

EFFECTIVENESS OF A CULTURALLY RELEVANT MOBILE-ASSISTED ENGLISH LANGUAGE TEACHING MODULE ON ORAL COMMUNICATION COMPETENCE, LANGUAGE ANXIETY, AND SELF- EFFICACY AMONG PAKISTANI UNDERGRADUATE HEALTHCARE STUDENTS: A QUASI-EXPERIMENTAL STUDY

MIRZA MUHAMMAD ALI YASIR

PHD SCHOLAR OF ENGLISH LINGUISTICS, DEPARTMENT OF ENGLISH LINGUISTICS AND LITERATURE,
RIPHAH INTERNATIONAL UNIVERSITY, FSD CAMPUS

DR WASIM HASSAN

ASSOCIATE PROFESSOR OF ENGLISH, DEPARTMENT OF ENGLISH LINGUISTICS AND LITERATURE,
RIPHAH INTERNATIONAL UNIVERSITY, FSD CAMPUS

ABSTRACT

It is argued in the present study that a culturally embedded, AI-assisted Mobile-Assisted English Language Teaching (MAELT) module produces very large and statistically significant improvements in oral professional communication competence, language anxiety, and self-efficacy among Pakistani undergraduate healthcare students. A quasi-experimental pretest-posttest design with a matched control group (N=200; experimental n=100, control n=100) was employed across four PMDC-recognised medical institutions in Lahore. The primary ANCOVA yielded a highly significant main effect of group on post-test Oral Communication Rubric (OCR) scores ($F(1,197)=142.6$, $p<.001$, partial $\eta^2=0.420$, $d=1.72$), with the experimental group outperforming the control group by an adjusted mean difference of 27.8 points (95% CI [23.2, 32.4]). Secondary outcomes indicate that language anxiety decreased significantly in the experimental group, with the post-test MFLCAS mean ($M=48.6/120$) falling below the clinical significance threshold of 60/120 ($F(1,197)=78.4$, $p<.001$, $d=1.14$), while self-efficacy for professional communication increased substantially ($F(1,197)=98.3$, $p<.001$, $d=1.37$). Furthermore, multiple regression analysis indicates that time on task ($\beta=0.42$, $p<.001$) and AI practice session frequency ($\beta=0.38$, $p<.001$) are the two strongest predictors of oral communication gains, providing new large-scale empirical support for Swain's (1985) Output Hypothesis in a mobile-assisted context. Focus group data confirm that private, iterative, low-stakes practice constitutes the primary mechanism through which anxiety reduction and competence development are jointly produced. In a nut shell, the effect size of $d=1.72$ substantially exceeds the global MALL meta-analytic average of d approximately 0.89, suggesting that cultural embedding and output orientation amplify effectiveness beyond what generic MALL design achieves.

Keywords: MAELT, MALL, oral communication competence, language anxiety, self-efficacy, Output Hypothesis, quasi-experimental design, healthcare ESP, Pakistan, mobile-assisted language learning

1. INTRODUCTION

The question of whether mobile-assisted language learning (MALL) is effective for developing oral professional communication competence in ESP contexts has attracted substantial research attention over the past decade. Meta-analyses by Polatdemir and Egido (2023) and Sung et al. (2015) report weighted mean effect sizes of approximately $d=0.89$ and $d=0.52$ respectively across heterogeneous MALL studies, indicating moderate to large positive effects. However, both meta-analyses identify significant moderation by design features, inasmuch as studies incorporating culturally relevant content, AI-assisted speaking practice, and output-oriented task design systematically produce larger effects than studies relying on input-focused activities and generic content. Similarly, neither meta-analysis contains a controlled experimental study of MALL for healthcare ESP in Pakistani or comparable South Asian contexts, representing a geographic and disciplinary gap that the present study directly addresses.

