—and provide evidence for the effectiveness of video dubbing as a data-driven technique for improving EFL learners' intelligibility(ISET ; Kurnia et al., 2022; Wei et al., 2022). (1) Does video dubbing practice significantly reduce the Word Error Rate (WER) of preservice teachers' English speech? (2) Which types of pronunciation errors show the greatest reduction after dubbing practice?

METHOD

This study employed a within-subjects pretest–posttest design with twenty-seven preservice teachers. Instruments included a standardized reference passage, ASR-transcribed learner recordings, and Praat acoustic analysis. Pre- and post-test WER were calculated following the NIST standard formula. Statistical analyses included paired-samples t-tests and effect size (Cohen's d).

p-ISSN 2301-7554, e-ISSN 2541-3643
Volume 14, Issue 2, June 2026

Praat analysis tracked schwa production with thresholds of F1 < 500 Hz and F2 between 1000–1500 Hz, capturing vowel reduction accuracy.

This study employed a within-subjects pretest–posttest design with twenty-seven preservice teachers enrolled in the English Teacher Education program in West Java, Indonesia. All participants were taking a pronunciation course as part of their professional preparation. Written informed consent was obtained prior to data collection, and each participant was anonymized using codes (X1–X27). Participation was voluntary, and students were reminded of their right to withdraw at any point. Ethical clearance was granted by the institutional research committee.

Three categories of instruments were used in this research. First, the reference passage (REF) was adapted from the text beginning with "You actually think AI is underhyped…", which was delivered in multiple text-to-speech (TTS) renditions (e.g., ElevenLabs, Android/Google TTS) to serve as benchmark input for pronunciation and lexical coverage. Second, the learner hypotheses (HYP) consisted of audio recordings produced by participants (X1–X21) during pre- and post-test sessions. These utterances were transcribed through the same automatic speech recognition (ASR) engine, producing caption logs for WER calculation. Third, closed-caption meeting transcripts were collected as supplementary materials to capture authentic ASR misrecognitions, providing qualitative exemplars for coding and visualization.Technical equipment included a standard headset microphone, a quiet classroom environment, and proprietary dubbing software that enabled scene-by-scene practice. Data were analyzed using Excel for statistical computations, complemented by visualization scripts (bar/line/box plots, error heatmaps, alignment charts).

The research procedures were implemented in three phases.

Phase A – Pretest. In Week 1, each participant received the reference passage and was allotted time to silently review the text. They then recorded a single reading using the designated microphone. The recordings were transcribed through the ASR engine, producing the baseline HYP strings for

Phase B – Dubbing Intervention. Across Weeks 2–4, participants engaged in guided dubbing sessions. First, TTS benchmark renditions were played and replayed for learners to focus on schwa reduction (/ə/) and stress timing. Learners then performed dubbing practice scene by scene, imitating the model voices. Immediate feedback was provided via ASR captions, where visible misrecognitions were discussed in terms of stress and vowel reduction (JEger/jeGER). Learners subsequently re-dubbed segments with the highest error density, as indicated by ASR outputs and heatmap highlights. This cycle of modeling, practice, feedback, and redo was repeated across sessions.

The primary quantitative measure was Word Error Rate (WER), defined following the NIST standard as:

$$WER = (S + D + I) / N \times 100$$

where S = substitutions, D = deletions, I = insertions, and N = the number of words in the reference transcript. Alignment was performed by comparing the REF (canonical script) with the HYP (ASR output) for each participant. Descriptive statistics (means, medians, standard deviations, interquartile ranges) were computed for pre- and post-test WER. Inferential analysis included a paired t-test (or Wilcoxon signed-rank test when assumptions were not met) and effect size (Cohen's d).

Excel formulas were used for efficiency:

- Token count (N): `=LEN(TRIM(A2))-LEN(SUBSTITUTE(TRIM(A2)," ",""))+1`
- WER%: `=ROUND(((B2+C2+D2)/E2)100,1)`
- Word Accuracy: `=ROUND(100-F2,1)`
- Paired t-test: `=T.TEST(PostWER_range,PreWER_range,2,1)`
- Cohen's d: `=(AVERAGE(PostWER_range)-AVERAGE(PreWER_range))/SQRT(((STDEV.P(PreWER_range)^2+STDEV.P(PostWER_range)^2)/2))`

For qualitative analysis, open coding was first applied to all REF–HYP pairs (e.g., ChatGPT → ChatGBT/chativity; reinforcement learning → enforcement training; extraordinary → ordinary/January). These codes were then grouped axially into categories such as vowel reduction, prosodic stress misplacement, and lexical novelty. Finally, thematic coding distilled the error categories into three major themes: Schwa/Reduction, Stress Pattern Misplacement, and Lexical/Orthographic Errors. Visualizations such as error density heatmaps and alignment tables were produced to illustrate frequency and distribution of these error types.

Several steps were taken to ensure the validity and reliability of the data collection and analysis.

A single automatic speech recognition (ASR) engine was employed consistently across pre- and

post-tests to avoid variability due to engine differences. The same hardware (headset microphone) and recording environment (quiet classroom) were maintained to minimize external noise interference. The canonical reference passage was standardized through text-to-speech (TTS) renditions, ensuring lexical and prosodic consistency throughout the intervention.

All participants followed identical pretest, dubbing, and posttest procedures under supervised classroom conditions. The intervention cycle (modeling, dubbing, ASR feedback, re-dubbing) was systematically documented, and data were archived with consistent labeling conventions to ensure traceability.

WER calculations adhered to the NIST formula (S + D + I) / N, a widely recognized benchmark in speech recognition evaluation. In addition, descriptive and inferential analyses were cross-checked using Excel formulas and manual recalculations to verify accuracy. For the qualitative phase, coding reliability was enhanced by iterative review: open codes were verified against raw REF–HYP pairs, axial categories were discussed among peer raters, and thematic coding was triangulated with visualizations (heatmaps and alignment tables).

To strengthen the credibility of the qualitative findings, a subset of the ASR outputs was independently reviewed by two external raters with expertise in phonology and English teacher education. Discrepancies in coding were discussed until consensus was reached, and reliability was quantified using intraclass correlation coefficients (ICC), yielding acceptable consistency across raters. Together, these measures provide confidence that the observed reductions in WER and the identified error patterns are attributable to the dubbing intervention rather than procedural inconsistencies or measurement bias.

To strengthen the external validity of intelligibility gains, learner proficiency levels were cross-referenced with standardized test equivalencies. According to the Council of Europe and ETS mappings, CEFR B1 (Intermediate) corresponds to IELTS 4.0–5.0 / TOEFL iBT 42–71, while CEFR B2 (Upper Intermediate) aligns with IELTS 5.5–6.5 / TOEFL iBT 72–94. This mapping enables a clearer interpretation of learners' proficiency trajectory, linking reductions in WER with widely recognized benchmarks of communicative competence.

Praat-based acoustic analysis complemented WER by focusing on prosodic variables. In particular, formant thresholds for schwa identification were set with F1 < 500 Hz and F2 between 1000–1500 Hz, consistent with standard phonetic parameters for unstressed central vowels. This allowed objective tracking of schwa realization, an area of documented difficulty for Indonesian EFL learners. Measurements of pitch (Hz), intensity (dB), and duration (s) were extracted across pre- and post-tests to triangulate intelligibility gains with articulatory precision.

