P-ISSN: 2790-6795	Annals of Human and Social Sciences	May-June 2026,Vol. 7, No.3
O-ISSN:2790-6809	http://doi.org/10.35484/ahss.2026(7-III)16	[164-178]

Annals of Human and Social Sciences
www.ahss.org.pk

RESEARCH PAPER

Effectiveness of AI-Based Mobile-Assisted Language Learning (MALL) Using SpeakingPal to Enhance English Speaking Skills of ESL Learners: A Quasi-Experimental Study

[1]Dr. Pir Suhail Ahmed Sarhandi, [2]Moomal Seelro and [3]Dr. Majid Hussain Buledi

1. Registrar, the University of Mirpurkhas, Sindh, Pakistan
2. PhD Scholar, Department of English Language & Literature, GC University Hyderabad, Sindh, Pakistan
3. Assistant Professor, AlRayyan National Colleges, Medina, Saudi Arabia.

Corresponding Author ssarhandi@gmail.com

ABSTRACT

This study examined the effectiveness of the AI-based SpeakingPal mobile application in improving English speaking proficiency among first-year undergraduate ESL learners. It was delimited to 60 female A2-B1 learners at one public-sector girls' degree college in Hyderabad, Sindh, Pakistan, and focused on fluency, pronunciation accuracy, communicative competence, and coherence and cohesion during an eight-week intervention. AI-supported Mobile-Assisted Language Learning offers repeated speaking practice, immediate feedback, and opportunities for autonomous learning beyond the classroom. However, controlled evidence concerning AI-based speaking applications remains limited in under-resourced South Asian higher-education contexts. This study addressed that gap by evaluating SpeakingPal under authentic classroom conditions. The study employed a quantitatively dominant mixed-method, quasi-experimental pre-test/post-test control-group design. Using purposive sampling, 60 first-year undergraduate ESL learners were assigned through two intact classes to an experimental group (n = 30) and a control group (n = 30). The experimental group used SpeakingPal for four approximately 30-minute sessions per week for eight weeks, whereas the control group received conventional speaking instruction for an equivalent period. Data were collected through a validated 20-point speaking rubric, a 20-item learner-engagement questionnaire, application-use records, and semi-structured interviews with 12 experimental-group participants. Quantitative data were analysed using descriptive statistics, paired- and independent-samples t-tests, ANCOVA, and Cohen's d, while interview data were analysed thematically. Both groups improved, but the experimental group achieved substantially greater gains across all four speaking dimensions. Its mean post-test speaking score ($M = 17.43$, $SD = 1.52$) exceeded that of the control group ($M = 13.27$, $SD = 1.84$), $t(58) = 9.71$, $p < .001$, $d = 2.51$. The group effect remained significant after controlling for pre-test performance, $F(1, 57) = 94.26$, $p < .001$, partial eta squared = .62. Learners reported high engagement and perceived usefulness ($M = 4.32/5.00$), while interviews highlighted immediate feedback, reduced speaking anxiety, and greater autonomy, alongside connectivity and speech-recognition challenges. SpeakingPal should be integrated as a structured supplement to teacher-led ESL speaking instruction rather than as a replacement for classroom teaching. Institutions should support reliable internet access and learner access to compatible devices, while developers should improve speech-recognition accuracy for regional accents and low-bandwidth settings. Future studies should use larger and more diverse samples, longer interventions, longitudinal follow-up, and detailed usage analytics.

Keywords: Mobile-Assisted Language Learning (MALL), Speakingpal, English Speaking Skills, Quasi-Experimental Design, AI In Education, ESL, Speech Recognition

Introduction

Mobile-Assisted Language Learning (MALL) is one of the most impactful changes in applied linguistics and educational technology over the last 20 years. MALL builds on the previous model of Computer-Assisted Language Learning (CALL) and takes advantage of the ubiquity and portability of smartphones and tablets to present flexible, context-aware, and truly learner-centred environments for language practice (Kukulska-Hulme & Shield, 2008; Viberg & Grönlund, 2013). The mobile infrastructure for mobile-based education is arguably more available than ever in the world with more than 6.9 billion active mobile subscriptions in 2024 (ITU, 2024).

In the four language skills, oral production is the most challenging skill for second language (L2) learners, particularly in formal academic contexts. When speaking, several areas of knowledge, such as phonological, lexical, syntactic and pragmatic, have to be integrated simultaneously, and the process must take place in real time under processing pressure (Skehan, 1998). The classroom is not normally the answer to this problem: in large classes individual students may get less than 2 minutes of speaking practice each class (Nation & Newton, 2009). The challenge is compounded by learner anxiety, inconsistent corrective feedback and a lack of opportunities for autonomous practice.

AI has paved the way for innovative solutions to these everlasting gaps. The use of automatic speech recognition (ASR), natural language processing (NLP), and interactive dialogue systems in the applications can replicate real communicative contexts and deliver instant and non-evaluative correction feedback to learners, something that classroom teachers can hardly do on a large scale in the classroom (Liu, 2016). However, controlled quasi-experimental studies measuring actual gains in speaking performance are limited especially in developing countries like Pakistan with infrastructural issues that add another complexity to technology integration (Memon et al., 2022).

