



CULTURALLY RELEVANT MOBILE-ASSISTED ENGLISH LANGUAGE TEACHING MODULE FOR PAKISTANI UNDERGRADUATE HEALTHCARE STUDENTS

Mirza Muhammad Ali Yasir
mirzaaliyasir7256@gmail.com

PhD Scholar, Department of English Linguistics and Literature, Riphah International University, Faisalabad Campus, Pakistan.

Dr Wasim Hassan

Associate Professor of English, Department of English Linguistics and Literature, Riphah International University, Faisalabad Campus, Pakistan.

Abstract

It is argued in the present study that culturally relevant, evidence-grounded module design constitutes a necessary condition for effective Mobile-Assisted Language Learning (MALL) in professional English for Specific Purposes (ESP) contexts, and that the Design-Based Research (DBR) methodology provides the appropriate framework for translating empirically documented needs into a validated educational product. This paper reports the design, development, and expert validation of a Mobile-Assisted English Language Teaching (MAELT) module for undergraduate healthcare students at Pakistani medical institutions. The module was designed through the ADDIE instructional design model nested within DBR, with the Four-D model providing the validation architecture. Nine non-negotiable design commitments, each traceable to at least two independent data sources from a preceding N=200 needs analysis, governed every design decision. Expert validation was conducted across two rounds by a ten-member specialist panel using the 45-item Expert Validation Scale (EVS). The overall Round 1 mean was 2.90/4.00 with a Content Validity Index (CVI) of 0.72, indicating that substantial revision was required; the overall Round 2 mean improved to 3.67/4.00 with CVI=0.90 following 21 documented revisions, indicating that all five dimensions met the required standards. Pilot testing with N=30 students at Shalimar Medical College (SMDC) Lahore, confirmed feasibility and produced six further targeted revisions. Furthermore, the validated module is argued to constitute the first culturally embedded, AI-assisted MALL resource specifically designed for Pakistani healthcare ESP, grounded in multi-instrument needs analysis, and validated to the standard required for experimental deployment.

Keywords: *Addie, Content Validity Index, Culturally Relevant Pedagogy, Design-Based Research, ESP Module Design, Expert Validation, Healthcare Education, Mall, Mobile Learning, Pakistan.*

Corresponding Author: Mirza Muhammad Ali Yasir (PhD Scholar, Department of English Linguistics and Literature, Riphah International University, Faisalabad Campus, Pakistan.)

Email: mirzaaliyasir7256@gmail.com

1. Introduction

The design of effective educational technology resources for language learning in professional contexts is a theoretically complex and practically demanding undertaking that requires simultaneous attention to learner needs, theoretical grounding, cultural relevance, technical accessibility, and pedagogical quality. In the specific context of English for Specific Purposes (ESP) for Pakistani healthcare students, commercially available mobile language learning applications address none of the documented communicative needs of this population specifically: patient history-taking in Lahore public hospitals, clinical case presentation before senior consultants in Pakistani teaching institutions, or interprofessional handover in Pakistani ICU settings. Likewise, the absence of a validated, contextually appropriate, mobile-assisted resource for this population is not merely a gap in the literature; it is a structural absence that the current curriculum has not filled and that commercially produced tools are not designed to fill (Hutchinson and Waters, 1987; Nomaan, 2022).

The present paper reports the second phase of a three-phase doctoral study on mobile-assisted English language teaching for Pakistani undergraduate healthcare students. Phase 1, reported separately, established through a convergent mixed-methods needs analysis (N=200 students across SMDC n=104 and LMDC n=96, N=20 instructors, five institutions' curricula) 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, with institutional differences of $t(198)=2.54$, $p=.013$, $d=0.45$), and that the HEC-mandated curriculum allocates an average of 2% of assessment to oral communication. The present paper addresses Research Questions 3 and 4: How can a culturally relevant MAELT module be designed to address these needs? and How can the module be developed to meet validated standards of content accuracy, cultural relevance, pedagogical soundness, technical usability, and assessment alignment?