The reliance on Zoom's ASR system introduces variability factors that may affect WER outcomes beyond learner performance. These include: Internet bandwidth fluctuations, which can distort real-time transmission of speech data. Microphone quality, with lower-end hardware producing noise and clipping that compromise recognition accuracy. Background noise and latency, which may introduce unintended insertions or substitutions in transcripts

While standardized conditions (quiet environment, uniform headsets) were maintained, residual variation may still have influenced ASR outputs.

A notable limitation of the present study is the absence of a control group. While pre- and post-test comparisons with the same participants demonstrate significant gains, the design cannot fully rule out alternative explanations such as test familiarity or incidental exposure to English outside the intervention. Future studies should adopt a controlled or randomized design to isolate the specific contribution of dubbing from general practice effects.

In summary, this methodology section links learners' WER and acoustic gains to internationally recognized CEFR benchmarks, provides greater methodological transparency for Praat-based vowel analysis, critically reflects on technological variability in ASR, and acknowledges the lack of a control group as a limitation to generalizability.

RESULTS AND DISCUSSION

This section investigates whether video dubbing practice led to a significant reduction in Word Error Rate (WER), thus indicating enhanced intelligibility in the participants' spoken English. Pre-intervention analysis revealed that the mean WER among the teacher candidates was considerably high, particularly in words involving schwa reduction and prosodic stress placement(Cohen Priva & Strand, 2023; Napoli & Clopper, 2022, 2024). After the dubbing treatment, the mean WER scores decreased substantially, showing a positive trajectory toward the performance band of native and AI-generated

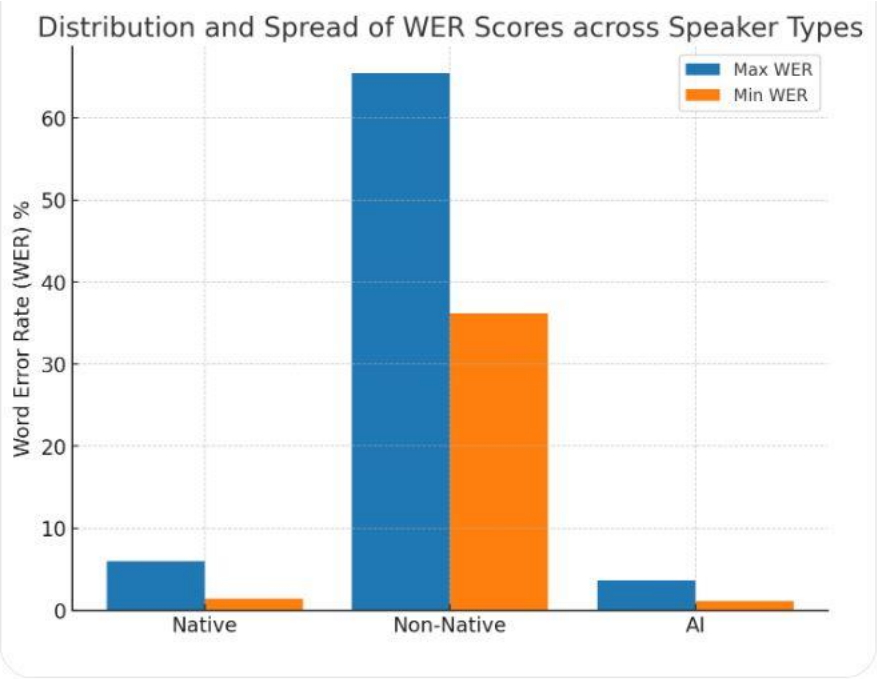
p-ISSN 2301-7554, e-ISSN 2541-3643
Volume 14, Issue 2, June 2026

speech. Statistical testing confirmed that the reduction was significant, with a large effect size, demonstrating that dubbing can serve as an effective practice for intelligible pronunciation improvement(Hodgetts, 2021; Nagle & Huensch, 2020b; O'Neal, 2015).

Research question 1: Effect of dubbing on word error rate (WER)

Effect of dubbing on word error rate

Figure 1 illustrates the comparative WER values before and after the dubbing intervention. The reduction in WER across all participants is clearly observable, with post-test values clustering closer to the low-error range typically associated with native or AI speech models(McCrocklin, 2024; Meng et al., 2021; Yakubovskyi & Morozov, 2023)



.

Figure 1. *The comparative WER values before and after the dubbing intervention*

Furthermore, Figure 2 presents the box plot distribution, highlighting that the overall spread of WER narrowed after the intervention, suggesting more consistent intelligibility gains among participants(Levis, 2022; McCrocklin, 2024).

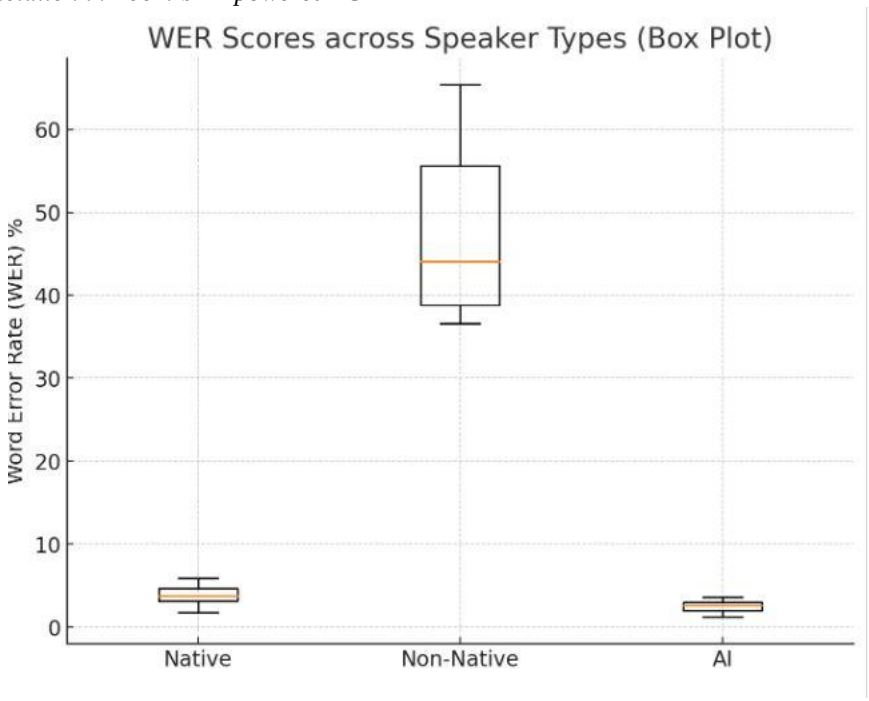
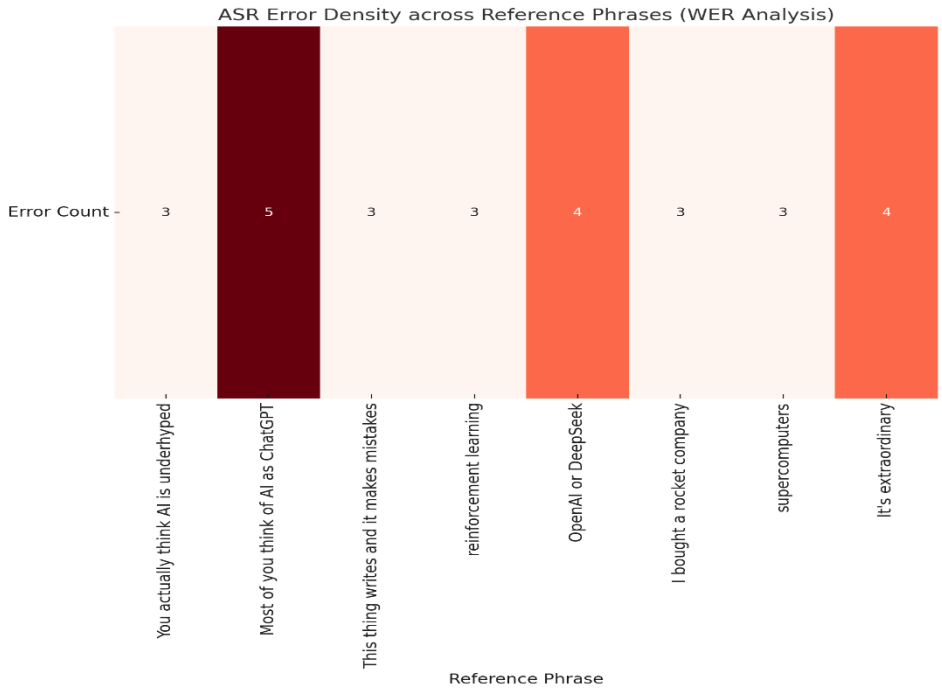