The present study is directly relevant to these gaps. Using a strict quasi-experimental pre-test/post-test control group design, we test the meaningfulness of the improvement in English speaking proficiency of first-year undergraduate ESL students of the University of Mirpurkhas, Sindh, Pakistan using SpeakingPal.

Many people use AI language learning applications now, but the obstacles still persist in Pakistan for even the most difficult aspect of learning English: speaking. Speaking skills is one of the least developed among the four language skills as evident in national assessment of undergraduate students in Pakistan (British Council, 2015; Shah et al., 2021). Structural elements are important: class sizes often are over 40 students; instruction time is short and providing anything but superficial individualized feedback in most institutions is impractical at best.

The implementation of commercial MALL like SpeakingPal is growing, but the evidence for what works in the controlled environment is limited. Most of the published research is based on self-reported perception data, do not incorporate control groups, or do not separate the effect of MALL from any other instructional factors (Rosell-Aguilar, 2018). Based on the researchers' knowledge, there was not any quasi-experimental research especially to assess SpeakingPal in the context of undergraduate ESL students in Pakistan. Teachers and decision-makers are therefore taking technology adoption risks on faith.

Research Hypotheses

H_0: There is no statistically significant difference in English speaking performance between students using SpeakingPal and those receiving conventional instruction ($\alpha = .05$).

H_1: Students using SpeakingPal will demonstrate significantly greater improvement in English speaking performance than students receiving conventional instruction ($\alpha = .05$).

Sub-hypotheses: Students in the experimental group will show significantly greater gains in (H_1a) fluency, (H_1b) pronunciation accuracy, (H_1c) communicative competence, (H_1d) coherence and cohesion, and (H_1e) overall engagement and motivation compared to the control group.

Literature Review

Theoretical Foundations

Constructivist Learning Theory

Constructivist learning theory (Piaget, 1954; Vygotsky, 1978) suggests that learners create knowledge, rather than simply imbibing it. In language learning, this translates to authentic, interactive practice and corrective feedback being not just helpful, but essential for acquisition. Constructivist principles are operationalized in mobile-assisted environments, allowing learners to interact with language input and output with diverse real-life contexts, out of the classroom (Jonassen, 1999). 2.1.2 Mobile-Assisted Language Learning (MALL).

Mobile-Assisted Language Learning (MALL)

Kukulska-Hulme and Shield (2008) describe MALL as a further development of CALL with mobility, continuity and situatedness as its qualities. Sharples et al. (2007) defined the ideal situation as "seamless learning" – learning occurring seamlessly between formal and informal learning situations. The findings from meta-analysis indicate that the effect of mobile learning on language achievement is significant with the values of d ranging from 0.50 to 0.82 (Sung et al., 2016; Liu et al., 2020). But for speaking specific outcomes, effects are less well reported than for vocabulary and grammar, which this study directly addresses.

Human–Computer Interaction and Usability

HCI theory (Norman, 1988) reminds us that pedagogical effectiveness is not only based on content quality but also on the interface usability, system responsiveness, and immediacy of feedback. A low usability application can lead to additional cognitive load that, even with good instructional content, hinders learning (Mayer, 2009). The quality of ASR feedback (accuracy, timeliness, and interpretability) is consequently a vital mediating variable between the amount of student use of the system and the amount of actual improvement.

AI-Driven Speaking Practice Systems

With the advent of AI in language learning, the capabilities of MALL tools in terms of speaking practice have significantly increased. Current applications use deep neural network-based ASR for phoneme-level pronunciation analysis, prosodic feedback, and fluency scoring, which was previously only afforded by expert human raters (Cucchiarini et al., 2009).

The evidence of AI speaking tools is positive. Neri et al. (2002) showed that computer-assisted pronunciation training can significantly increase the phonological accuracy of Dutch L2 speakers. Liakin et al. (2015) concluded that mobile ASR-based feedback had a positive impact on L2 learner awareness of pronunciation errors in French.

However, the available studies vary in design, duration, and assessment tools, so more controlled classroom-based evidence is still needed.

However, there are significant caveats to the positive results. Most studies consider AI feedback to be a one-size-fits-all variable, which means that it is unclear which aspects of AI feedback led to gains. Developing educational contexts are systematically underrepresented in the studies. Even matched control groups and uniform assessment instruments are still the exception.

SpeakingPal in Educational Settings

SpeakingPal is an interactive mobile application that uses AI to aid in English speaking practice via video dialogue scenarios. SpeakingPal is not a generic ASR system, but a conversational agent system, where learners converse with the video characters, and get an immediate phonological analysis and are evaluated based on pronunciation accuracy, fluency, and intelligibility. The self-paced, repeatable format is purposefully structured to overcome the time and anxiety issues of conventional speaking coursework.

Limited but promising empirical studies of SpeakingPal are conducted. Liakin et al. (2017) found significant gains in oral fluency in intermediate French learners after six weeks but without the presence of a control group. In Saudi Arabia, Al-Harbi (2019) performed a quasi-experimental study, which revealed that the post-test results for the SpeakingPal group was significantly higher than that for the other groups ($p < .05$); however, effect sizes were not reported and the intervention lasted for 4 weeks. Overall, the available literature indicates some success, but with limitations of small sample size, brief intervention duration and a lack of long-term follow-up.