2. Theoretical Framework

The module design draws on eight theoretical frameworks, three of which are most centrally operative. Swain's (1985) Output Hypothesis establishes that communicative competence develops through active language production under conditions that push learners to deploy their full linguistic resources, not through passive input exposure alone. The module operationalises this principle through a minimum 60% output orientation across all units, AI-assisted speaking practice with targeted feedback, and peer review of recorded clinical communication performances. Vygotsky's (1978) Zone of Proximal

Development establishes that learning occurs in the space between what a learner can accomplish independently and what they can accomplish with the guidance of a more knowledgeable other. The module operationalises this through AI pronunciation and fluency feedback as a responsive scaffold on every speaking attempt, structured peer review, and an instructor integration framework that positions classroom instruction as the elaboration of mobile practice rather than its replacement.

Ladson-Billings's (1995) culturally relevant pedagogy framework establishes that learners engage more effectively with content whose cultural context they recognise and inhabit, inasmuch as cultural familiarity reduces cognitive load and frees attentional resources for language acquisition (Sweller, 1988). Consequently, all module scenarios, diseases, settings, and professional communication conventions are drawn exclusively from Pakistani healthcare contexts. Furthermore, this cultural embedding is argued to constitute a cognitive load reduction strategy rather than merely a motivational amenity, inasmuch as the Phase 1 SSIG data confirm that students spend significantly less cognitive effort processing culturally familiar scenarios, thereby freeing resources for the language acquisition the task is designed to produce.

3. Design-Based Research Framework

Design-Based Research (DBR), as conceptualised by Brown (1992) and elaborated by the Design-Based Research Collective (2003), is a methodology that integrates the development and testing of educational innovations with the generation of theoretical knowledge about learning. It is distinguished from pure experimental research, which tests pre-specified interventions without the iterative revision process DBR requires, and from pure design research, which produces educational artefacts without the controlled evaluation that establishes their effectiveness. The three-phase architecture of the present study reflects DBR's defining commitments: Phase 1 generates the empirical foundation that Phase 2 translates into a design; Phase 2 validates and refines the design through expert review and pilot testing; Phase 3 evaluates the effectiveness of the validated design through quasi-experimental methodology.

The ADDIE model (Analysis, Design, Development, Implementation, Evaluation) provides the operational framework within which the DBR process is conducted. The Analysis phase corresponds to Phase 1 of the study; the Design, Development, and partial Evaluation phases correspond to Phase 2; Implementation and the remaining Evaluation functions correspond to Phase 3. The Four-D model (Define, Design, Develop, Disseminate) of Thiagarajan et al. (1974), applied to healthcare ESP module development by Rochmawati et al. (2024) in the Indonesian nursing context, provides the specific validation architecture for the expert review process.

4. Module Design Specification

4.1. Define Stage: Nine Non-Negotiable Design Commitments

The Define stage translated the Phase 1 convergent needs profile into nine non-negotiable design commitments, each traceable to at least two independent Phase 1 data sources. Cultural embedding (all scenarios from Pakistani healthcare contexts) responds to the SSIG finding that 18/20 students identify Pakistani clinical content as a critical need, and to the ISIG finding that 8/10 instructors attribute student disengagement specifically to contextual unfamiliarity. Communicative output focus (minimum 60% of active learning time in production tasks) responds to the SNCQ finding that speaking is the most severely underdeveloped skill ($M=2.35$). AI-assisted private practice (private recording mode as default; unlimited attempts before submission) responds to the SNCQ-F finding that private practice receives the highest MALL attitude endorsement ($M=4.53$) and to the SSIG finding that 18/20 students explicitly request the opportunity to practise privately before public performance.