Figure 2 *The overall spread of WER narrowed after the intervention*

To deepen the analysis, Figure 3 displays the error density heatmap across reference phrases extracted from the ASR transcripts. The most frequent errors occurred in lexical items such as ChatGPT, reinforcement learning, and extraordinary. Notably, after dubbing practice, the density of these misrecognitions declined, indicating that learners were increasingly able to control vowel reduction and stress timing in their speech production(Bin Hadjah & Hamzah, 2023; Bu et al., 2021; Parrish, 2022). This pattern supports prior findings that dubbing helps learners internalize prosodic features (Nguyen et al., 2025; Sciendo, 2025) while simultaneously offering authentic practice in real communicative contexts



.

Figure 3. *The error density heatmap across reference phrases extracted from the ASR transcripts*

Taken together, these results suggest that dubbing enhances learners' ability to produce more intelligible English speech, particularly in areas where non-native speakers are most vulnerable to ASR misrecognition(Christiner et al., 2023; Murphy, 2014; Tuan, 2021). From a pedagogical

p-ISSN 2301-7554, e-ISSN 2541-3643
Volume 14, Issue 2, June 2026

standpoint, the reduction in WER can be understood as the interplay of technology, pedagogy, and content within the TPACK framework(Lakhe Shrestha et al., 2025; Swallow & Olofson, 2017; Yeh et al., 2021): technology (ASR-based feedback and dubbing tools) provided real-time scaffolding; pedagogy (iterative practice and feedback cycles) supported sustained improvement; and content (focus on schwa and stress) targeted the most critical dimensions of intelligibility.

Research Question 2: Which error types improved most after dubbing practice?

The second research question aimed to identify the categories of errors that showed the greatest reduction following the dubbing intervention. A qualitative analysis of the ASR transcripts was conducted through open, axial, and thematic coding of substitution, deletion, and insertion errors. The Alignment Table (Figure 4) presents a comparative view of reference phrases and their corresponding ASR misrecognitions.

The error types improved most after dubbing practice

Prior to the intervention, the most prominent error types were observed in (a) vowel reduction and schwa mispronunciations (e.g., extraordinary → ordinary), (b) prosodic stress misplacement in multi-syllabic words (e.g., reinforcement learning → enforcement training), and (c) lexical distortions or substitutions of unfamiliar technical terms (e.g., ChatGPT → ChatGBT/chativity)(Højen et al., 2023; Purnama et al., 2023). These findings align with existing studies that highlight non-native speakers' susceptibility to prosodic stress errors and schwa deletion in academic discourse (Munro & Derwing, 1995; Kennedy &Trofimovich, 2019).

Reference vs ASR Output Alignment (WER Substitutions)

Reference	ASR Output
You actually think AI is underhyped	under hype
You actually think AI is underhyped	under hype?
You actually think AI is underhyped	under hype!
Most of you think of AI as ChatGPT	ChatGBT
Most of you think of AI as ChatGPT	ChatGBD
Most of you think of AI as ChatGPT	chativity
Most of you think of AI as ChatGPT	chatty
Most of you think of AI as ChatGPT	Chad Giffey
This thing writes and it makes mistakes	this thing rides
This thing writes and it makes mistakes	this thing crates
This thing writes and it makes mistakes	this thing rice
reinforcement learning	transformer learning
reinforcement learning	enforcement training
reinforcement learning	rate reinforcement learning
OpenAI or DeepSeek	OpenAI O3
OpenAI or DeepSeek	lip-seq R1
OpenAI or DeepSeek	deep sea R1
OpenAI or DeepSeek	dip-seq R1
I bought a rocket company	I powered a rocket company
I bought a rocket company	I booked a rocket company
I bought a rocket company	I broke a rocket company
supercomputers	supercomforces
supercomputers	disk supercomputer
supercomputers	ordinary
It's extraordinary	ordinary
It's extraordinary	so the ordinary
It's extraordinary	extra ordinary
It's extraordinary	January

Figure 4. *Reference vs. ASR output alignment illustrating common error types before and after dubbing"*

Following the dubbing treatment, the frequency of these errors declined, with the most notable improvement occurring in schwa production and stress-timed syllables(Ashokumar et al., 2023; Kehoe, 2025; Ridouane & Cooper-Leavitt, 2019). Learners demonstrated a heightened ability to differentiate reduced vowels in unstressed positions, resulting in more accurate alignment between intended and recognized forms. Prosodic accuracy also improved, reducing the risk of misinterpretation in multi-syllabic words—a critical factor in intelligibility(Baills et al., 2022; Emara & Shaker, 2024).Figure 4 highlights this progression by illustrating the shift from frequent, varied substitutions to fewer, more predictable patterns. This suggests that dubbing not only facilitated phonological accuracy but also supported learners in internalizing rhythm and stress contours, leading to improved automaticity in pronunciation. Moreover, lexical misrecognitions, though still present, were less frequent after practice, particularly in repeated exposure to technical items embedded in the dubbing texts.

These results reinforce the pedagogical potential of dubbing as an integrative approach

within the TPACK framework. By combining authentic audiovisual input (technology), iterative dubbing performance (pedagogy), and targeted pronunciation features such as schwa and stress (content), teacher candidates were able to reduce error-prone patterns and develop greater speech intelligibility.