The following four gaps result from the above review. First, although there is ample evidence of the role of MALL in vocabulary and grammar development, the use of AI-based MALL for the development of speaking skills has received limited controlled experimental research. Second, no published quasi-experimental study with matched groups and standardized assessments and reporting of effect size has focused on the use of SpeakingPal in South Asian ESL contexts. Third, the existing research seldom breaks down the overall speaking achievements into the specific sub-skills, which leads to poor application of the research. Fourthly, little empirical research has been done on the mediation between the use of MALL and speaking improvement. The present study is designed to tackle with all four gaps in a single shot.

Material and Methods

Research Paradigm

The study have a positivist research paradigm which was operationalized quantitatively. The study's goal of testing falsifiable hypotheses concerning the measurable impacts of an AI-based instructional intervention on speaking performance in a controlled setting is compatible with positivism. The epistemological position is that it is possible to measure educational outcomes reliably using standardized assessment instruments and analyse the results using inferential statistics, which can be provisionally generalized (Creswell & Creswell, 2018).

Research Design

The design used was a quasi-experimental pre-test/post-test control group design (Campbell & Stanley, 1963). It was used to compare the impact of the intervention with conventional instruction because random allocation to conditions was not possible at the

participating college because intact classroom groups were used. The structure of the design is given in Table 1.

Table 1
Research Design Structure

Group	Pre-test	Intervention (8 weeks)	Post-test	n
Experimental	O_1	SpeakingPal AI MALL	O_2	30
Control	O_1	Conventional Instruction	O_2	30

Note. O_1 = pre-test observation; O_2 = post-test observation. Both groups received identical total instructional time (3 hours/week).

Participants

The subjects of the study were 60 first-year undergraduate students (N=60) of the public-sector girls' degree college in Hyderabad, Sindh, Pakistan, (Academic Year 2025-2026). Only first-year undergraduate students were involved in this study and none of the second or higher year students. The sampling method used was purposive sampling and inclusion criteria were: (a) students enrolled in first-year English language course, (b) students with a CEFR level A2–B1 as verified by pre-test standardized assessment and (c) students who owned or had access to a compatible smartphone. Two intact first-year classroom groups of students were assigned to the experimental (n = 30) and control (n = 30) groups. The demographic characteristics are presented in Table 2.

Table 2
Participant Demographic Characteristics

Characteristic	Experimental (n = 30)	Control (n = 30)	Total (N = 60)
Gender: Female	30 (100 %)	30 (100 %)	60 (100 %)
CEFR Level: A2	12 (40.0%)	11 (36.7%)	23 (38.3%)
CEFR Level: B1	18 (60.0%)	19 (63.3%)	37 (61.7%)
Prior MALL use: Yes	9 (30.0%)	8 (26.7%)	17 (28.3%)
Smartphone ownership	30 (100%)	30 (100%)	60 (100%)

Note. MALL = Mobile-Assisted Language Learning; CEFR = Common European Framework of Reference for Languages.

Intervention Description

Experimental Group: SpeakingPal Intervention

SpeakingPal (iOS/Android) is a mobile app for practicing English, which uses AI technology to enable interactive video dialogues with virtual characters, based on a script. It features: a deep learning-based Automatic Speech Recognition (ASR) module for processing the phonological input in real time, a Pronunciation Evaluation Engine (PEE) that generates the accuracy metrics of word level and phoneme level, an Interactive Dialogue System (IDS) that simulates turn-taking conversational practice, and an adaptive difficulty system that adapts the difficulty of tasks according to the history of individual performances.

There was a predetermined eight-week intervention schedule. Participants were asked to engage in a minimum of four SpeakingPal sessions per week (which was about 30 minutes per session), resulting in a total of about 16 hours of AI-assisted speaking practice over the intervention period. Themes were developed weekly in relation to the syllabus and comprised of daily usage, classroom dialogue, description of narrative, and expression of opinion. Application usage logs were tracked by the course instructor, and brief weekly encouragement was given; however, the course instructor did not intervene in the AI feedback process to not interfere with the treatment.

Control Group: Conventional Instruction

Conventional instruction is an integrated curriculum that is centred on the teacher. Students in the control group were given the same amount of in-class contact time (3 hours/week) as the experimental group that used the application, but instruction was conventional teacher-led speaking instruction. The conventional teaching method included teacher-modelled pronunciation, pair and group speaking work, role play, and oral presentation. Instructor feedback was also given in a conventional way (whole-class or deferred individual feedback), and there was no use of mobile applications or AI-based tools.

Instruments

Speaking Assessment Rubric

A standardized speaking assessment rubric (adapted from the IELTS Speaking Band Descriptors and Heaton, 1988) was developed for the assessment of four dimensions: (1) Fluency (5 points), (2) Pronunciation Accuracy (5 points), (3) Communicative Competence (5 points), and (4) Coherence and Cohesion (5 points) with total of 20. A pilot test of the rubric was conducted on 15 students outside the main sample with high interrater reliability (Cohen's $\kappa = .87$) between two trained raters. Individual assessments were carried out on the participants using oral production task which lasted for three minutes, consisting of Picture Description (1 minute), Narrative Retelling (1 minute), and Opinion Expression (1 minute). The pre- and post-test tasks were parallel forms of equal difficulty (see Appendix A for the complete rubric).