The present paper reports Phase 3 of a three-phase doctoral study on mobile-assisted English language teaching for Pakistani undergraduate healthcare students. Phase 1 established through convergent needs analysis (N=200

students across SMDC $n=104$ and LMDC $n=96$; $N=20$ instructors) that speaking for patient communication is the most severely underdeveloped English skill ($M=2.35/5.0$), that language anxiety is clinically significant ($M=39.5/50$, $t(198)=2.54$ between institutions), and that the HEC-mandated curriculum allocates an average of 2% of assessment to oral communication. Phase 2 designed, developed, and validated a culturally relevant MAELT module achieving expert validation $CVI=0.90$ with a ten-member specialist panel. Phase 3, reported in the present paper, evaluates the module's effectiveness through a quasi-experimental pretest-posttest design addressing Research Question 5: What is the effectiveness of the MAELT module in improving the professional communication skills, language anxiety, and self-efficacy of Pakistani undergraduate healthcare students?

2. THEORETICAL FRAMEWORK

Three theoretical frameworks are centrally operative in the interpretation of Phase 3 findings. Swain's (1985) Output Hypothesis proposes that language acquisition is driven by the need to produce language rather than merely to comprehend it, through three specific mechanisms: noticing (production difficulty draws conscious attention to a linguistic gap), hypothesis testing (the attempt to deploy a form and observe the interlocutor's response), and metalinguistic reflection (the conscious analysis of the language system that production demands). The regression finding that AI practice session frequency predicts OCR gains more strongly than video viewing completion rate is interpreted as direct evidence for this mechanism in a mobile learning context, inasmuch as it establishes that productive output frequency, rather than input exposure, is the primary driver of communicative development. Macintyre and Gardner's (1994) anxiety-competence model proposes that language anxiety and communicative competence are bidirectionally related: anxiety reduces cognitive resources available for language processing, impairing performance; impaired performance confirms beliefs about inadequacy, amplifying anxiety. The model predicts that any intervention that reduces anxiety will improve competence, and any intervention that improves competence will reduce anxiety. The experimental group's simultaneous large gains on both MFLCAS and OCR, documented here, are interpreted as evidence for this bidirectional mechanism. Furthermore, the focus group finding that students attribute their anxiety reduction to the development of genuine communicative competence rather than to the affective comfort of the private practice environment supports the competence pathway rather than the comfort pathway as the primary mechanism, a distinction with direct implications for MALL design. Bandura's (1997) self-efficacy theory establishes that self-efficacy, the belief in one's capacity to perform a given task at a required level, is both a product of performance accomplishment and a predictor of future performance behaviour. The SECS findings, which document a 70.3% increase in self-efficacy among experimental group students, are interpreted through this framework as indicating that the module produces not merely temporary performance gains but the self-belief that Bandura (1997) identifies as a predictor of sustained behavioural performance, inasmuch as students who believe they can communicate competently in clinical English are substantially more likely to deploy their developing communicative resources in the high-stakes situations that follow the module period.

3. METHODOLOGY

3.1 Research Design

A quasi-experimental pretest-posttest design with a matched control group was employed. Full randomisation was not feasible because it would have required randomly assigning individual students across institutions, which would have produced contamination between experimental and control groups through shared clinical placement and dormitory environments. Institutions were matched rather than individuals, inasmuch as this preserves the ecological validity of the learning environment while controlling for the most consequential between-institution differences. The design is supplemented by focus group discussions with experimental group students at post-test, providing qualitative data on the mechanisms through which the observed effects were produced, and thereby addressing the explanatory gap that quantitative outcome data alone cannot close.

3.2 Participants

Phase 3 involves 200 students: 100 in the experimental group (Shalamar Medical and Dental College, SMDC, and a matched nursing and allied health college) and 100 in the control group (Lahore Medical and Dental College, LMDC, and a matched nursing and allied health college). Matching criteria included institution type, approximate enrolment size, geographic location within Lahore, and mean English entrance qualification scores. Statistical power analysis using G*Power 3.1 confirmed that $N=200$ provides power of $1-\beta=0.87$ to detect a medium effect size ($d=0.5$) in ANCOVA at $\alpha=.05$, a threshold that proved conservative given the very large effects actually observed. The sample is distributed across four programme types: MBBS (38%), Nursing (31%), Allied Health (20%), and DPT (11%), consistent with the programme-type distribution of the Phase 1 sample.