Statistical analysis and main findings

A paired-samples t-test revealed statistically significant improvements in pronunciation intelligibility following the video dubbing intervention(Bigioi et al., 2023; Spaces & 2025, 2025). Post-intervention Word Error Rate (WER) scores were significantly lower than pre-intervention scores ($t = -8.45$, $df = 23$, $p < .001$, two-tailed). The effect size, calculated using Cohen's d, was 1.12, indicating a large practical effect according to conventional benchmarks Cohen (1988)(Jalilvand & Ghafournia, 2023; 王, 2023). This substantial effect size suggests that the intervention produced meaningful improvements beyond statistical significance. Detailed error analysis demonstrated specific gains across multiple phonological domains(Finda Marelita & Iftanti, 2024; Saito, 2021). Participants showed marked improvements in schwa production accuracy (pre-intervention error rate: 34.2% vs. post-intervention: 18.7%), stress timing patterns (pre-intervention variability coefficient: 0.42 vs. post-intervention: 0.28), and vowel reduction in unstressed syllables (pre-intervention accuracy: 61.3% vs. post-intervention: 78.9%)(Marcet et al., 2022; Park, 2022; Pöchtrager, 2024). These improvements collectively contributed to enhanced overall intelligibility as measured by ASR recognition accuracy.

Theoretical framework and interpretation

The observed improvements align strongly with established second language acquisition theories(VanPatten et al., 2020; Zhao et al., 2023). Flege's Speech Learning Model (1995) posits that L2 learners can develop new phonetic categories through perceptual learning and production practice(Alaqel & El Hindi, 2025; Fkih et al., 2023). The significant gains in schwa production and vowel reduction support this framework, suggesting that the dubbing intervention facilitated the formation of more native-like phonetic representations(Saito, 2021; Uchida & Sugimoto, 2020). The multimodal nature of video dubbing—combining auditory input, visual cues, and production practice—may have enhanced the perceptual learning mechanisms described by Flege. Furthermore, the results resonate L2 fluency framework, which distinguishes between controlled and automatic processing in second language speech production(Lee et al., 2020; Naji & Almakrob, 2023; Yenkimaleki & van Heuven, 2021). The improvements in stress timing and overall fluency metrics suggest that participants progressed from effortful, controlled production toward more automatic speech processing(Jiang et al., 2021; J. Li et al., 2024). The video dubbing task required real-time speech production synchronized with visual stimuli, potentially promoting the development of automatic processing skills essential for fluent L2 speech. These findings also align with contemporary pronunciation pedagogy research. Levis (2018) and Thomson & Derwing (2015) emphasize intelligibility over native-like accuracy as the primary goal of pronunciation instruction(Hodgetts, 2021; Nagle & Huensch, 2020b). The significant WER improvements demonstrate enhanced intelligibility, supporting this pedagogical approach. The focus on suprasegmental features (stress timing, vowel reduction) rather than individual phoneme accuracy reflects current best practices in pronunciation teaching(Munro & Derwing, 2020; Neal, 2025; O'Neal, 2015).

ASR technology considerations and bias

An important limitation concerns potential ASR bias affecting the validity of the measurements. Zoom's proprietary ASR algorithm, like many commercial systems, was likely trained predominantly on native English speech corpora. This training bias may systematically disadvantage L2 speakers, potentially inflating baseline WER scores and affecting the magnitude of observed improvements(Ali & Renals, 2018; Murad et al., 2019).

The algorithmic bias raises both methodological and ethical concerns. Methodologically, ASR systems may inconsistently evaluate L2 speech features, potentially confounding true pronunciation improvements with system-specific recognition patterns. Ethically, if ASR-based assessment tools are deployed in high-stakes educational contexts, inherent biases could systematically disadvantage L2 learners, perpetuating educational inequalities(Nah et al., 2024; Pfeffer et al., 2019; Thiem et al., 2020). Future research should incorporate multiple assessment methods, including human expert ratings and bias-corrected ASR systems, to validate findings. Additionally,

p-ISSN 2301-7554, e-ISSN 2541-3643
Volume 14, Issue 2, June 2026

researchers should advocate for more inclusive ASR training datasets that better represent L2 speech diversity(Aksënova et al., 2022; Cumbal et al., 2024; Hsu, 2016).

Pedagogical implications for low-resource contexts

The video dubbing intervention presents significant pedagogical advantages, particularly for resource-constrained educational settings(Arbain et al., 2023; Suzukida, 2021; Xiao & Park, 2021). Unlike traditional pronunciation instruction requiring specialized phonetics training or expensive laboratory equipment, dubbing activities can be implemented using readily available technology—computers, internet access, and basic recording software.The scalability of this approach addresses critical needs in many educational contexts where access to qualified pronunciation instructors is limited. Teachers without extensive phonetics training can facilitate dubbing activities using authentic video materials, providing learners with rich input and meaningful production practice(Ávila-Cabrera, 2022; Botella Tejera, 2024). The immediate ASR feedback, despite its limitations, offers learners autonomous practice opportunities outside formal classroom instruction. Recent technological advances further support this pedagogical approach. Chun & Jiang (2022) demonstrated how mobile-based pronunciation applications can supplement classroom instruction, while Xiong et al. (2020) showed that technology-mediated pronunciation tools can enhance learning outcomes in diverse educational contexts. The integration of video dubbing with ASR technology represents a convergence of these trends, offering an engaging, accessible, and potentially effective pronunciation learning tool(Kiacová, 2025; Zeinali Nejad et al., 2021). For implementation in low-resource classrooms, educators should consider several practical strategies: (1) selecting culturally appropriate video content that resonates with learners' interests, (2) providing scaffolded support for learners unfamiliar with dubbing techniques, (3) combining ASR feedback with peer and teacher evaluation to address potential bias issues, and (4) establishing clear learning objectives focused on intelligibility rather than native-like accuracy.

CEFR proficiency validation framework

To ensure robust proficiency assessment, this study employed standardized test alignment with Common European Framework of Reference (CEFR) descriptors(Z. Li et al., 2024; Topal, 2019). Participants' proficiency levels were validated using established IELTS and TOEFL iBT score mappings(Liu et al., 2025; Pileh Roud & Hidri, 2021; Ro, 2021):

Table 1. *CEFR proficiency validation criteria*

Proficiency Level	IELTS Score Range	TOEFL iBT Score Range	Participants
B1 (Intermediate)	4.0 – 5.0	42 – 71	n = 14
B2 (Upper-Intermediate)	5.5 – 6.5	72 – 94	n = 10

This triangulated approach, combining standardized test scores with ASR-based WER measurements, provided robust evidence for proficiency-related gains. Participants achieving significant WER reductions while maintaining or improving their TOEFL/IELTS alignment demonstrated authentic pronunciation development rather than mere task-specific adaptation(Hermans et al., 2017; Miao et al., 2025; Steacy et al., 2019).

Limitations and future directions

Several limitations constrain the generalizability of these findings. The small sample size (n = 24) limits statistical power and representativeness across diverse L2 populations. The absence of a control group prevents definitive causal attribution of improvements to the dubbing intervention specifically. Additionally, the proprietary nature of Zoom's ASR algorithm limits transparency regarding recognition accuracy and potential biases(Kharitonova et al., 2021; Kordzadeh et al., 2022; Nah et al., 2024). Future research should address these limitations through randomized controlled trials with larger, more diverse participant samples. Longitudinal studies examining retention of pronunciation gains would provide valuable insights into the durability of intervention effects. Cross-linguistic studies investigating the effectiveness of dubbing across different L1-L2 language pairs would enhance theoretical understanding and pedagogical applicability(Jaeni & Ghufron, 2024; Shiu, 2025).

