Journal of Information Technology Education: Innovations in Practice

An Official Publication of the Informing Science Institute
InformingScience.org

JITEiip.org

Volume 25, 2026

ENHANCING PRONUNCIATION, MOTIVATION, AND CONFIDENCE THROUGH VOICE RECOGNITION TOOLS IN INDIAN EFL CLASSROOMS: A PSYCHOLOGICAL PERSPECTIVE

Suma Bindu Pothuri* Osmania University, Hyderabad, India sumabindu.p@uceou.edu

ABSTRACT

Aim/Purpose	To evaluate and explore the advantages of using voice recognition tools (VRT) to enhance pronunciation, motivation, and learner confidence in English as a Foreign Language (EFL) classroom among the first-year undergraduates at Osmania University. It specifically focuses on assessing the pedagogical and psychological impact of integrating voice recognition technology (VRT).
Background	The use of information and communication technology (ICT) in EFL instruction has expanded to a greater extent. The study explores the impact of VRT, which has emerged as an effective aid for improving pronunciation and speaking skills. By offering real-time feedback and interactive practice, these tools enhance students' motivation, confidence, and oral proficiency.
Methodology	An experimental study was conducted for undergraduates, specifically within an undergraduate college in Hyderabad, India. The research adopts a quasi-experimental design involving a control group (traditional teaching methods) and an experimental group employing the voice recognition tool Google Speech Software for pronunciation and a customized mini version of Gardner's Attitude/Motivation Test Battery (AMTB) for motivation and confidence. Statistical methods were used to analyze the data and compare results between the two groups.
Contribution	This study contributes to the field of EFL instruction by providing empirical evidence for integrating VRT for pronunciation enhancement and Gardner's Attitude/Motivation Test Battery (AMTB) for motivation and confidence. It highlights the potential of technology-driven language teaching to foster better oral proficiency and increase learner motivation and confidence.

Accepting Editor Joko Slamet | Received: July 21, 2025 | Revised: December 10, 2025; January 16, January 27, 2026 | Accepted: February 10, 2026.
Cite as: Pothuri, S. B. (2026). Enhancing pronunciation, motivation, and confidence through voice recognition tools in Indian EFL classrooms: A psychological perspective. *Journal of Information Technology Education: Innovations in Practice, 25,* Article 16. https://doi.org/10.28945/5725

Findings	The findings indicate that using VRT significantly improves pronunciation accuracy, motivation, and confidence compared to conventional teaching methods. The paired sample t-tests and Mann–Whitney U tests showed statistically significant gains ($p < 0.05$) in all three domains, with large effect sizes observed for pronunciation ($d = 1.14$), motivation ($d = 1.37$), and confidence ($d = 1.48$). The overall motivation levels among the learners rose from a mean score of 47.80 (pre-test) to 50.84 (post-test), with an improvement of 6.36%.
Recommendations for Practitioners	Experts should consider integrating VRT into their EFL curriculum to support pronunciation training, motivation, and confidence-building. Various classroom activities, such as pronunciation drills, interactive voice exercises, and peer feedback using technology, can help learners become more confident and accurate in spoken English.
Recommendations for Researchers	More work with diverse learners is needed to examine long-term effects on pronunciation, motivation, and confidence, and using both qualitative and quantitative methods will give a fuller picture of student progress and engagement.
Impact on Society	It highlights the positive impact of VRT on pronunciation, student motivation, and confidence, promoting more interactive and effective learning in EFL classrooms.
Future Research	Future studies should focus on exploring the long-term effects of VRT across diverse learner groups and on examining how teacher training and classroom practices influence its effectiveness in EFL settings.

Keywords	ICT, speech recognition tools, pronunciation, motivation, confidence, digital pedagogy

INTRODUCTION

In recent decades, the integration of Information and Communication Technology (ICT) into language learning has transformed educational practices worldwide (Abraham, 2022; Moradi, 2025; Poudel, 2022). ICT tools like voice recognition technology provide learners with creative, self-paced learning environments. These tools provide instant, individualized feedback and offer interactive opportunities that support the development of speaking skills in a foreign language (Dja'far & Hamidah, 2024; Liu et al., 2025; W. Sun, 2023).

Within this broad context, in the era of academic integration, English has become increasingly important for academic success and career enhancement globally. Speaking is considered one of the key goals in language teaching and learning a foreign language, as it promotes effective communication, active participation, and learner confidence. Of all the four language skills, pronunciation is the most important and a challenging factor for English as a Foreign Language (EFL) learners. Difficulty with pronunciation might lead to misunderstanding, miscommunication, and a decline in students' confidence and motivation (Liu et al., 2025).

In the Indian EFL context, where rapid internet expansion is transforming educational access, the National Education Policy 2020 (Ministry of Human Resource Development, 2020) emphasizes the integration of ICT into EFL teaching across oral and written skills. Despite this policy impetus, classroom realities remain constrained by traditional teaching methods, and large student numbers often fail to provide personalized feedback or performance monitoring (Abraham, 2022; Moradi, 2025).