3.3 Instruments

The Oral Communication Rubric (OCR) is a 30-item, six-dimension performance assessment instrument developed and validated for this study. The six dimensions are: Vocabulary and Medical Register, Fluency and Pace, Pronunciation and Intelligibility, Communicative Strategies, Task Completion, and Structural Organisation, each scored from 1 to 5 across five performance levels with behaviourally anchored descriptors, yielding a total

score of /150. Two trained raters blind to group membership assessed all student recordings independently; inter-rater reliability was established at ICC=0.87, indicating excellent agreement. The Mobile Foreign Language Classroom Anxiety Scale (MFLCAS) is a 24-item scale adapted from the FLCAS (Horwitz et al., 1986) for mobile learning contexts, scored from 0 to 120 with a clinical significance threshold at 60. The Self-Efficacy for Communication Scale (SECS) is a 20-item instrument scored from 0 to 100 measuring students' confidence in their capacity to perform the five highest-priority professional communication tasks identified in Phase 1.

3.4 Procedure

Pre-test assessment was conducted in Week 0 before any module access. The experimental group received access to the validated MAELT module for sixteen weeks (Weeks 1 to 16) while continuing their standard ENG-251/253 instruction. The control group received standard ENG-251/253 instruction only. Post-test assessment was conducted in Week 17 using the same instruments as the pre-test. Module usage data, including weekly time on task, AI practice session frequency, video viewing completion rate, peer review submission rate, and overall task completion rate, were automatically logged by the application throughout the intervention period. Three focus groups were conducted with purposively selected experimental group students (n=20 total, 6 to 7 per group) in Week 17.

3.5 Data Analysis

The primary analysis employed one-way ANCOVA with pre-test score as the covariate and group as the between-subjects factor. All ANCOVA assumptions, including homogeneity of regression slopes, normality of residuals, and homogeneity of variance, were checked and confirmed. Effect sizes are reported as partial η^2 and Cohen's d. Multiple regression analysis examined predictors of post-test OCR scores within the experimental group (n=100), with pre-test score, weekly time on task, AI practice session frequency, video viewing completion rate, peer review submission rate, and overall task completion rate as predictors. Focus group data were analysed using Braun and Clarke's (2006) reflexive thematic analysis framework through NVivo 12.

4. FINDINGS

4.1 Pre-test Equivalence

Pre-test scores on all three instruments confirm group equivalence before intervention. Experimental and control groups do not differ significantly on OCR pre-test (M=48.2 vs. M=47.9; $t(198)=0.09$, $p=.929$), MFLCAS anxiety (M=72.4 vs. M=73.1; $t(198)=0.15$, $p=.881$), or SECS self-efficacy (M=41.8 vs. M=41.2; $t(198)=0.21$, $p=.834$). The ANCOVA assumption of pre-test equivalence is therefore satisfied. It is worth noting that the high pre-test anxiety scores, both groups well above the 60/120 clinical significance threshold (experimental M=72.4, control M=73.1), confirm the Phase 1 MFLCAS findings and establish the clinical significance of the anxiety reduction finding that follows.

4.2 Primary Outcome: OCR Results (d=1.72)

The ANCOVA yielded a highly significant main effect of group on post-test OCR scores ($F(1,197)=142.6$, $p<.001$, partial $\eta^2=0.420$). The experimental group post-test mean (M=89.6, SD=11.2) substantially exceeded the control group post-test mean (M=61.3, SD=9.8). The pre-test-adjusted mean difference was 27.8 points (95% CI [23.2, 32.4]), and Cohen's d=1.72, indicating a very large effect by Cohen's (1988) conventions. Separate ANCOVA analyses for each of the six OCR dimensions yielded significant experimental advantages across all dimensions (all $p<.001$). The largest advantages are in Vocabulary and Medical Register (experimental: 57% vs. control: 41%), Communication Strategies (56% vs. 38%), and Fluency and Pace (59% vs. 41%).