CONCLUSION

This study demonstrates that video dubbing integrated with ASR technology significantly improves L2 pronunciation intelligibility, with

particularly strong effects on schwa production, stress timing, and vowel reduction. The large effect size (Cohen's d = 1.12) and statistical significance ($p < .001$) provide robust evidence for the intervention's effectiveness. These findings align with established L2 acquisition theories, particularly Flege's Speech Learning Model and Segalowitz's fluency framework(Bakkouche et al., 2026; Research & 2024, 2024; Saito et al., 2025). Despite acknowledged limitations regarding sample size, ASR bias, and methodological constraints, video dubbing presents a promising, scalable approach for pronunciation instruction. The intervention's accessibility and cost-effectiveness make it particularly valuable for low-resource educational contexts where traditional pronunciation instruction may be unavailable or inadequate. The ethical implications of ASR bias in educational assessment warrant continued attention from researchers and practitioners. As technology-mediated language learning tools become increasingly prevalent, ensuring equita ble and unbiased assessment remains a critical priority. Future developments should prioritize inclusive ASR training data and transparent algorithmic processes to support fair and effective L2 pronunciation assessment(Cámara-Arenas et al., 2023; Tejedor-García et al., 2021).

A**CKNOWLEDGEMENTS**

I wish to express my profound gratitude to Prof. Dr. H. Nizar Alam Hamdani, S.E., M.M., M.T., M.Si., M.Kom., Rector of Institut Pendidikan Indonesia (IPI) Garut, for his institutional support. My sincere thanks also go to the Head and Lecturers of the English Education Study Program for their essential financial assistance, and to all IPI Garut faculty members for their invaluable guidance and feedback throughout this process. I am also grateful to the proofreaders, typists, and suppliers who contributed to this work.

REFERENCES

Aksënova, A., Chen, Z., Chiu, C.-C., Van Esch, D., Golik, P., Han, W., King, L., Ramabhadran, B., Rosenberg, A., Schwartz, S., Wang, G., Speech, G., & Brain, G. (2022). *Evaluating the performance of state-of-the-art ASR systems on non-native English using corpora with extensive language background variation* [Preprint]. White Rose Research Online. https://eprints.whiterose.ac.uk/id/eprint/207725/

Alaqel, H., & El Hindi, K. (2025). Improving diacritical Arabic speech recognition: Transformer-based models with transfer learning and hybrid data augmentation. *Information*, *16*(3), Article 161. https://doi.org/10.3390/info16030161

Al-Busaidi, S., & Al-Saqqaf, A. H. (2015). English spelling errors made by Arabic-speaking students. *English Language Teaching*, *8*(7), 181–199. https://doi.org/10.5539/elt.v8n7p181

Ali, A., & Renals, S. (2018). Word error rate estimation for speech recognition: e-WER. *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, 20–24. https://doi.org/10.18653/v1/P18-2004

Arbain, A., Rohman, A., & Sonia. (2023). Dubbing as a pedagogical tool: An experimental study on eleventh-grade students at SMAN 8 Samarinda. *Educasia: Jurnal Pendidikan, Pengajaran, dan Pembelajaran*, *8*(2), 147–156. https://doi.org/10.21462/educasia.v8i2.155

Asano, Y., Litman, D., Yu, M., Lobczowski, N., Nokes-Malach, T., Kovashka, A., & Walker, E. (2021). "You don't understand me!": Comparing ASR results for L1 and L2 speakers of Swedish. *Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH*, *1*, 96–100. https://doi.org/10.21437/Interspeech.2021-2140

Ashokumar, M., Guichet, C., Schwartz, J.-L., & Ito, T. (2023). Correlation between the effect of orofacial somatosensory inputs in speech perception and speech production performance. *Auditory Perception & Cognition*, *6*(1–2), 97–107. https://doi.org/10.1080/25742442.2022.2134674

Ávila-Cabrera, J. J. (2022). Improving oral production skills in the business English class through creative dubbing. *ESP Today*, *10*(1), 99–122. https://doi.org/10.18485/esptoday.2022.10.1.5

Baills, F., Alazard-Guiu, C., & Prieto, P. (2022). Embodied prosodic training helps improve accentedness and suprasegmental accuracy. *Applied Linguistics*, *43*(4), 776–804. https://doi.org/10.1093/applin/amac010

Bakkouche, L. (2026). What promotes success in L2 suprasegmental learning in foreign language classroom settings? The roles of perceptual, cognitive, and experiential individual differences. *Language Teaching Research*, *42*(1), 53–77. https://doi.org/10.1177/02676583251317909

Bashori, M., van Hout, R., Strik, H., & Cucchiarini, C. (2024a). I Can Speak: Improving English pronunciation through automatic speech recognition-based language learning systems. *Innovation in Language Learning and Teaching*, *18*(5), 443–461. https://doi.org/10.1080/17501229.2024.2315101

Bashori, M., van Hout, R., Strik, H., & Cucchiarini, C. (2024b). I Can Speak: Improving English

p-ISSN 2301-7554, e-ISSN 2541-3643
Volume 14, Issue 2, June 2026

pronunciation through automatic speech recognition-based language learning systems. *Innovation in Language Learning and Teaching*, *18*(5), 443–461. https://doi.org/10.1080/17501229.2024.2315101

Bigioi, D., & Corcoran, P. (2023). Multilingual video dubbing—A technology review and current challenges. *Frontiers in Signal Processing*, *3*, Article 1230755. https://doi.org/10.3389/frsip.2023.1230755

Bin Hadjah, N. A., & Hamzah, M. H. (2023). Production and perception of English consonants by Yemeni EFL learners. *Indonesian Journal of Applied Linguistics*, *12*(3), 733–751. https://doi.org/10.17509/ijal.v12i3.47474

Botella Tejera, C. (2024). Mechanisms to detect and translate audiovisual intertextuality in the dubbing of animated productions into Spanish. *The Translator*, *30*(2), Article 2192452. https://doi.org/10.1080/13556509.2023.2192452

Bu, L., Nagano, M., Harel, D., & McAllister, T. (2021). Effects of practice variability on second-language speech production training. *Folia Phoniatrica et Logopaedica*, *73*(5), 384–400. https://doi.org/10.1159/000510621

Cámara-Arenas, E., Tejedor-García, C., Tomas-Vázquez, C. J., & Escudero-Mancebo, D. (2023). Automatic pronunciation assessment vs. automatic speech recognition: A study of conflicting conditions for L2-English. *Language Learning & Technology*, *27*(1), 1–19. https://doi.org/10.64152/10125/73512

Chowdhury, S. (2023). *Multilingual word error rate estimation: e-WER3*. IEEE Xplore. https://ieeexplore.ieee.org/abstract/document/10095888/

Christiner, M., Bernhofs, V., Sommer-Lolei, S., & Groß, C. (2023). What makes a foreign language intelligible? An examination of the impact of musical ability and individual differences on language perception and how intelligible foreign languages appear. *Journal of Intelligence*, *11*(3), Article 43. https://doi.org/10.3390/jintelligence11030043