Offline functionality (all core content downloadable) responds to the SNCQ-E finding that 90% of students want offline access and 38% report unreliable internet connectivity. Bilingual scaffolding (847-entry English-Urdu medical glossary with colloquial register for patient-level terms) responds to the SSIG finding that 15/20 students identify vocabulary retrieval failure as the primary trigger of anxiety spikes in clinical communication. Proficiency differentiation (two pathways from Unit 0 diagnostic) responds to the statistically significant proficiency gap between SMDC and LMDC students ($t(198)=2.12$, $p=.036$, $d=0.37$). Healthcare ESP task authenticity (all tasks drawn from genres identified in the SNCQ-C priority ranking) responds to the convergent Phase 1 finding that oral professional genres constitute all five highest-priority communication tasks. Instructor integration (instructor guide, 16-week integration plans, 4-hour professional development module) responds to the IPS finding that high conditional MALL adoption willingness ($M=4.40$) is paired with severely inadequate technology training ($M=1.75$).

4.2. Design Stage: Module Architecture

The module comprises nine content units plus an orientation unit, structured across sixteen weeks. Unit sequencing is determined by the SNCQ-C needs priority ranking: Unit 2 (Patient History-Taking) and Unit 3 (Clinical Case Presentation), addressing the two highest-priority tasks ($M=4.70$ and $M=4.68$ respectively), are positioned in weeks 3 to 6 to maximise practice time. Within each unit, a five-step pedagogical sequence operationalises the theoretical framework: contextualised input (Krashen, 1985), language analysis and vocabulary pre-loading, AI-assisted private speaking practice (Swain, 1985; low anxiety condition), structured peer review (Vygotsky, 1978), and assessed communicative performance task. This sequence is applied consistently across all eight content units, providing the procedural scaffolding that supports self-regulated learning (Zimmerman, 2000).

Video scenario production involved eight original productions with Pakistani actors, Lahore hospital settings, and medical consultant review for clinical accuracy. The scenarios address diseases of high prevalence in Pakistan: dengue fever, pulmonary tuberculosis, Type 2 diabetes with complications, acute myocardial infarction, hypertension, and road traffic accident injuries. The Speechace API was selected for AI speech recognition and feedback following comparative evaluation of four systems on Pakistani-accented English accuracy. The bilingual glossary comprises 847 entries, medically reviewed, with colloquial Urdu register for patient-level terms, based on the Phase 1 finding that formal written Urdu medical terminology does not reflect the register patients use in clinical consultations.

5. Expert Validation

5.1. Expert Panel Composition and Process

A ten-member expert panel reviewed the completed module prototype using the 45-item Expert Validation Scale (EVS) over two rounds. The panel comprised five ESP specialists (three Pakistani university faculty and two international), two healthcare ESP specialists with medical education backgrounds, and three educational technology specialists including two MALL experts and one mobile application developer. Experts accessed the full module through a reviewer account and completed the EVS with mandatory qualitative justifications for all ratings of 1 (poor) or 2 (adequate). This mandatory justification requirement is identified as a critical feature of the validation design, inasmuch as it converts expert review from a rating exercise into a diagnostic process that specifies what requires revision and why. The EVS addresses five dimensions: Content Accuracy (nine items), Pedagogical Quality (ten items), Cultural Relevance (ten items), Usability and Technical Quality (nine items), and Assessment Alignment (seven items).

5.2. Round 1 Results and Revision Process

Round 1 EVS results showed that only Dimension A (Content Accuracy) marginally met the acceptable threshold ($M=3.12$, $CVI=0.78$), while four dimensions fell below: Pedagogical Quality ($M=2.88$, $CVI=0.72$), Cultural Relevance ($M=2.75$, $CVI=0.68$), Usability ($M=2.71$, $CVI=0.65$), and Assessment Alignment ($M=3.05$, $CVI=0.76$). Twenty-one of 45 items fell below the required CVI of 0.80. It is argued that this outcome is not a failure of the development process; it is the process functioning as designed, inasmuch as a two-round validation protocol producing uniformly high Round 1 scores would indicate insufficient critical scrutiny rather than genuine quality assurance.