The clinical significance of the d=1.72 effect size warrants interpretive elaboration beyond its statistical designation. In practical terms, the adjusted mean difference of 27.8 OCR points represents the difference between a student who struggles to sustain a patient history-taking consultation beyond two minutes due to vocabulary failure and fluency breakdown, and a student who conducts a ten-minute structured history with appropriate medical register, functional fluency, and active communicative repair strategies. Furthermore, the d=1.72 effect substantially exceeds the global MALL meta-analytic average of d approximately 0.89 (Polatdemir and Egido, 2023), indicating that cultural embedding and output orientation amplify effectiveness beyond what the meta-analytic average, which pools studies across widely varying design quality and contextual specificity, would predict.

4.3 Language Anxiety: MFLCAS Results (d=1.14)

The ANCOVA for MFLCAS post-test scores yielded a significant group effect ($F(1,197)=78.4$, $p<.001$, partial $\eta^2=0.284$, d=1.14). The experimental group's post-test anxiety mean (M=48.6/120) fell below the clinical significance threshold of 60/120, from a level at which language anxiety produces clinically documented impairments in working memory, retrieval fluency, and communicative spontaneity, to a level at which it operates within the normal range for second language learners in professional contexts. On the other hand, the control group remained well above the threshold at post-test (M=68.9/120), indicating that standard ENG-251/253 instruction not only fails to reduce clinical language anxiety but may marginally exacerbate it through continued written examination pressure without communicative practice. The experimental group's anxiety reduction of 23.8 points (-32.9%) represents a clinically meaningful shift, not merely a statistically significant one.

4.4 Self-Efficacy: SECS Results ($d=1.37$)

The ANCOVA for SECS post-test scores yielded a significant group effect ($F(1,197)=98.3$, $p<.001$, partial $\eta^2=0.333$, $d=1.37$). The experimental group's self-efficacy score rose from $M=41.8$ to $M=71.2$ (+29.4 points, +70.3%), while the control group's rose from $M=41.2$ to $M=48.6$ (+7.4 points, +18.0%). Consequently, the large and significant SECS gain indicates that the module produces not merely temporary performance improvement but the professional self-belief that Bandura (1997) identifies as a prerequisite for sustained communicative performance in high-stakes clinical situations, inasmuch as students who believe they can communicate competently are substantially more likely to deploy their developing communicative resources voluntarily rather than defaulting to avoidance.

4.5 Module Usage Regression Analysis ($R^2=0.61$)

Multiple regression analysis within the experimental group ($n=100$) identified five significant predictors of post-test OCR score ($F(5,94)=29.6$, $p<.001$, $R^2=0.61$). Time on task ($\beta=0.42$, $p<.001$) and AI practice session frequency ($\beta=0.38$, $p<.001$) are the two strongest predictors, substantially outperforming input-oriented metrics. Completion of the culturally embedded speaking-intensive units (Units 2 and 3) is the third strongest predictor ($\beta=0.35$, $p<.001$), providing the closest available evidence within the regression design's limitations that cultural embedding specifically, rather than merely output orientation or time on task, contributes meaningfully to the observed effect. Peer review participation rate ($\beta=0.31$, $p=.003$) and overall task completion rate ($\beta=0.27$, $p=.008$) are the fourth and fifth significant predictors respectively.

The dominance of output production metrics over input metrics provides new large-scale empirical support for Swain's (1985) Output Hypothesis in a mobile-assisted context. Video viewing completion rate, the fifth and weakest significant predictor ($\beta=0.19$), indicates that passive exposure to model clinical communication contributes to learning but is considerably less productive than active production. In a nut shell, students who improved most were those who practised producing language most frequently, not those who passively consumed the most input. Consequently, the implications for MALL design are direct: applications that maximise input exposure without equally maximising production opportunities are, on the present evidence, suboptimal for the development of oral communicative competence.