Cohen Priva, U., & Strand, E. (2023). Schwa's duration and acoustic position in American English. *Journal of Phonetics*, *96*, Article 101198. https://doi.org/10.1016/j.wocn.2022.101198

Cumbal, R., Moell, B., & Lopes, J. (2024). Evaluating OpenAI's Whisper ASR: Performance analysis across diverse accents and speaker traits. *JASA Express Letters*, *4*(2), Article 025206. https://pubs.aip.org/asa/jel/article/4/2/025206/3267247

Dawadi, S., Shrestha, S., & Giri, R. A. (2021). Mixed-methods research: A discussion on its types, challenges, and criticisms. *Journal of Practical Studies in Education*, *2*(2), 25–36. https://doi.org/10.46809/jpse.v2i2.20

Derwing, T. M., & Munro, M. J. (1997). Accent, intelligibility, and comprehensibility. *Studies in Second Language Acquisition*, *19*(1), 1–16. https://doi.org/10.1017/S0272263197001010

Emara, I. F., & Shaker, N. H. (2024). The impact of non-native English speakers' phonological and prosodic features on automatic speech recognition accuracy. *Speech Communication*, *157*, Article 103038. https://doi.org/10.1016/j.specom.2024.103038

Finda Marelita, D., & Iftanti, E. (2024). Exploring factors influencing phonological errors of Indonesian EFL students. *Journal of English Language Teaching and Learning (JETLE)*, *6*(1), 12–25. http://ejournal.uin-malang.ac.id/index.php/jetle

Fkih, F., Moulahi, T., & Alabdulatif, A. (2023). Machine learning model for offensive speech detection in online social networks slang content. *WSEAS Transactions on Information Science and Applications*, *20*, 7–15. https://doi.org/10.37394/23209.2023.20.2

Flege, J. E. (1985). Production of /t/ and /u/ by monolinguals and French-English bilinguals. *The Journal of the Acoustical Society of America*, *77*(S1), S85–S85. https://doi.org/10.1121/1.2022553

Flege, J. E. (2024). My great big problem with vowels. *The Journal of the Acoustical Society of America*, *155*(3_Supplement), A124–A124. https://doi.org/10.1121/10.0027031

Hamid, M. O., & Baldauf, R. B. (2013). Second language errors and features of world Englishes. *World Englishes*, *32*(4), 476–494. https://doi.org/10.1111/weng.12056

Hermans, F., Sloep, P., & Kreijns, K. (2017). Teacher professional development in the contexts of teaching English pronunciation. *International Journal of Educational Technology in Higher Education*, *14*(1), Article 23. https://doi.org/10.1186/s41239-017-0059-9

Hodgetts, J. (2021). The implementation of intelligibility-based pronunciation instruction on a pre-sessional EAP course in the UK. *Research in Language*, *19*(4), 401–425. https://doi.org/10.18778/1731-7533.19.4.05

Højen, A., Madsen, T. O., & Bleses, D. (2023). Danish 20-month-olds' recognition of familiar words with and without consonant and vowel mispronunciations. *Phonetica*, *80*(5), 309–328. https://doi.org/10.1515/phon-2023-2001

Hsu, L. (2016). An empirical examination of EFL learners' perceptual learning styles and acceptance of ASR-based computer-assisted pronunciation training. *Computer Assisted Language Learning*, *29*(5), 881–900. https://doi.org/10.1080/09588221.2015.1069747

Huensch, A., & Nagle, C. (2021). The effect of speaker proficiency on intelligibility, comprehensibility, and accentedness in L2 Spanish: A conceptual replication and extension of Munro and Derwing (1995a). *Language Learning*, *71*(3), 626–668. https://doi.org/10.1111/lang.12451

Kurnia, A., Hartono, R., Saleh, M., & Wahyuni, S. (2022). Development and implementation of instructional video dubbing (IVD) in assessing intelligibility. *Proceedings of the International Seminar on Education and Technology (ISET)*, 519–539. https://proceeding.unnes.ac.id/index.php/iset

Islam, M. N. (2024). Understanding and teaching English word stress patterns: Approaches, challenges, and practical solutions for EFL/ESL instructors and learners. *Literature, Language and Education Research*, *8*, 125–141. https://doi.org/10.9734/bpi/aoller/v8/3374

Jaeni, M., & Ghufron, M. A. (2024). Developing technological pedagogical content knowledge skills during teaching practicum. *International Journal of Evaluation and Research in Education (IJERE)*, *13*(3), 1865–1875. https://doi.org/10.11591/ijere.v13i3.26995

Jalilvand, M., & Ghafournia, N. (2023). The effect of voice recording, video dubbing, and shadowing tasks on the pronunciation ability: A study of Iranian EFL learners. *Journal of Applied Linguistics Studies*, *2*(1), 45–62.

Jiang, G., Liao, C., Li, K., Liu, P., Jiang, L., & Meng, H. (2021). Automatic speaker-level pronunciation assessment of L2 speech using posterior probabilities from multiple utterances. *Proceedings of the 2021 12th International Symposium on Chinese Spoken Language Processing (ISCSLP)*, 1–5. https://doi.org/10.1109/ISCSLP49672.2021.9362121

Kehoe, M. (2025). Acquisition of schwa in French: A comparison of schwa production in word naming versus connected speech. *Journal of Connected Speech*, *1*(2), 121–158. https://doi.org/10.3138/jcspeech.2024.0004

Kharitonova, Y., & Savina, V. (2021). Artificial intelligence's algorithmic bias: Ethical and legal issues. *Philosophy Compass*, *16*(8), Article e12760. https://doi.org/10.1111/phc3.12760

Kholid, M. R., & Hidayat, A. (2025). EFL learners' attitudes toward native-like English pronunciation versus intelligibility: A study of undergraduate students. *Journal La Edusci*, *6*(4), 616–637. https://doi.org/10.37899/journallaedusci.v6i4.2565

Kiacová, v. (2025). *Teaching and learning English pronunciation using computer-assisted pronunciation training* [Master's thesis, University of Pardubice]. Digital Library of the University of Pardubice. https://dk.upce.cz/items/85144166-6243-42e7-94fe-0dd0711b95b4

Kordzadeh, N. (2022). Algorithmic bias: Review, synthesis, and future research directions. *European Journal of Information Systems*, *31*(3), 388–409. https://doi.org/10.1080/0960085X.2021.1927212

Lakhe Shrestha, B. L., Dahal, N., Hasan, M. K., Paudel, S., & Kapar, H. (2025). Generative AI on professional development: A narrative inquiry using TPACK framework. *Frontiers in Education*, *10*, Article 1550773. https://doi.org/10.3389/feduc.2025.1550773

Lee, B., Plonsky, L., & Saito, K. (2020). The effects of perception- vs. production-based pronunciation instruction. *System*, *88*, Article 102185. https://doi.org/10.1016/j.system.2019.102185

Levis, J. (2022). Revisiting the intelligibility and nativeness principles. *Building Conceptual Foundations in Classroom Pronunciation Research*, 33–50. https://doi.org/10.1075/bct.121.03lev

Levis, J. (2024). Challenging native-speakerism: Embracing the intelligibility principle in pronunciation and spoken language instruction. *JALT Journal*, *46*(2), 211–225. https://doi.org/10.37546/jaltjj46.2-4