Learner Engagement Questionnaire

After the intervention, a 20-item Likert scale questionnaire (1 = Strongly Disagree and 5 = Strongly Agree) was given to the experimental group to assess perceived engagement, motivation and usability. Items were grouped into four subscales: (a) Perceived Usefulness (5 items); (b) Motivation and Enjoyment (5 items); (c) Ease of Use/Usability (5 items); and (d) Preference over Conventional Methods (5 items). Content validity was done by taking expert opinion from three faculty members of the field of Applied Linguistics from the University of Mirpurkhas. Internal consistency of the instrument was high (Cronbach's $\alpha = .89$; see Appendix B for the full instrument).

Semi-Structured Interviews

The participants of the experimental group who entered the individual semi-structured interview (20–30 minutes) were a subsample (40%) purposively selected during the last week of the intervention. The interview revealed the learners' experience and perceived challenges and offered practical suggestions to complement the quantitative findings and provide texture. Thematic analysis (Braun & Clarke, 2006) was used, with qualitative results described as supplementary evidence to inform the context in which the quantitative findings occurred, in keeping with the quantitative focus of the study (see Appendix C for interview protocol).

Variables

Independent Variable: Use of the SpeakingPal AI-based MALL application (experimental group) vs. Conventional instruction (control group).

Dependent Variables: (1) Total speaking performance score (0–20); (2) Fluency sub-score (0–5); (3) Pronunciation accuracy sub-score (0–5); (4) Communicative

competence sub-score (0–5); (5) Coherence and cohesion sub-score (0–5); (6) Learner engagement score (1–5).

Controlled Variables: Total instructional time, syllabus content coverage, assessment rubric, instructor, and assessment conditions.

Extraneous Variables: Individual differences in prior mobile learning experience, language aptitude, and motivation were monitored through pre-test baseline comparisons and demographic survey data.

Procedures

All respondents consented to participate in the study and due approval for the collection of the Data was sought from the principal of the college. Procedures were as follows: (1) Week 1—Participant orientation and pre-test was administered; (2) Weeks 2-9—Eight-week intervention period; (3) Week 10—Post-test was administered, questionnaires were distributed and interviews were conducted. All groups were evaluated under the same conditions. Two trained raters, blind to group assignment, independently scored pre-test and post-test tasks.

Data Analysis

IBM SPSS Statistics v.29 was used to analyse all quantitative data. The mean, standard deviation, and range were computed for pretest and post-test scores for both groups and all sub dimensions. These are the following Inferential procedures that have been used:

Paired-samples t-tests: to assess within-group pre-test to post-test gains separately for the experimental and control groups.

Independent-samples t-tests: to compare pre-test baseline equivalence and post-test performance between groups.

Cohen's d: to quantify effect size for all significant comparisons, interpreted as small ($d \geq 0.20$), medium ($d \geq 0.50$), or large ($d \geq 0.80$) per Cohen (1988).

One-way ANCOVA: conducted with post-test total score as the dependent variable and pre-test score as the covariate, to control for residual baseline differences.

Normality of distributions was confirmed via Shapiro-Wilk tests (all $p > .05$). Levene's test confirmed equality of variances for all independent-samples comparisons (all $p > .05$). Alpha was set at .05 for all inferential tests.

Results and Discussion

Pre-test Baseline Equivalence

It was critical to establish whether the two groups had a similar baseline prior to analysing the effects of the intervention. Independent-samples t-tests showed that there were no significant differences in pre-test overall scores on the speaking test, $t(58) = 0.43$, $p = .671$, or in each of the four sub-scales (all $p > .05$). This similarity of groups at pre-test is a required precondition for interpreting the subsequent differences between groups as being due to the intervention. The full pre-test comparison statistics are shown in Table 3.

Table 3
Pre-Test Score Comparison Between Groups

Dimension	Exp. M (SD)	Ctrl. M (SD)	t(58)	p	Cohen's d
Fluency	2.43 (0.57)	2.37 (0.61)	0.40	.693	0.10
Pronunciation Accuracy	2.30 (0.60)	2.23 (0.57)	0.47	.638	0.12
Communicative Competence	2.27 (0.58)	2.33 (0.55)	−0.42	.677	0.11
Coherence & Cohesion	2.37 (0.56)	2.30 (0.60)	0.47	.638	0.12
Total Score (0–20)	9.37 (1.88)	9.23 (1.94)	0.29	.770	0.07

Note. M = Mean; SD = Standard Deviation. All p-values > .05, confirming pre-test equivalence between groups.

Within-Group Pre-Test to Post-Test Gains

Experimental Group

In the experimental group, statistically significant difference between pre-test and post-test was found on all the speaking dimensions ($p < .001$). The mean total score rose from M = 9.37 (SD = 1.88) at pre-test to M = 17.43 (SD = 1.52) at post-test, representing a mean gain of 8.07 points, $t(29) = 24.33$, $p < .001$, $d = 4.44$. This is a very large effect and is indicative of the type of practice that can be obtained from SpeakingPal, with feedback. The greatest increase was observed for Fluency (Δ = 1.87 points on the sub-score scale) followed closely by Communicative Competence (Δ = 1.83) and Coherence and Cohesion (Δ = 1.83).