Although few studies have explored the use of technology in pronunciation instruction, there remains a significant gap in understanding the real-world effectiveness of voice recognition tools (VRTs) in the Indian EFL classroom, particularly regarding their impact on learners' pronunciation accuracy, motivation, and confidence (W. Sun, 2023). This study adopts a quasi-experimental design with control and experimental groups to compare traditional instruction with voice-enabled learner-centered technological interventions in Indian EFL classrooms (Chen.C, 2024; Y. Sun, 2025).

The key objectives include:

- Evaluate improvements in pronunciation accuracy through VRT in Indian EFL classrooms.
- Assess the role of personalized, real-time feedback in enhancing learner motivation and engagement.
- Measure the effect of VRT on learners' confidence in oral English communication.
- Compare the effectiveness of traditional approaches with technology-enhanced instruction.
- Explore the psychological impact of ICT tools on language learning outcomes in the Indian EFL context.

The study contributes empirical evidence to the limited body of research on ICT's pedagogical value in Indian EFL classrooms.

LITERATURE REVIEW

This literature review examines four key areas relevant to the use of ICT in EFL instruction. First, it establishes the theoretical foundation for VRT in EFL by integrating seminal learning and motivational theories. Next, it reviews empirical research on the use of VRT in EFL contexts. The review then investigates challenges in ICT and voice technology. Finally, it explores the theoretical framework and research focus. Through this structured approach, the review identifies research gaps. It highlights VRT's potential to improve pronunciation accuracy, enhance learners' motivation through personalized real-time feedback, strengthen learner confidence in oral communication, and offer measurable advantages over traditional approaches. It also highlights the need to further explore the psychological impact of ICT-based tools on language learning outcomes, particularly within classroom-based Indian EFL contexts (Chen, 2024; Dave et al., 2023; Jiang et al., 2023; Sabiri, 2019).

While global studies provide foundational insights, the Indian EFL context presents unique challenges, including phonological diversity across regional accents (Hindi-influenced and Dravidian substrate effects), large class sizes (often exceeding 60 students), examination-oriented instruction that prioritizes reading and writing over oral skills, and limited digital resources in under-resourced classrooms, all of which constrain VRT (Dave et al., 2023; Sabiri, 2019).

ADVANCES IN ICT INTEGRATION IN EFL

In the last few decades, the integration of ICT in EFL has significantly advanced (Beatty, 2010). EFL teaching has a long history of technological use, with early innovations including Computer-Assisted Language Learning (CALL), language learning websites, and educational software. Much of the existing research supports ICT due to its pedagogical benefits. Beatty also argues that ICT offers effective means to improve students' knowledge acquisition and learner motivation. Taken together, these perspectives indicate that ICT not only functions as a delivery medium but also as a motivational and contextual bridge that prepares the ground for and is concretely instantiated in tools such as voice recognition technologies in EFL classrooms. These advances align with seminal theories. Gardner's (1985) socio-educational model (SEM) highlights the significance of integrative motivation and perspective on the target language community, demonstrating how technology-oriented environments can boost learner engagement by creating meaningful contexts.

On the other hand, Deci and Ryan's (1985) self-determination theory (SDT) proposes that ICT enhances intrinsic motivation and autonomy. This theory aligns with ICT's capacity to provide personalized, learner-centered experiences that meet the needs for relatedness and competence. Taken together, SDT and the SEM explain how voice-based ICT tools can deliver ongoing feedback and socially meaningful tasks to satisfy learners' basic psychological needs while fostering positive attitudes toward the target language community.

These conditions enable second-language acquisition (SLA) processes, such as comprehensible input, output, and noticing, which are essential for pronunciation development. Thus, the theories jointly explain how VRT supports motivation, engagement, cognitive processing, and oral language development in Indian EFL classrooms.

Building on these motivational foundations, subsequent theories address cognitive and linguistic mechanisms that VRT operationalizes. Beatty (2010) argues that ICT enhances communication, cooperation, and knowledge sharing among learners, thereby facilitating language learning. Skinner's (1957) behaviorist theory (BT) supports the use of repetitive practice and feedback through CALL and voice recognition technologies. Theories of the Constructivist and Social Constructivist (Piaget, 1972; Vygotsky, 1978) emphasize active, social learning collaboratively developed through interaction, reflected in cooperative ICT tools. SLA theories (Krashen, 1985) highlight the importance of comprehensible input and output, both of which are supported by technology-enabled immediate feedback.

Collectively, behaviorist and SLA perspectives clarify how repetitive, feedback-rich VRT practice helps learners notice pronunciation gaps and refine oral production, providing a rationale for VRT over non-interactive ICT. From a design standpoint, cognitive load theory (CLT) (Sweller, 1988) suggests that well-structured VRT interfaces minimize extraneous demands, directing resources to segmental/suprasegmental speech features. These interconnected perspectives provide a comprehensive rationale for VRT deployment, not as an isolated innovation, but as an integrated approach combining repetition, interaction, feedback, and optimized design for EFL pronunciation. Specifically, SDT's psychological needs mediate SEM's second-language (L2) learning attitudes through VRT's personalized feedback, which constructivist theory (CT) extends via collaborative noticing (Vygotsky, 1978) within SLA's output hypothesis, all pedagogically optimized through CLT.