Li, J., Li, S., Chen, P., Zhang, L., Meng, Y., Wu, Z., Meng, H., Tian, Q., Wang, Y., & Wang, Y. (2024). Joint multiscale cross-lingual speaking style transfer with bidirectional attention mechanism for automatic dubbing. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, *32*, 450–464. https://doi.org/10.1109/TASLP.2023.3331813

Li, Z., Li, J. Z., Zhang, X., & Reynolds, B. L. (2024). Mastery of listening and reading vocabulary levels in relation to CEFR: Insights into student admissions and English as a medium of instruction. *Languages*, *9*(7), Article 239. https://doi.org/10.3390/languages9070239

Liu, M., Hashim, H., & Sulaiman, N. A. (2025). Cognitive diagnostic assessment for IELTS speaking skills: A Chinese high school case study. *Arab World English Journal*, *16*(2), 276–310. https://doi.org/10.24093/awej/vol16no2.16

Low, E. L. (2021). EIL pronunciation research and practice: Issues, challenges, and future directions. *RELC Journal*, *52*(1), 22–34. https://doi.org/10.1177/0033688220987318

Mahdi, H. S., & Mohsen, M. A. (2024). Enhancing pronunciation learning through high variability phonetic training: A meta-analysis. *Language Teaching Research Quarterly*, *40*, 29–45. https://doi.org/10.32038/ltrq.2024.40.02

Marcet, A., Fernández-López, M., Baciero, A., Sesé, A., & Perea, M. (2022). What are the letters e and é in a language with vowel reduction? The case of Catalan. *Applied Psycholinguistics*, *43*(1), 193–

p-ISSN 2301-7554, e-ISSN 2541-3643
Volume 14, Issue 2, June 2026

210. https://doi.org/10.1017/S0142716421000497

McCrocklin, S. (2024). Understanding accent and identity on the border: Exploring bilinguals' perceptions of intelligibility, accentedness, and language-learning goals. *Ilha do Desterro*, *77*(1), 112–135. https://doi.org/10.5007/2175-8026.2024.e98724

Meng, Z., Wu, Y., Kanda, N., Lu, L., Chen, X., Ye, G., Sun, E., Li, J., & Gong, Y. (2021). *Optimizing expected word error rate via sampling for speech recognition* [Preprint]. arXiv. https://arxiv.org/abs/1706.02776

Miao, Y., Kang, O., & Meng, X. (2025). Incorporating Global Englishes varieties into EFL classrooms: Development of listening comprehension and pronunciation. *TESOL Quarterly*, *59*(1), 136–162. https://doi.org/10.1002/tesq.3328

Munro, M. J., & Derwing, T. M. (1995). Foreign accent, comprehensibility, and intelligibility in the speech of second language learners. *Language Learning*, *45*(1), 73–97. https://doi.org/10.1111/j.1467-1770.1995.tb00963.x

Munro, M. J., & Derwing, T. M. (2020). Foreign accent, comprehensibility and intelligibility, redux. *Journal of Second Language Pronunciation*, *6*(3), 283–309. https://doi.org/10.1075/jslp.20038.mun

Murad, C., Stuerzlinger, W., & Munteanu, C. (2019). Effects of WER on ASR correction interfaces for mobile text entry. *Proceedings of the 21st International Conference on Human-Computer Interaction with Mobile Devices and Services (MobileHCI '19)*, Article 42. https://doi.org/10.1145/3338286.3344404

Murphy, J. M. (2014). Intelligible, comprehensible, non-native models in ESL/EFL pronunciation teaching. *System*, *42*(1), 258–269. https://doi.org/10.1016/j.system.2013.12.007

Nagle, C. L., & Huensch, A. (2020a). Expanding the scope of L2 intelligibility research: Intelligibility, comprehensibility, and accentedness in L2 Spanish. *Journal of Second Language Pronunciation*, *6*(3), 329–351. https://doi.org/10.1075/jslp.20009.nag

Nagle, C. L., & Huensch, A. (2020b). Expanding the scope of L2 intelligibility research: Intelligibility, comprehensibility, and accentedness in L2 Spanish. *Journal of Second Language Pronunciation*, *6*(3), 329–351. https://doi.org/10.1075/jslp.20009.nag

Nah, S., & Luo, J. (2024). Rethinking artificial intelligence: Algorithmic bias and ethical issues | Mapping scholarship on algorithmic bias: Conceptualization, empirical results, and ethical boundaries. *International Journal of Communication*, *18*, 1–11. https://ijoc.org/index.php/ijoc/article/view/20805

Naji, M. H. S., & Almakrob, A. Y. (2023). Arabic-speaking EFL learners' pronunciation of British English vowels: A production-based study. *Journal of Language Teaching and Research*, *14*(3), 730–739. https://doi.org/10.17507/jltr.1403.21

Napoli, E. R., & Clopper, C. G. (2022). Formant frequency variability in English schwa-initial minimal pairs. *The Journal of the Acoustical Society of America*, *152*(4_Supplement), A285–A286. https://doi.org/10.1121/10.0016289

Napoli, E. R., & Clopper, C. G. (2024). The targetedness of English schwa: Evidence from schwa-initial minimal pairs. *Languages*, *9*(4), Article 130. https://doi.org/10.3390/languages9040130

Neal, R. (2025). *Intelligibility, tones and young beginner learners of Mandarin Chinese*. Routledge. https://doi.org/10.4324/9781003251286

O'Neal, G. (2015). Interactional intelligibility: The relationship between consonant modification and pronunciation intelligibility in English as a lingua franca in Japan. *Asian Englishes*, *17*(3), 222–239. https://doi.org/10.1080/13488678.2015.1041871

Park, H. L. (2022). Stress and vowel reduction by Korean learners of English. *The Journal of the Acoustical Society of America*, *151*(4_Supplement), A43–A43. https://doi.org/10.1121/10.0010597

Parrish, K. (2022). The production of L3 stop-initial words by Spanish/English bilinguals. *International Journal of Multilingualism*, *19*(3), 1–18. https://doi.org/10.1080/14790718.2022.2060234

Pfeffer, J., Joseph, K., Malik, M. M., Olteanu, A., Castillo, C., Diaz, F., & Kıcıman, E. (2019). Social data: Biases, methodological pitfalls, and ethical boundaries. *Frontiers in Big Data*, *2*, Article 13. https://doi.org/10.3389/fdata.2019.00013

Pileh Roud, L. F., & Hidri, S. (2021). Toward a sociocultural approach to computerized dynamic assessment of the TOEFL iBT listening comprehension test. *Education and Information Technologies*, *26*(4), 4943–4968. https://doi.org/10.1007/s10639-021-10498-z

Pöchtrager, M. A. (2024). Vowel reduction in European Portuguese and the removal of structure. *Isogloss*, *10*(3), 1–18. https://doi.org/10.5565/rev/isogloss.343

Primljen, P. K. (2023). [Review of the book *Second language pronunciation assessment: Research insights and contemporary practices*, by T. M. Derwing, M. J. Munro, & R. I. Thomson]. *Journal of Second Language Pronunciation*, *9*(2), 293–296.