Control Group

The group also showed statistically significant pre-test to post-test gains, $t(29) = 7.21$, $p < .001$, $d = 1.32$, and this was attributed to the effects of regular classroom instruction. This is an important discovery: it shows that conventional teaching also has an impact. However, the increase in magnitude was significantly less for this group than for the experimental group (total gain ? = 4.03 vs. ? = 8.07 points), and this difference is directly considered by the between-group analysis. Within-group statistics are provided in full in Table 4. analysis. Within group statistics are provided in full in Table 4.

Table 4
Within-Group Pre-Test to Post-Test Gains (Paired-Samples t-tests)

Dimension	Pre M (SD)	Post M (SD)	Δ	t(29)	p	d	Group
Fluency	2.43 (0.57)	4.30 (0.46)	1.87	25.12	<.001	4.58	Exp.
Pronunciation Acc.	2.30 (0.60)	3.83 (0.53)	1.53	19.73	<.001	3.60	Exp.
Communicative Comp.	2.27 (0.58)	4.10 (0.48)	1.83	23.47	<.001	4.29	Exp.
Coherence & Cohesion	2.37 (0.56)	4.20 (0.48)	1.83	23.19	<.001	4.24	Exp.
Total Score (0–20)	**9.37 (1.88)**	**17.43 (1.52)**	**8.07**	**24.33**	**<.001**	**4.44**	**Exp.**
Fluency	2.37 (0.61)	3.43 (0.57)	1.07	9.11	<.001	1.66	Ctrl.
Pronunciation Acc.	2.23 (0.57)	3.30 (0.54)	1.07	8.80	<.001	1.61	Ctrl.
Communicative Comp.	2.33 (0.55)	3.27 (0.52)	0.93	8.53	<.001	1.56	Ctrl.
Coherence & Cohesion	2.30 (0.60)	3.27 (0.58)	0.97	8.49	<.001	1.55	Ctrl.
Total Score (0–20)	**9.23 (1.94)**	**13.27 (1.84)**	**4.03**	**7.21**	**<.001**	**1.32**	**Ctrl.**

Note. Exp. = Experimental Group (n = 30); Ctrl. = Control Group (n = 30). Δ = post-test minus pre-test mean. Effect sizes: small $d \geq 0.20$; medium $d \geq 0.50$; large $d \geq 0.80$ (Cohen, 1988).

Between-Group Post-Test Comparison

Independent-samples t-tests were performed to compare post-test scores between groups and statistically significant differences were found across all dimensions (all $p < .001$) and the experimental group scored higher than the control group on each dimension. The largest between-group difference was observed in Fluency ($\Delta = 0.87$, $d = 1.74$), followed by Coherence and Cohesion ($\Delta = 0.93$, $d = 1.74$), Communicative Competence ($\Delta = 0.83$, $d = 1.66$), and Pronunciation Accuracy ($\Delta = 0.53$, $d = 1.00$). The total score difference ($\Delta = 4.17$, $d = 2.51$) of the overall post-test score is a very large effect of the intervention by any interpretive standard used (Table 5).

Table 5
Between-Group Post-Test Comparison (Independent-Samples t-tests)

Dimension	Exp. M (SD)	Ctrl. M (SD)	Δ	t(58)	p	Cohen's d
Fluency	4.30 (0.46)	3.43 (0.57)	0.87	6.63	<.001	1.74
Pronunciation Accuracy	3.83 (0.53)	3.30 (0.54)	0.53	3.90	<.001	1.00
Communicative Comp.	4.10 (0.48)	3.27 (0.52)	0.83	6.55	<.001	1.66
Coherence & Cohesion	4.20 (0.48)	3.27 (0.58)	0.93	6.99	<.001	1.74
Total Score (0–20)	**17.43 (1.52)**	**13.27 (1.84)**	**4.17**	**9.71**	**<.001**	**2.51**

Note. Δ = Experimental post-test M minus Control post-test M. All $p < .001$. Effect sizes $d \geq 0.80$ are considered large (Cohen, 1988).

ANCOVA Controlling for Pre-Test Scores

A one-way ANCOVA was performed with post-test total as the dependent variable, group membership as the fixed variable, and pre-test total as the covariate, to check whether there was any effect of the pre-test baseline differences on the post-test results. The group effect was still highly significant after controlling for baseline performance, $F(1, 57) = 94.26$, $p < .001$, partial $\eta^2 = .62$. That is, the SpeakingPal intervention explains about 62% of the variance in post-test scores over and above what participants contributed to the study. The results are presented in Table 6.

Table 6
ANCOVA Results – Post-Test Total Score Controlling for Pre-Test

Source	SS	df	MS	F	Partial η^2
Pre-test (Covariate)	12.47	1	12.47	3.82*	.063
Group (Exp. vs. Ctrl.)	**307.49**	**1**	**307.49**	**94.26*** **	**.623**
Error	185.97	57	3.26	—	—
Total	8,234.00	60	—	—	—

Note. SS = Sum of Squares; df = degrees of freedom; MS = Mean Square. *p < .05. ***p < .001.