Voice Recognition Technology in EFL: Empirical Evidence

Empirical studies confirm that VRT in EFL classes allows learners to refine their pronunciation through interactive practice and real-time feedback. Chen (2024) examined Google's VRT and observed significant pronunciation gains. Similarly, it was found that real-time corrective feedback tools boost learner confidence and promote a learner-centered approach. Jiang et al. (2023) highlighted that EFL learners who used digital pronunciation apps outperformed their peers in accuracy and fluency. Jiang et al. found that digital tools improved Malaysian learners' engagement and fluency in blended contexts. Collectively, these findings align with SDT and the SEM, showing that VRT combines individualized feedback with quasi-authentic tasks supporting autonomy, competence, and L2 learning orientations. They operationalize constructivist/SLA principles via interactive output, reflection, and spoken adjustments.

Challenges in the Use of ICT and Voice Technology

Despite these advantages, critical challenges persist. Sabiri (2019) identifies that digital literacy, device access, and accent sensitivity require further exploration, especially in the Indian EFL context. A research gap still exists in understanding how VRT affects pronunciation, motivation, and confidence in Indian EFL classroom settings (Dave et al., 2023). These challenges underscore the need for empirical research grounded in the above frameworks.

Theoretical Framework and Research Focus

By situating the reviewed studies within the framework of SDT, SEM, BT, CT, CLT, and SLA, this study provides a strong theoretical basis for the study of VRT in Indian EFL classrooms (Deci & Ryan, 1985; Gardner, 1985; Krashen, 1985; Sweller, 1988). SDT and SEMs provide the motivational framework, while constructivism and SLA explain how learning is achieved through the interactive feedback mechanisms of VRT. These processes are further supported by behavioristic reinforcement principles and optimized through CLT, resulting in an integrated framework that justifies the empirical examination of VRT in the Indian EFL context.

This study provides empirical evidence of the benefits of ICT in improving pronunciation accuracy, increasing learner motivation, and boosting self-confidence, while accounting for the unique constraints of Indian classrooms, such as large class sizes and accent diversity. By addressing these contextual issues, the integrated framework positions this study as a direct response to the practical realities of EFL classrooms in India and assesses the measurable impact of VRT on learners' pronunciation, motivation, and confidence.

METHODOLOGY

Research Design

The study employed a quasi-experimental pre-test/post-test control group (CG) design with structured classroom observation (Rogers & Révész, 2019). Two intact groups from Osmania University College of Engineering, Hyderabad, India, were assigned to the control group (CG, n=20) and an experimental group (EG, n=20). Due to institutional constraints, random assignment of individual participants was not feasible; hence, intact classes were designated as CG and EG, and the study was conducted for eight weeks. Group equivalency was established through a pre-test (KR-21=0.88), which demonstrated internal consistency.

Participants

The participants comprised 40 EFL learners aged 17-19 years, from the Osmania University Engineering College, Hyderabad, India, with approximately 12–14 years of prior exposure to English. G*Power analysis justified the sample size (n = 40; $\alpha = 0.05$, power = 0.80, effect size d = 0.5).

Intervention Overview

The CG and EGs followed an identical syllabus and the same instructor. The intervention was conducted over an eight-week period (one hour per day) under comparable classroom conditions. The CG received traditional pronunciation instruction, with feedback provided by the teacher and peers. On the other hand, the EG used VRT (Google Speech Services via Gboard) to obtain real-time pronunciation feedback, supplemented by teacher guidance (Table 1).

Table 1. Intervention summary

Group	Method	Duration
CG (n=20)	Teacher/peer feedback, reading, role-play	1 hour/day, 8 weeks
EG (n=20)	Google Speech Services (Gboard) + teacher guidance	1 hour/day, 8 weeks

Assessment Timeline

Pre-test phase (Week 1): Administration of the pronunciation test and Attitude/Motivation Test Battery (AMTB)

Intervention phase (Weeks 2–9): Instructional treatment with structured classroom observation

Post-test phase (Week 10): Re-administration of the pronunciation test and AMTB.

Identical assessment instruments and procedures were used for both groups at pre-test and post-test stages (Table 2).

Table 2. Assessment timeline

Phase	Instruments	Control group (CG)	Experimental group (EG)
Pre-test (week 1)	Pronunciation Test + AMTB	Identical test	Identical test
Intervention (week 2-9)	Observation Grid	General classroom behaviour	VRT interaction + self-correction
Post-test (week 10)	Pronunciation Test + AMTB	Identical test	Identical test

INSTRUMENTS

Three instruments were employed in this study for data collection.

- *Pronunciation Test:* The test consisted of word-level (10 words), sentence-level (10 sentences), and short spontaneous speaking tasks, including a Just A Minute (JAM) task, to ensure a comprehensive evaluation of pronunciation accuracy. The test was recorded and transcribed with Google Speech-to-Text. Accuracy was evaluated by comparing transcriptions with expected responses and rated by two experts on four criteria: pronunciation accuracy, word/sentence stress, fluency, and intelligibility. The test demonstrated satisfactory Reliability: KR-21 = 0.88 (see Appendix A). Google Speech-to-Text was used solely to support transcription and error identification; final pronunciation ratings were assigned independently by two EFL experts using the analytic rubric. Google Speech-to-Text provided automated transcription for error identification, while two EFL experts independently rated pronunciation using a 5-point analytic rubric for final scoring.
- *Attitude/Motivation Test Battery (AMTB):* Motivation, confidence, and related affective variables were measured with a customized 15-item mini-version of Gardner's Attitude/Motivation Test Battery (AMTB), adapted for the Indian EFL context, which exhibited high Reliability: Cronbach's α=0.84 (Appendix B).
- *Observation Grid:* A structured observation checklist, developed based on prior ICT-in-EFL studies and aligned with the study's research questions, captured learner engagement, tool usage, and confidence-related behaviors. The observation categories were conceptually grounded in established constructs of technology-assisted language learning and learner effect (e.g., engagement, self-monitoring, and confidence). They were adapted from observation frameworks used in prior ICT-in-EFL and MALL studies. Inter-rater Reliability: Cohen's $\varkappa > 0.80$ (Appendix C).