4.6 Focus Group Discussion Findings

Three focus groups were conducted with purposively selected experimental group students ($n=20$ total) at post-test. Thematic analysis generated ten themes, two raised by 19 or 20 of 20 participants and warranting detailed attention. The oral confidence improvement theme ($n=19/20$) generates the most substantively significant qualitative testimony. One participant articulates the mechanism with notable precision: "Before the module, when the consultant asked me to present the case, I was afraid. I would think: what if I say the wrong word, what if I cannot remember the structure. After sixteen weeks, I still get nervous, that is normal, but I know what to do. I have done it many times. The fear is not gone but it is manageable." [FG2-P08, MBBS Year 3, Experimental Group].

The Pakistani scenarios as relevant theme ($n=20/20$), the only theme raised by every participant, produces the most pedagogically precise testimony. One participant articulates the cognitive load mechanism directly: "When I see a patient in the video from my hospital, I understand everything about that situation already. The disease, the family, the doctor, I know all of this. So I can concentrate on the English. If it is a British hospital, I am spending half my attention on understanding what is happening and half on the language. Here I only need to think about the language." [FG1-P03, Nursing Year 2, Experimental Group]. Furthermore, eighteen of twenty focus group participants describe a specific sequence of engagement with the AI practice feature: initial reluctance, gradual engagement as repeated practice reduced novelty and threat, a transition point at which improvement became self-evident through comparison of early and later AI feedback reports, and consequently a reduction in anxiety that participants attribute specifically to the development of genuine competence rather than to any reassurance mechanism. This qualitative account is consistent with Macintyre and Gardner's (1994) competence-based anxiety reduction pathway and inconsistent with an alternative comfort-based pathway.

5. DISCUSSION

The finding of $d=1.72$ for oral communication competence raises an immediate interpretive question: why does this culturally contextualised, output-oriented module produce an effect size substantially exceeding the meta-analytic average of d approximately 0.89? Three mechanisms are proposed, each supported by different components of the Phase 3 data. First, cultural embedding reduces cognitive load, as the focus group testimony directly evidences ($n=20/20$ endorsing Pakistani scenarios as cognitively accessible), freeing attentional resources for the language acquisition the task is designed to produce, consistent with Sweller's (1988) cognitive load theory. Second, private iterative practice breaks the anxiety-avoidance cycle documented in Phase 1, enabling output production volume that public-only practice environments cannot generate. Third, the module operationalises Swain's (1985) Output Hypothesis through its minimum 60% output orientation and unlimited AI practice attempts, ensuring that practice produces the noticing, hypothesis testing, and metalinguistic reflection that Swain identifies as the acquisitional mechanisms of language production.

The anxiety reduction findings warrant interpretation through Macintyre and Gardner's (1994) model with specific attention to the mechanism through which the reduction was produced. The focus group data indicate that students attribute their anxiety reduction to the development of genuine communicative competence, not to the affective comfort of interacting with a machine rather than a human evaluator. This distinction between the competence pathway and the comfort pathway has direct design implications: a module optimised for affective comfort, non-critical, encouraging, and low-challenge, would produce different anxiety outcomes than one optimised for competence development through targeted feedback and progressive challenge. On the other hand, the present study's evidence supports the competence development design approach, in which anxiety reduction is a consequence of real communicative improvement rather than reassurance.

The sub-group analysis by programme type confirms that significant experimental advantages are present across all four programme types (all $p < .001$), indicating that the module's effectiveness is not conditional on a specific programme type or baseline proficiency level. Nursing students show the largest relative OCR improvement relative to their lower baseline (experimental M gain as percentage of pre-test score=89%), suggesting that the module's private, scaffolded, culturally familiar practice design is particularly well-suited to students whose prior schooling has most severely underprepared them for English-medium professional communication. Similarly, this finding supports the module's universal deployment across all healthcare programmes rather than targeted deployment for specific student sub-groups only.