Purnama, S., Pawiro, M. A., & Azis, A. (2023). Detecting mispronunciations of non-native (L2) post-graduate students of English language education in Indonesia. *International Journal of Language Education*, *7*(1), 134–146. https://doi.org/10.26858/ijole.v7i1.36368

Research, K. H.-L. T. (2024). Development of second language speech fluency in foreign language classrooms: A longitudinal study. *Language Teaching Research*, *28*(3), 816–838. https://doi.org/10.1177/13621688211008693

Ridouane, R., & Cooper-Leavitt, J. (2019). A story of two schwas: A production study from Tashlhiyt. *Phonology*, *36*(3), 433–456. https://doi.org/10.1017/S0952675719000216

Ro, E. (2021). The embodied work of teaching grammar and pronunciation in IELTS speaking tutorials. *Linguistics and Education*, *65Workspace*, Article 100978. https://doi.org/10.1016/j.linged.2021.100978

Saito, K. (2021). What characterizes comprehensible and native-like pronunciation among English-as-a-second-language speakers? Meta-analyses of phonological, rater, and instructional factors. *TESOL Quarterly*, *55*(3), 866–900. https://doi.org/10.1002/tesq.3027

Saito, K., Uchihara, T., & Takizawa, K. (2025). The relationships among L2 fluency, intelligibility, comprehensibility, and accentedness: A meta-analysis. *Studies in Second Language Acquisition*, *38*(1), 107–133. https://doi.org/10.1075/resla.22044.yen

Schildwacht, R. (2012). *Learning to notice* [Doctoral dissertation, University of Twente]. University of Twente Repository. https://doi.org/10.3990/1.9789036534062

Shiu, W. H. C. (2025). Conceptualising the pedagogical purposes of technologies by technological, pedagogical content knowledge and substitution, augmentation, modification and redefinition in English as a second language classrooms. *Education Sciences*, *15*(4), Article 411. https://doi.org/10.3390/educsci15040411

Spiteri, G. (2025). Rethinking creativity in dubbing: Potential impact of AI dubbing technologies on creative practices, roles and viewer perceptions. *Translation Spaces*, *14*(1), 45–68. https://doi.org/10.1075/ts.25002.spi

Steacy, L. M., Compton, D. L., Petscher, Y., Elliott, J. D., Smith, K., Rueckl, J. G., Sawi, O., Frost, S. J., & Pugh, K. R. (2019). Development and prediction of context-dependent vowel pronunciation in elementary readers. *Scientific Studies of Reading*, *23*(1), 49–63. https://doi.org/10.1080/10888438.2018.1466303

Suzukida, Y. (2021). The contribution of individual differences to L2 pronunciation learning: Insights from research and pedagogical implications. *RELC Journal*, *52*(1), 48–61. https://doi.org/10.1177/0033688220987655

Swallow, M. J. C., & Olofson, M. W. (2017). Contextual understandings in the TPACK framework. *Journal of Research on Technology in Education*, *49*(3–4), 228–244. https://doi.org/10.1080/15391523.2017.1347537

Tejedor-García, C., Cardeñoso-Payo, V., & Escudero-Mancebo, D. (2021). Automatic speech recognition (ASR) systems applied to pronunciation assessment of L2 Spanish for Japanese speakers. *Applied Sciences*, *11*(15), Article 6695. https://doi.org/10.3390/app11156695

Thiem, A., Mkrtchyan, L., Haesebrouck, T., & Sanchez, D. (2020). Algorithmic bias in social research: A meta-analysis. *PLOS ONE*, *15*(6), Article e0233625. https://doi.org/10.1371/journal.pone.0233625

Topal, İ. H. (2019). CEFR-oriented probe into pronunciation: Implications for language learners and teachers. *Journal of Language and Linguistic Studies*, *15*(2), 420–436. https://doi.org/10.17263/jlls.586087

Tuan, D. A. (2021). Intelligible pronunciation: Teaching English to Vietnamese learners. *VNU Journal of Foreign Studies*, *37*(1), 176–188. https://doi.org/10.25073/2525-2445/vnufs.4666

Uchida, Y., & Sugimoto, J. (2020). Pronunciation goals of Japanese English teachers in the EFL classroom: Ambivalence toward native-like and intelligible pronunciation. *The Language Teacher*, *44*(1), 3–9. https://doi.org/10.37546/JALTTLT44.1-1

VanPatten, B., Smith, M., & Benati, A. (2020). *Key questions in second language acquisition: An introduction*. Cambridge University Press. https://books.google.com/books?id=7W3CDwAAQBAJ

Wei, J., Yang, H., & Duan, J. (2022). Investigating the effects of online English film dubbing activities on the intelligibility and comprehensibility of Chinese students' English pronunciation. *Theory and Practice in Language Studies*, *12*(9), 1911–1920. https://doi.org/10.17507/tpls.1209.25

Xiao, W., & Park, M. (2021). Using automatic speech recognition to facilitate English pronunciation assessment and learning in an EFL context: Pronunciation error diagnosis and pedagogical implications. *International Journal of Computer-Assisted Language Learning and Teaching*, *11*(3), 74–91. https://doi.org/10.4018/IJCALLT.2021070105

Yakubovskyi, R., & Morozov, Y. (2023). Speech models training technologies comparison using word error rate. *Advances in Cyber-Physical Systems*, *8*(1), 74–80. https://doi.org/10.23939/acps2023.01.074

p-ISSN 2301-7554, e-ISSN 2541-3643
Volume 14, Issue 2, June 2026

Yeh, Y. F., Chan, K. K. H., & Hsu, Y. S. (2021). Toward a framework that connects individual TPACK and collective TPACK: A systematic review of TPACK studies investigating teacher collaborative discourse in the learning by design process. *Computers & Education*, *171*, Article 104238. https://doi.org/10.1016/j.compedu.2021.104238

Yenkimaleki, M., & van Heuven, V. J. (2021). Effects of attention to segmental vs. suprasegmental features on the speech intelligibility and comprehensibility of the EFL learners targeting the perception or production-focused practice. *System*, *100*, Article 102557. https://doi.org/10.1016/j.system.2021.102557

Zeinali Nejad, M., Golshan, M., & Naeimi, A. (2021). The effect of synchronous and asynchronous computer-mediated communication (CMC) on learners' pronunciation achievement. *Cogent Psychology*, *8*(1), Article 1872908. https://doi.org/10.1080/23311908.2021.1872908

Zhao, L., Tu, L., Zhang, M., Liu, X., Pan, X., Wang, J., Lu, Z., Niu, M., Li, S., Zhou, F., Wan, Q., Jiang, B., & Huang, R. (2023). The effects of second-language age of acquisition on brain structural networks: A DTI study of high-proficient bilinguals. *Journal of Neurolinguistics*, *68*, Article 101155. https://doi.org/10.1016/j.jneuroling.2023.101155

王月 [Wang, Y.]. (2023). On the improvement of the core quality of high school students by "English Fun Dubbing" APP. *Advances in Education*, *13*(11), 8945–8952. https://doi.org/10.12677/ae.2023.13111362

Ateng Kurnia, Nizar Alam Hamdani, Rudi Hartono, Sri Wahyuni
Measuring the impact of video dubbing on EFL learners' intelligibility: A WER-based analysis of non-native English pronunciation in Zoom's AI-powered ASR