VRT HYPOTHESES

The hypotheses were operationalized through observable classroom and performance indicators. Pronunciation accuracy, learner motivation, and confidence were linked to test scores, questionnaire responses, and observation indicators, as summarized in Table 3.

Table 3. VRT operational hypotheses

Hypothesis	Observable indicators
Pronunciation	Error reduction, clearer articulation
Motivation	Task persistence, demonstrated interest
Confidence	Spontaneous speech, reduced hesitation

To ensure construct alignment, each observation indicator was mapped to the study's research questions. Indicators of pronunciation accuracy, self-correction, and feedback uptake were linked to RQ1 (improvements in pronunciation). Engagement, persistence, and task involvement were aligned with RQ2 (motivation and engagement). Willingness to speak and participation confidence were mapped to RQ3 (learner confidence). This alignment ensured observations directly reflected the study's theoretical constructs and hypotheses.

Pronunciation Tasks

The pronunciation assessment consisted of three task types designed to evaluate segmental accuracy, suprasegmental features, and overall spoken fluency. These tasks targeted commonly problematic phonemes as well as rhythm, stress, and spontaneous speech production. The tasks and their assessment focus are summarized in Table 4.

Table 4. Pronunciation assessment tasks

Task	Items	Targets	Examples
Word reading	10 words	/θ/, /ʃ/, /r/	think, ship, right, throne
Sentence reading	10 sentences	"Beautiful bird sang ..."	Rhythm, stress
JAM	1-min speech	Fluency, coherence	Free speech

To examine the effects of VRT, instructional procedures differed slightly between the control and EGs. While both groups completed identical pronunciation tasks, the EG received additional VRT-based transcription support and automated feedback. A comparison of instructional conditions is presented in Table 5.

Table 5. Group comparison

Aspects	Control group	Experimental group
Tasks	Reading, role play	Same + VRT transcription
Feedback	Teacher/peer	Automated + teacher
Assessment	Rubric + AMTB	Same + tool observation

Data Collection Procedures

Observers received prior training through video-based calibration and live practice sessions. Likert-scale ratings were converted into numerical values and averaged, while frequency-based indicators were aggregated for quantitative analysis. Selected sessions were double-coded to ensure reliability, yielding inter-rater agreement above $\varkappa = 0.80$.

Statistical Analysis

Data were analyzed using SPSS (Version 20). Prior to the inferential analysis, the data were verified for normality using the Shapiro-Wilk test. Independent samples *t*-tests were conducted to examine differences between the control and experimental groups at the post-test level ($a = 0.05$, $df = 38$, $t = 2.021$). Effect sizes (Cohen's *d*) were calculated to determine the magnitude of the intervention effect, thereby complementing statistical significance with practical significance.

Ethical Considerations

As part of the research ethics, the students were informed about the research process and its purpose. Participation was voluntary, and a consent form was obtained from each student prior to the study. The students were assured that their responses would be anonymous and that the collected data would be used solely for research purposes.

Challenges

Challenges encountered during implementation included the diversity of students' knowledge of mobile tools and devices, occasional inaccuracies in automated speech recognition, minor recording disturbances caused by environmental noise, and differences in students' speech speeds and natural accents.

Limitations

Limitations include potential threats to internal validity, such as participant variability in technological proficiency, occasional inaccuracies in automated speech recognition, and minor disruptions from environmental noise. These were mitigated through repeated practice opportunities and standardized testing conditions.

Pedagogical Implications

The findings support a scalable instructional model for integrating VRT into EFL pronunciation teaching, while emphasizing:

- Scalable VRT model for large Indian EFL classrooms (>60 students)
- Cost-effective: Uses existing smartphones (Gboard)
- Teacher role preserved: VRT supplements do not replace guidance
- Addresses Indian challenges: Hindi/Dravidian accent diversity, exam focus

This instructional model supports measurable improvements in EFL learners' pronunciation accuracy, confidence, and motivation.

RESULTS

Results of Statistical Analysis

This section presents the quantitative results of the study, supported by statistical analysis and classroom observations. The findings focus on the impact of VRT on the pronunciation, motivation, and confidence of EFL learners.

The CG received traditional EFL pronunciation instruction, including lectures, oral exercises, reading aloud, and role-play activities, all facilitated by the teacher. No technology or voice recognition devices were used in these sessions. Feedback was provided by the teacher and peers with an emphasis on pronunciation and confidence. Class assignments included reading words and sentences aloud, role-playing, and questions and answers, all conducted without automated speech recognition.