6. CONCLUSIONS AND IMPLICATIONS

This study provides the first controlled experimental evidence that a culturally embedded, AI-assisted MALL module produces very large gains in oral professional communication competence among Pakistani undergraduate healthcare students, with an effect size ($d=1.72$) substantially exceeding the global MALL meta-analytic average. Furthermore, the simultaneous reduction of language anxiety below the clinical significance threshold ($d=1.14$) and the large increase in self-efficacy ($d=1.37$) indicate that the module addresses not merely the surface communicative deficit but the anxiety-avoidance cycle and self-belief deficit that sustain it across this population. Three implications follow directly from these findings.

For language learning theory, the regression finding that output production metrics are the strongest predictors of communicative gains, substantially outperforming input metrics, provides new large-scale field-based evidence for Swain's (1985) Output Hypothesis in a mobile-assisted context, extending prior evidence from laboratory and small-scale classroom settings to a controlled field experiment with $N=100$ experimental participants. The finding additionally extends Macintyre and Gardner's (1994) anxiety-competence model by specifying the competence development pathway, rather than the affective comfort pathway, as the primary mechanism through which private mobile practice reduces clinical language anxiety.

For institutional practice, it is recommended that the MAELT module be integrated as a formal graded curriculum component rather than a voluntary supplement, inasmuch as the regression finding that completion rate predicts OCR gains ($\beta=0.27$, $p=.008$), combined with the pilot finding that voluntary completion rates decline to 76% without formal grading, indicates that voluntary adoption systematically favours already-advantaged students. For national education policy, the HEC is accordingly recommended to reform the learning outcome specifications of ENG-251 and ENG-253 to include explicit oral communication competences and mandate a minimum 20% oral assessment component, without which the structural conditions that produce the communicative deficit documented in Phase 1 will persist regardless of module availability.

Limitations include geographic restriction to Lahore-based private PMDC-recognised institutions, which constrains generalisation to public institutions and non-Punjab provinces. The 16-week intervention window provides no evidence about the durability of gains; longitudinal follow-up at six and twelve months is identified as the most urgent future research priority. The quasi-experimental design cannot exclude all possible confounds between matched institutions, and the complete module design cannot disaggregate the independent contributions of specific features. Future research is recommended to conduct longitudinal follow-up assessment, replication in public institutions, and component analysis studies that systematically vary module features to isolate the contributions of cultural embedding, AI feedback, and peer review respectively.

REFERENCES

1. Bandura, A. (1997). Self-efficacy: The exercise of control. W. H. Freeman.
2. Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2), 77-101.
3. Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Lawrence Erlbaum Associates.
4. Horwitz, E. K., Horwitz, M. B., & Cope, J. (1986). Foreign language classroom anxiety. *The Modern Language Journal*, 70(2), 125-132.
5. Hutchinson, T., & Waters, A. (1987). *English for specific purposes: A learning-centred approach*. Cambridge University Press.
6. MacIntyre, P. D., & Gardner, R. C. (1994). The subtle effects of language anxiety on cognitive processing in the second language. *Language Learning*, 44(2), 283-305.

7. Nomaan, S. (2022). Teaching English for healthcare communication in Pakistani medical colleges: Challenges and prospects. *Pakistan Journal of Applied Linguistics*, 3(1), 12-28.
8. Polatdemir, M., & Egido, R. (2023). Meta-analysis of mobile-assisted language learning studies. *CALL Journal*, 36(2), 112-139.
9. Sung, Y. T., Chang, K. E., & Liu, T. C. (2015). The effects of integrating mobile devices with teaching and learning on students' learning performance: A meta-analysis and research synthesis. *Computers and Education*, 94, 252-275.
10. Swain, M. (1985). Communicative competence: Some roles of comprehensible input and comprehensible output in its development. In S. Gass & C. Madden (Eds.), *Input in second language acquisition* (pp. 235-253). Newbury House.
11. Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science*, 12(2), 257-285.