Learner Engagement Survey Results

Descriptive statistics for the Learner Engagement Questionnaire (experimental group, n = 30) revealed high levels of perceived engagement, motivation, and usability across all four subscales. The overall mean engagement score was M = 4.32 (SD = 0.41), indicating that students strongly agreed SpeakingPal had enhanced their speaking learning experience. The highest-rated subscale was Ease of Use (M = 4.51, SD = 0.38)—a finding that matters practically, because an application that students find difficult to use will not be used enough to produce gains. Perceived Usefulness followed (M = 4.43, SD = 0.42), then Motivation and Enjoyment (M = 4.28, SD = 0.47), and Preference over conventional methods (M = 4.07, SD = 0.52). Item-level statistics are presented in Table 7.

Table 7
Learner Engagement Questionnaire Results (Experimental Group, n = 30)

Item	M	SD	Subscale
SpeakingPal improved my spoken English fluency.	4.60	0.49	Perceived Usefulness
AI feedback helped me identify my pronunciation errors.	4.53	0.51	Perceived Usefulness
Using SpeakingPal increased my speaking practice time.	4.47	0.57	Perceived Usefulness
The app helped me understand my speaking strengths/weaknesses.	4.40	0.50	Perceived Usefulness
I believe SpeakingPal is an effective learning tool.	4.17	0.65	Perceived Usefulness
I enjoyed practicing speaking through SpeakingPal.	4.43	0.50	Motivation & Enjoyment
SpeakingPal made learning English more interesting.	4.33	0.55	Motivation & Enjoyment
I felt motivated to practice because of the application.	4.27	0.58	Motivation & Enjoyment
I felt less anxious speaking English with the app than in class.	4.20	0.61	Motivation & Enjoyment
I would recommend SpeakingPal to other English learners.	4.17	0.59	Motivation & Enjoyment
The SpeakingPal interface was easy to navigate.	4.67	0.47	Ease of Use
The AI pronunciation feedback was clear and understandable.	4.57	0.50	Ease of Use
I could use SpeakingPal without technical assistance.	4.50	0.51	Ease of Use
The application ran smoothly on my phone.	4.43	0.57	Ease of Use
Feedback was provided quickly without noticeable delay.	4.37	0.61	Ease of Use
I prefer SpeakingPal over Conventionalspeaking drills.	4.23	0.57	Preference
SpeakingPal offers more practice than class instruction alone.	4.17	0.59	Preference
I would choose mobile learning over textbook-based practice.	4.03	0.67	Preference
SpeakingPal is more beneficial than teacher-only feedback.	3.97	0.72	Preference
I plan to continue using SpeakingPal after this study.	3.90	0.80	Preference

Note. Scale: 1 = Strongly Disagree to 5 = Strongly Agree. All items n = 30.

Qualitative Interview Insights

The analysis of the 12 semi-structured interviews, thematically, led to three main themes that validated and enriched the quantitative results.

Theme 1: Immediate Feedback as a Catalyst for Improvement. This theme focuses on the idea of immediate feedback as a catalyst for improvement. Most participants, 91.7% (11/12), saw the AI pronunciation feedback as the most helpful feature of SpeakingPal. A representative comment was given by one participant: "My teacher can't do that for all the other kids in a class of 40, but the app told me exactly what sound I said wrong." Repeatedly, immediate feedback, which large-enrolment classrooms do not usually have in their structure, appeared as a mechanism of learning.

Theme 2: Reduced Speaking Anxiety and Increased Autonomy. Nine participants (75.0%) found that they felt less anxious when speaking with SpeakingPal as opposed to speaking in class. Students reported that they were able to experiment with pronunciation with fewer inhibitions and, in turn, take risks. What many of the others said in other terms was summarized by one participant: "The app does not mock my errors." This perceived psychological safety is important because one of the best documented inhibitors of oral L2 performance is speaking anxiety (Horwitz et al., 1986).

Theme 3: Technical and Contextual Challenges. Practical concerns were found to be 58.3% with 7 participants reporting intermittent connectivity problem which impacted on the performance of applications, which is an intrinsic part of the infrastructure condition at the University of Mirpurkhas. Four participants (33.3%) reported experiencing

occasional misrecognition of their speech by the ASR system, especially in noisy backgrounds or when the regional accent characteristics differed from the acoustic model of the system. These are all known problems of existing ASR in multilingual scenarios (Liu, 2016), which indicate clear directions for future developments.

Discussion

Effectiveness of SpeakingPal on Overall Speaking Performance

The major result of the study was that the E group significantly outperformed the C group in terms of the increase in English speaking scores ($t(58) = 9.71$, $p < .001$, $d = 2.51$), thus offering strong empirical evidence for H_1. This effect size is significantly higher than the effect sizes found in recent meta-analyses on MALL for language achievement (0.50 – 0.82, Sung et al., 2016; Liu et al., 2020), indicating that the SpeakingPal mode (interactive and feedback-intensive) generates stronger gains than more passive mobile learning approaches.

The results form a cohesive theoretical framework. Vygotsky (1978) in his constructivist perspective, states that the iterative process of practice, feedback and revision, which SpeakingPal allows, is the key to skill acquisition. These gains are confirmed by the ANCOVA results ($F(1,57) = 94.26$, $p < .001$, $\eta^2 = .62$) which indicates that these gains are due to the intervention and not due to group differences.