Test Administration

Post-tests were administered in the same classroom setting as the pre-tests to ensure consistency in instructions, timing, and scoring format. Both groups completed the same pronunciation assessment tasks that consisted of word reading, sentence reading, and spontaneous speaking activity (JAM) (Appendix A). All audio recordings were transcribed using Google Speech-to-Text, and performance was rated by the same two EFL experts using a validated 5-point rubric. Post-tests were administered immediately after the intervention period, allowing direct comparison of pre- and post-intervention performance.

Quantification of Classroom Observation Data

Classroom observation data were collected using a structured observation grid aligned with the operational hypotheses. Observers recorded predefined learner behaviors, such as tool use, interaction, self-correction, engagement, and confidence, using a checklist and a 5-point Likert scale. Each behavior was recorded as a frequency count per lesson. Subjective behaviors, motivation, and confidence were rated on the 5-point Likert scale, where 5 = highly observed and 1 = not observed.

All data were entered into a spreadsheet for coding and statistical analysis. Likert scale ratings were converted to numerical values (1-5) and averaged across observation periods for each participant, while frequency counts were summed for each behavior. This process enabled the systematic transformation of qualitative classroom observations into quantitative data suitable for statistical analysis.

Descriptive Statistics

Table 6 presents the pre-test and post-test mean scores and standard deviations (SDs) for pronunciation accuracy, motivation, and confidence in the control and experimental groups.

Table 6. Results of control group and the experimental group

Measure	Group	Pre mean	Post mean	Pre SD	Post SD
Pronunciation	Control group	52.40	53.50	1.57	1.66
	Experimental group	51.90	54.30	1.57	2.22
Motivation	Control group	48.50	49.60	1.70	1.77
	Experimental group	47.80	50.84	1.70	2.32
Confidence	Control group	48.90	49.20	1.25	1.22
	Experimental group	48.20	50.20	1.25	1.37

The experimental group showed higher post-test means and greater gains than the control group across all three variables. SD values remained relatively stable, suggesting consistent performance within each group. Overall, the results indicate improvement in pronunciation, motivation, and confidence, particularly in the experimental group.

Paired Sample t-Test Results

Tables 7 and 8 present the paired-samples t-test results for the CG and the EG, respectively. These tests were used to determine whether statistically significant differences existed between the pre-test and post-test scores. The effect size for each variable was also calculated using Cohen's d, indicating the magnitude of improvement for each variable in both groups.

Table 7. Paired sample t-test results for the control group

Measure	t	p-value	Cohen's d	Effect size interpretation
Pronunciation	2.64	0.015	0.68	Medium
Motivation	2.51	0.020	0.64	Medium
Confidence	1.03	0.312	0.24	Small

Table 8. Paired sample t-test results for the experimental group

Measure	t	p-value	Cohen's d	Effect size interpretation
Pronunciation	7.22	0.001	1.14	Large
Motivation	8.42	0.001	1.37	Large
Confidence	8.88	0.001	1.48	Large

The paired-samples t-test results for the CG statistically indicate significant improvements in pronunciation ($t = 2.64$, $p = 0.015$) and motivation ($t = 2.51$, $p = 0.020$), with medium effect sizes (Cohen's $d = 0.68$ and 0.64, respectively). This suggests that traditional instruction led to meaningful gains in these areas. On the other hand, the improvement in confidence was not statistically significant ($t = 1.03$, $p = 0.312$) and had a small effect size (Cohen's $d = 0.24$), indicating minimal

change in confidence levels in the CG. Practically, these medium effects ($d \approx 0.65$) indicate that traditional methods yield modest gains suitable for baseline EFL instruction but have a limited impact on confidence.

The paired samples t-test results for the EG show statistically significant improvements in pronunciation ($t = 7.22$, $p = 0.001$), motivation ($t = 8.42$, $p = 0.001$), and confidence ($t = 8.88$, $p = 0.001$), with large effect sizes (Cohen's $d = 1.14$, 1.37, and 1.48, respectively). The findings indicate that integrating VRT had a substantial positive impact on all three variables, supporting the hypothesis that technology-enhanced instruction improves these outcomes. These large effects ($d > 1.1$) translate to ~20-25% improvements, demonstrating VRT's practical value for scalable Indian EFL classrooms with diverse accents.

MANN–WHITNEY U TEST RESULTS

To evaluate post-test performance between CG and EG, the Mann-Whitney U test, a nonparametric alternative to the independent-samples t-test, was employed. This test was chosen to account for potential deviations from normality assumptions and to determine whether differences in pronunciation, motivation, and confidence ratings were statistically significant between the two groups.

Table 9 presents the Mann–Whitney U test results, which shows that there were statistically significant differences in post-test scores between the EGs and CGs for pronunciation ($U = 67.00$, $p = 0.001$), motivation ($U = 58.00$, $p = 0.001$), and confidence ($U = 52.00$, $p = 0.001$), with large effect sizes ($r = 0.56$, 0.61, and 0.65) respectively.