Fluency, Pronunciation Accuracy, and Communicative Competence

The experimental group showed highest levels of gains in fluency ($\Delta = 1.87$ on the sub-score compared to $\Delta = 1.07$ for controls, and $d = 1.74$ between the groups). This is consistent with Liakin et al. (2015, 2017) in that ASR-based feedback mainly benefits fluency development by increasing the frequency of practice and decreasing hesitation anxiety, which was directly attested to by the qualitative theme of reduced speaking anxiety in the present study. A quarter of the interview participants indicated that they were less anxious in AI feedback than in classroom speaking tasks, and it's important to stop and think about this one practically: if anxiety is suppressed, students practice more, and practice with less cognitive interference and improve faster. It is the feedback-anxiety-practice loop that SpeakingPal establishes that is perhaps more important than the phonological analysis itself.

Pronunciation accuracy gains (between-groups $d = 1.00$) were significant but slightly lower in comparison to fluency gains. This is probably a true technical limitation – seven speakers indicated that they had experienced times when SpeakingPal had misrecognised their voice, resulting in inaccurate feedback. This is in line with the reported problems with ASRs trained mainly on standard American or British English while performing ASR on L2 speech with a Sindhi accent (Liu, 2016; Cucchiarini et al., 2009). In the South Asian setting, arguably the greatest technical challenge that the SpeakingPal developers must overcome is that of enhancing the accuracy of ASR for the various regional accent varieties.

Learner Engagement as a Mediating Factor

The quantitative performance data is supported by the high engagement scores (overall $M = 4.32/5.00$; Cronbach's $\alpha = .89$) which suggest that engagement is not only an outcome of using the SpeakingPal but is likely a key mediating mechanism – ease of use, usefulness, and enjoyment increase use, and use increases performance. Ease of Use ($M = 4.51$) was the highest rated subscale signifying the application's interface did not present any meaningful barrier to learning in line with HCI theory (Norman, 1988; Mayer, 2009). The ratings on Preference over Conventional Methods ($M = 4.07$) are low and serve as a

valuable reminder to not overestimate the use of SpeakingPal as a pedagogy; students appreciated it as a supplement to teacher-led instruction, rather than as a replacement for that instruction. Teachers who want to integrate MALL into their lessons should not take it lightly.

Implications for Practice

For ESL educators: For ESL teachers, the results suggest that SpeakingPal can be used as an additional speaking practice resource, and that it is best used in a more structured approach in larger classes where there is not much scope for individual feedback. The model employed here is of eight weeks, four sessions per week, which can be implemented in a regular academic semester without having to change the existing syllabuses.

For curriculum designers: MALL components are designed best to complement, not replace, teacher-led instruction. Thematic alignment used in this study offer a replicable framework for principled integration of MALL in first-year undergraduate language programs.

For EdTech developers: EdTech developers should take this study as a reminder of two key points: (1) make ASR more accurate with regional varieties of accents and (2) make applications perform better on low bandwidth mobile connections. Both can be addressed as engineering issues, and both would make a quantifiable difference to the learning outcomes in situations such as Mirpurkhas.

For institutional policymakers: Infrastructure is important. To achieve equitable MALL access, reliable internet connection and devices provision must be taken as preconditions, and digital infrastructure should be viewed not as an administrative add-on but as a resource for teaching and learning in institutions in developing contexts.

Conclusion

The findings of this quasi-experimental study give strong empirical evidence that AI-based MALL using SpeakingPal can improve the speaking proficiency of first-year undergraduate students of ESL at the University of Mirpurkhas as compared to the conventional classroom only approach. The results revealed statistically significant differences between the different groups on all four aspects of speaking (fluency, pronunciation accuracy, communicative competence, and coherence/cohesion) with large effect size ($d = 1.00$ to 2.51) and good level of acceptance by the learners (engagement $M = 4.32/5.00$). ANCOVA verified that about 62% of the variance in the post-test scores is due to the intervention alone.

Qualitatively, the picture is also clear: students appreciated the immediate feedback from AI, the psychological comfort of practising without being judged, and the convenience of using a mobile device, all of which are structural hallmarks of AI-based speaking training in high-volume settings as opposed to conventional classroom-based speaking instruction. However, technical problems such as accuracy in ASR for Sindhi-accented speech and intermittent Internet connectivity are real issues, which future versions of the tool and future studies should take into direct account.

The results add to the empirical studies on the use of AI in education as it shows that an ASR-based speaking practice tool can be feasible and meaningful in a real, resource-limited environment. They form a basis for evidence-based technology use in ESL programs and a template for future developments of intelligent tutoring systems for language learning. Ideal next steps include longitudinal follow-up, larger and more diverse samples,

and a proper examination of usage analytics as mediators, to better understand the mechanisms at play in the relationship between AI-based MALL and speaking development.

Recommendation

SpeakingPal should be incorporated as a structured supplement to teacher-led speaking instruction, particularly in large ESL classes where individual oral practice and immediate feedback are limited. A practical implementation model is four guided sessions of approximately 30 minutes per week, aligned with course themes and followed by teacher-led reflection and communicative activities.

Institutions should ensure equitable access to compatible devices, stable internet connectivity, and basic technical orientation before adopting AI-based MALL. Teachers should monitor participation and learner progress while retaining responsibility for contextualized feedback, communicative strategy development, and support when automated feedback is inaccurate.

Developers should improve automatic speech-recognition performance for Pakistani and other regional English accents and optimize the application for noisy and low-bandwidth environments. Future research should include larger, more diverse, and multi-institutional samples; longer interventions and delayed post-tests; and analyses of application-use logs to determine how practice frequency, feedback uptake, anxiety, and engagement contribute to speaking development.

References

Al-Harbi, S. S. (2019). The impact of SpeakingPal on developing Saudi EFL learners' oral fluency. *International Journal of English Language Teaching, 7*(3), 12–28.

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology, 3*(2), 77–101. https://doi.org/10.1191/1478088706qp063oa

British Council. (2015). *The state of English in Pakistan.* British Council.

Campbell, D. T., & Stanley, J. C. (1963). *Experimental and quasi-experimental designs for research.* Rand McNally.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Lawrence Erlbaum Associates.

Creswell, J. W., & Creswell, J. D. (2018). *Research design: Qualitative, quantitative, and mixed methods approach* (5th ed.). SAGE Publications.

Cucchiarini, C., Strik, H., & Boves, L. (2009). Oral proficiency training in Dutch L2: The contribution of read-aloud exercises. *Language Learning, 59*(4), 874–902. https://doi.org/10.1111/j.1467-9922.2009.00527.x

Heaton, J. B. (1988). *Writing English language tests*. Longman.

Horwitz, E. K., Horwitz, M. B., & Cope, J. (1986). Foreign language classroom anxiety. *The Modern Language Journal, 70*(2), 125–132. https://doi.org/10.1111/j.1540-4781.1986.tb05256.x

ITU. (2024). *Measuring digital development: Facts and figures 2024.* International Telecommunication Union.

Jonassen, D. H. (1999). *Designing constructivist learning environments. In C. M. Reigeluth (Ed.), Instructional design theories and models: A new paradigm of instructional theory* (Vol. 2, pp. 215–239). Lawrence Erlbaum Associates.

Kukulska-Hulme, A., & Shield, L. (2008). An overview of mobile assisted language learning: From content delivery to supported collaboration and interaction. *ReCALL, 20*(3), 271–289. https://doi.org/10.1017/S0958344008000335

Liakin, D., Cardoso, W., & Liakina, N. (2015). Learning L2 pronunciation with a mobile speech recognizer: French /y/. *CALICO Journal, 32*(1), 1–25. https://doi.org/10.1558/cj.v32i1.25962

Liakin, D., Cardoso, W., & Liakina, N. (2017). The pedagogical use of mobile speech synthesis (TTS): Focus on French liaison. *Computer Assisted Language Learning, 30*(3–4), 348–365. https://doi.org/10.1080/09588221.2017.1308062

Liu, M., Abe, K., Cao, M., Liu, S., Ok, D., Park, J., Parrish, C., & Sardegna, V. G. (2020). An analysis of social presence effects in one-on-one online English video tutoring. *CALICO Journal, 37*(1), 1–21.

Liu, P. (2016). Mobile English vocabulary learning based on concept-mapping strategy. *Language Learning & Technology, 20*(3), 128–141.

Mayer, R. E. (2009). *Multimedia learning (2nd ed.).* Cambridge University Press.

Memon, M. A., Bhutto, S. A., & Soomro, N. H. (2022). Technology integration in ESL classrooms: Challenges in Pakistani public sector universities. *Pakistan Journal of Education, 39*(2), 87–104.

Nation, I. S. P., & Newton, J. (2009). *Teaching ESL/EFL listening and speaking*. Routledge.

Neri, A., Cucchiarini, C., Strik, H., & Boves, L. (2002). The pedagogy-technology interface in computer assisted pronunciation training. *Computer Assisted Language Learning, 15*(5), 441–467. https://doi.org/10.1076/call.15.5.441.13473

Norman, D. A. (1988). *The design of everyday things*. Basic Books.

Piaget, J. (1954). *The construction of reality in the child.* Basic Books.

Rosell-Aguilar, F. (2018). Autonomous language learning through a mobile application: A user evaluation of the Busuu app. *Computer Assisted Language Learning, 31*(8), 854–881. https://doi.org/10.1080/09588221.2018.1456465

Shah, S. A., Abidin, M. J. Z., & Javed, F. (2021). ESL speaking anxiety among Pakistani undergraduates: Causes and coping strategies. *International Journal of English Language Education, 9*(1), 43–62.

Sharples, M., Taylor, J., & Vavoula, G. (2007). A theory of learning for the mobile age. In R. Andrews & C. Haythornthwaite (Eds.), *The SAGE handbook of e-learning research* (pp. 221–247). SAGE Publications.

Skehan, P. (1998). *A cognitive approach to language learning*. Oxford University Press.

Sung, Y. T., Chang, K. E., & Liu, T. C. (2016). The effects of integrating mobile devices with teaching and learning on students' learning performance: A meta-analysis and research synthesis. *Computers & Education, 94*, 252–275.

Viberg, O., & Grönlund, Å. (2013). Systematising the field of mobile assisted language learning. *International Journal of Mobile and Blended Learning*, 5(4), 72–82.

Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Harvard University Press.