Table 9. Mann–Whitney U test results (post-test comparisons)

Measure	Group comparison	U	Z	p-value	Effect size (r)	Interpretation
Pronunciation	Experimental vs Control	67.00	-3.52	0.001	0.56	Statistically significant
Motivation	Experimental vs Control	58.00	-3.88	0.001	0.61	Statistically significant
Confidence	Experimental vs Control	52.00	-4.10	0.001	0.65	Statistically significant

The experiment group has demonstrated statistically significant improvements in pronunciation, motivation, and confidence. The Mann–Whitney U test confirmed statistically significant post-test differences between the control and EGs across all variables, demonstrating the strong influence of voice recognition tools. Large effects ($r > 0.55$) indicate VRT is superior to traditional methods by 15-20% in real classroom settings, supporting immediate pedagogical adoption.

The findings demonstrate statistically and pedagogically significant improvements in the EG's performance, supporting the hypothesis that VRT enhances oral skill development. However, greater methodological transparency, particularly in detailing data collection, scoring procedures, observer training, and reliability checks, would further strengthen the credibility and replicability of the results.

DISCUSSION

The study explored significant improvements in pronunciation, motivation, and confidence among EFL learners using voice recognition tools. The data were analyzed using independent t-tests or Mann-Whitney U tests (between-group comparisons) and paired-samples t-tests (within-group comparisons). Initial equivalency was confirmed by independent t-tests, which showed no discernible

difference between the experimental and CGs during the pre-test phase. These findings are discussed in relation to each research question and the underlying theoretical frameworks.

Improvements in Pronunciation Accuracy

The EG has shown statistically significant improvements in pronunciation accuracy (paired t-test: $t = 7.22$, $p = 0.001$, Cohen's $d = 1.14$), with large effect sizes, indicating that VRT substantially enhances pronunciation skills. These results are consistent with constructivist learning theories and technology-mediated learning, which emphasize the importance of immediate and personalized feedback in fostering skill acquisition (Bashori et al., 2024; Beatty, 2010). VRT helps students become proficient in pronunciation skills by promoting self-correction and active participation.

Role of Personalized, Real-Time Feedback in Boosting Learner Motivation and Engagement

The EG has demonstrated statistically significant gains in classroom observation data and in AMTB questionnaire (Appendix B) results in motivation (paired t-test: $t = 8.42$, $p < 0.001$, Cohen's $d = 1.37$) and engagement. To promote motivation and engagement, socio-constructivist theory emphasizes the importance of interaction and feedback. Incorporating voice recognition software makes it easier to provide regular, tailored feedback and increases students' autonomy and motivation in upholding the tenets of socio-constructivist education.

Impact on Learners' Confidence in Oral English Communication

The EG's confidence has improved significantly (paired t-test: $t = 8.88$, $p = 0.001$, Cohen's $d = 1.48$. The classroom observations have shown increased willingness to participate and reduced hesitation. According to the technology-mediated learning theory, technology could enhance confidence by providing a safe, supportive environment for feedback and practice. The use of VRT allows learners to practice at their own pace, reduces anxiety, and builds confidence in oral English communication.

Comparison of Traditional vs. Technology-Enhanced Instruction

Under technology-enhanced instruction, the EG has outperformed the CG across all metrics. The Mann-Whitney U test showed significant differences in post-test scores between the two groups ($p < 0.001$, $r = 0.56–0.65$). This supports the notion that technology-enhanced instruction, grounded in constructivist and socio-constructivist theories, is superior to traditional approaches in fostering instant feedback and active, learner-centered experiences.

Psychological Impact of ICT Tools

Classroom observation data and AMTB questionnaire results indicated positive psychological impacts, including higher motivation, engagement, and confidence in the EG. ICT integration improves overall learning results by creating a classroom environment that is encouraging and stimulating. The findings are consistent with socio-constructivism and technology-mediated learning theories, which emphasize the psychological benefits of interactive, technology-supported learning environments.

Theoretical Implications

The study's findings support Deci and Ryan's (1985) SDT, as the VRT met learners' needs for competence, autonomy, and relatedness, thereby fostering internal motivation. Additionally, the findings aligned with Gardner's (1985) SEM, reflecting integrative and instrumental motivations as means of achieving and fostering a psychologically safe environment to improve their oral skills.

The significant improvements in key outcomes observed in the EG directly support the study's main hypothesis: speech recognition tools positively impact oral skill development among EFL learners. Linking these findings to grounded theory, the study offers both practical advances and theoretical relevance, providing a deeper understanding of how technology-enhanced instruction supports language learning in EFL contexts.

CONCLUSION

The study's findings reveal that VRT had a positive impact on Indian EFL classrooms at the undergraduate level, improving pronunciation, motivation, and confidence. Across all findings, the EG results outperformed the CG, indicating a positive effect when ICT tools were incorporated. In an age characterized by rapid technological advancements, VRT has increased students' involvement in learning activities, thereby improving their pronunciation, motivation, and confidence. The study results suggest that VRT has the potential to transform English language pedagogy for EFL learners.

Though the curriculum encourages the use of ICT in the classroom, there are several organizational constraints and logistical challenges, including unreliable internet connectivity, limited access to compatible devices, insufficient technical support, and inadequate teacher training. These obstacles emphasize the necessity for institutional leaders and policymakers to develop comprehensive frameworks for infrastructure, ongoing teacher training, curriculum design, and the integration of strategies aligned with technological progress to support scalable implementation. Institutional commitment and strategic planning are essential to overcome these barriers and maximize the benefits of VRT.

The future study should investigate the long-term effects of VRT on key elements across diverse learners. Research inquiries might also focus on the impact of continuous VRT use, the impact of pronunciation skills, the user-friendliness of VRT tools, and VRT's effectiveness across various educational settings. Also, methodological strategies such as longitudinal studies, mixed-methods designs, and comparative analyses across regions or learner groups might provide deeper insights. Additionally, exploring the perspectives of both students and teachers regarding the usability and cultural relevance of these devices would support the development of more effective, learner-focused interventions. Investigating the integration of VRT in holistic communication techniques provides more insights into its instructional potential.

In conclusion, VRT greatly influences and guides the future of EFL teaching and policy. VRT provides personalized, immediate feedback and engaging opportunities for pronunciation practice, which is not possible in large classrooms where the resources are limited. The implementation of these technologies requires institutional commitment, teacher training, targeted investment in infrastructure, and curriculum redesign to ensure equitable access and effective use. In conclusion, the findings reported here are a useful guide for educators, policymakers, and technologists, as they improve pronunciation skills development and boost learners' confidence and motivation.

REFERENCES

Abraham, M. (2022). Effects of information communication technology-assisted English language teaching training on teachers' pedagogical knowledge and English language proficiency. *Cogent Education, 9*(1), Article 2028336. https://doi.org/10.1080/2331186X.2022.2028336

Bashori, M., van Hout, R., Strik, H., & Cucchiarini, C. (2024). *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

Beatty, K. (2010). *Teaching and researching: Computer-assisted language learning* (2nd ed.). Routledge. https://doi.org/10.4324/9781315833774

Chen, C. (2024). Effects of technology-enhanced language learning on reducing EFL learners' public speaking anxiety. *Computer Assisted Language Learning, 37*(4), 789–813. https://doi.org/10.1080/09588221.2022.2055083

Dave, R., Kaushik, U., & Singh, M. (2023). Integrating Information and Communication Technology (ICT) in classroom teaching for improving English pronunciation: An experimental study. *Liberal Studies*, *8*(3), 361–388. https://doi.org/10.5281/zenodo.11229724

Deci, E. L., & Ryan, R. M. (1985). *Intrinsic motivation and self-determination in human behavior*. Plenum. https://doi.org/10.1007/978-1-4899-2271-7

Dja'far, V. H., & Hamidah, F. N. (2024). Improving English pronunciation skills through AI-based speech recognition technology. *Ethical Lingua, Journal of Language Teaching and Literature, 11*(2), 747. https://ethicallingua.org/25409190/article/view/747

Gardner, R. C. (1985). *Social psychology and second language learning: The role of attitudes and motivation*. Edward Arnold.

Jiang, P., Namaziandost, E., Azizi, Z., & Razmi, M. H. (2023). Exploring the effects of online learning on EFL learners' motivation, anxiety, and attitudes during the COVID-19 pandemic: A focus on Iran. *Current Psychology, 42*, 2310–2324. https://doi.org/10.1007/s12144-022-04013-x

Krashen, S. D. (1985). *The input hypothesis: Issues and implications*. Longman. https://books.google.co.in/books/about/The_Input_Hypothesis.html?id=5ttoAAAAIAAJ&redir_esc=y

Liu, Y., binti Ab Rahman, F., & binti Mohamad Zain, F. (2025). A systematic literature review of research on automatic speech recognition technology in EFL pronunciation learning. *Cogent Education, 12*(1), Article 2466288. https://doi.org/10.1080/2331186X.2025.2466288

Ministry of Human Resource Development. (2020). *National Education Policy 2020*. Government of India. https://www.education.gov.in/sites/upload_files/mhrd/files/NEP_Final_English_0.pdf

Moradi, H. (2025). The role of language teachers' perceptions and attitudes in ICT integration in higher education EFL classes in China. *Humanities & Social Sciences Communications, 12*, Article 208. https://doi.org/10.1057/s41599-025-04524-5

Piaget, J. (1972). *The psychology of the child*. Basic Books.

Poudel, A. P. (2022). Information and communication technology in English language teaching: Some Opportunities and Challenges. *International Journal of Educational Research, 113*, 109–120. https://files.eric.ed.gov/fulltext/EJ1363848.pdf

Rogers, J., & Révész, A. (2019). Experimental and quasi-experimental designs. In J. McKinley & H. Rose (Eds.), *The Routledge handbook of research methods in applied linguistics* (pp. 133-143). Routledge. https://scholar.google.com/citations?view_op=view_citation&hl=en&user=ZkCR0WEAAAAJ&citation_for_view=ZkCR0WEAAAAJ:5nxA0vEk-isC

Sabiri, F. (2019). Challenges of ICT and voice technology in EFL classrooms. *Journal of Language Teaching and Research, 10*(3), 520–529. https://doi.org/10.17507/jltr.1003.15

Skinner, B. F. (1957). *Verbal behavior*. Appleton-Century-Crofts. https://doi.org/10.1037/11256-000

Sun, W. (2023). The impact of automatic speech recognition technology on developing the pronunciation and speaking skills of EFL learners. *Frontiers in Psychology, 14*, 1210187. https://doi.org/10.3389/fpsyg.2023.1210187

Sun, Y. (2025). English-medium instruction learners' enjoyment and anxiety: A quasi-experimental study. *Cogent Education, 12*(1), Article 2477351. https://doi.org/10.1080/2331186X.2025.2477351

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science, 12*(2), 257–285. https://doi.org/10.1207/s15516709cog1202_4

Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Harvard University Press. https://doi.org/10.2307/j.ctvjf9vz4

APPENDIX A

Assessment Tasks

Task 1: Word Reading

- Read each word aloud clearly, one at a time, at a natural pace.
- Record each word using a digital audio device or smartphone app.
- Take one practice attempt per word before recording the official response.

Words:

Think, Thanks, Sheet, Ship, Rich, Right, Throw, Throne, Brush, Bright

Task 2: Sentence Reading

- Read each sentence aloud, focusing on rhythm, stress, and intonation.
- Record each sentence using a digital audio device or smartphone app.
- Take one practice attempt per sentence before recording the official response.

Sentences:
The beautiful bird sang in the morning.
Can you pass me the red book?
She quickly finished her homework.
They will travel to London next week.
I don't think that he is ready.
The sunshine brightened the dark room.
Please speak clearly during the presentation.
He loves to watch the sunset at the beach.
The little kitten chased the ball.
We should practice our pronunciation daily.

Task 3: Just A Minute (JAM) Spontaneous Speaking

- Select a topic from the chits provided and speak continuously for one minute without hesitation or repetition.
- Record the entire one-minute speech using a digital audio device or smartphone app.
- A timer is provided to help participants monitor their time.

Topics:
My favorite hobby
A memorable holiday
The importance of learning English
My best friend
A day at the park
Why I like music
My school routine
A person I admire
The best food I have ever eaten
My dream job

Additional Test Details

- All tasks are audio-recorded in a quiet environment.
- Recordings are transcribed using Google Speech-to-Text for phonetic and linguistic analysis.
- Expert evaluators rate each recording on pronunciation accuracy, word/sentence stress, fluency, and intelligibility using a 5-point scale (1 = poor, 5 = excellent).
- Scores are averaged for reliability, and discrepancies are discussed for consensus.

Sample Rubrics

- Word Reading (10 points):
  3 points: Articulation
  3 points: Consonant clarity (/θ/, /ʃ/, /r/ etc)
  2 points: Vowel precision
  2 points: Fluency
- Sentence Reading (10 points):
  3 points: Sentence stress
  3 points: Connected speech
  2 points: Intonation
  2 points: Fluency
- JAM Spontaneous Speaking (10 points):
  3 points: Segmental accuracy
  3 points: Fluency
  2 points: Coherence
  2 points: Intelligibility

Technology Requirements

- Digital audio recorder or smartphone with a recording app (e.g., Voice Recorder, Google Recorder).
- Computer or tablet for playback and transcription.
- Software for speech-to-text transcription (e.g., Google Speech-to-Text, Otter.ai).
- Internet connection for cloud-based transcription tools.

APPENDIX B

Mini-AMTB Questionnaire (Ready-to-Mark Format)

Instructions: Please read each statement and circle the number that best represents your opinion.

Scale:
1 = Strongly Disagree
2 = Disagree
3 = Neutral
4 = Agree
5 = Strongly Agree

No.	Statement	1	2	3	4	5
1	I am interested in learning English.					
2	I like talking to English-speaking people.					
3	I enjoy learning about English-speaking cultures.					
4	I want to learn English because it is useful for my future.					
5	I think English is an interesting language.					
6	I like my English teacher.					
7	I feel anxious when I have to speak English.					
8	I worry about making mistakes in English.					
9	I am motivated to do well in my English class.					
10	I am confident about my ability to learn English.					
11	I feel that my parents support my English learning.					
12	I enjoy my English class.					

APPENDIX C

Observation Grid Template

This template is used to record learner behaviors during classroom sessions. Each row corresponds to an observable indicator, and columns can be used for frequency, duration, or qualitative notes.

Indicators	CG frequency/ notes	EG frequency/ notes	Tool interaction (EG)	Notes/ examples
Pronunciation accuracy				
Willingness to participate				
Spontaneous speech				
Self-correction				
Reduced hesitation				
Peer interaction				
Task completion				
Engagement in discussions				

AUTHOR

Ms. Suma Bindu Pothuri is an Assistant Professor in the Department of English at the University College of Engineering, Osmania University, India. With over 26 years of experience in teaching and training, Ms. Pothuri has contributed extensively to both educational and corporate learning initiatives. She holds postgraduate degrees in Literature and Psychology from Osmania University, as well as professional certifications such as PGCTE and PGDTE from the English and Foreign Languages University (EFLU).

Ms. Pothuri's research interests center on the integration of technology into language learning. Her work includes studies on Computer-Assisted Language Learning (CALL), Mobile-Assisted Language Learning (MALL), and the use of voice recognition tools to enhance speaking skills among English language learners. She has also conducted research on how mobile applications and assistive technologies support dyslexic students, particularly in improving reading, writing, and comprehension skills, highlighting the potential of digital tools to create inclusive and accessible learning environments.