The revision matrix documented 21 revisions, of which six were systemic. The most consequential revision addressed the AI practice prompt design: experts noted that overly structured prompts allowed students to read prepared scripts rather than produce spontaneous output, thereby undermining the Output Hypothesis rationale. Prompts were

redesigned to present scenario context only, requiring students to formulate their own response in real time. A second systemic revision addressed the Unit 3 scenario, which depicted an English-only interaction that experts identified as unrealistic for Pakistani clinical contexts where code-switching between English and Urdu is normative. The scenario was re-filmed with authentic code-switching, including an Urdu aside between the consultant and registrar, consistent with García and Wei's (2014) translanguaging framework. Furthermore, the bilingual glossary was revised from formal written Urdu to the colloquial register that patients actually use, following the expert finding that formal Urdu medical terminology misrepresents authentic patient communication.

5.3. Round 2 Results

Following the 21 documented revisions, all five dimensions exceeded both the mean threshold (3.0) and the CVI threshold (0.80) in Round 2. The overall module Round 2 mean is 3.67/4.00 with an overall CVI of 0.90, directly comparable to Rochmawati et al.'s (2024) benchmark of 3.7/4.00 for the E-SPAI nursing ESP application. The largest improvement was in Cultural Relevance, which moved from M=2.75, CVI=0.68 in Round 1 to M=3.68, CVI=0.91 in Round 2, reflecting the impact of the scenario re-filming and glossary revision. Assessment Alignment achieved the highest Round 2 CVI (0.94), indicating that the revised module's assessment design was the most consistently endorsed dimension across the expert panel. Similarly, the improvement across all dimensions confirms that the iterative two-round process was substantive rather than ceremonial: the revised module that entered Phase 3 was measurably superior to the initial design in all five quality dimensions.

6. Pilot Testing

A 16-week pilot with N=30 students at Shalimar Medical College (SMDC), Lahore an institution not included in Phase 1 or Phase 3, identified six issues requiring revision before full deployment. Completion rates ranged from 97% (Unit 0) to 76% (Unit 8), with the late-semester decline attributable to competing clinical placement demands rather than module quality, inasmuch as qualitative data from the two pilot focus group discussions confirm high satisfaction with all units including Unit 8. Satisfaction ratings ranged from M=3.7 to M=4.4 per unit. Pakistani scenario content was rated positively by n=28/30 pilot participants; offline functionality was endorsed by n=27/30, confirming that the equity-driven design decision to prioritise offline access was both valued and used.

Four revisions were implemented: AI loading speed was improved through additional compression and buffering, reducing latency from 8 to 12 seconds to 4 to 6 seconds on 3G connections; Unit 2 was split into sessions 2A and 2B to reduce cognitive load in the most linguistically demanding unit; a 90-second animated peer review onboarding tutorial was added; and Unit 6 documentation writing was expanded with templates and worked examples. The pilot focus group data independently confirm three

Phase 1 SSIG findings: private AI practice was credited specifically with reducing the anxiety that had previously prevented oral English practice; Pakistani clinical scenarios received enthusiastic endorsement as cognitively accessible in a way that contextually foreign scenarios are not; and peer review was initially resisted but subsequently endorsed by the majority of participants as the activity most revealing of their own communication patterns.

7. Discussion

The expert validation process demonstrates that expert consensus on a culturally contextualised healthcare ESP mobile module is achievable through a structured iterative design process, and that the CVI=0.90 achieved in Round 2 represents a genuinely high standard. The improvement from Round 1 (CVI=0.72) to Round 2 (CVI=0.90) indicates that the two-round process produced substantive rather than cosmetic change: expert feedback identified genuine weaknesses that the development team could not detect from inside their own work, and the revisions addressed those weaknesses in ways the expert panel could subsequently verify. On the other hand, a single-round validation process, which is common in the ESP module development literature, would have resulted in a module with CVI=0.72 entering experimental evaluation, with all the validity threats that entails.

The nine design commitments produced by the Phase 1-to-Phase 2 translation process are not theoretical postulates imported from the international MALL literature; they are specific responses to specific documented needs in the N=200 Phase 1 sample. This derivation logic is itself a methodological contribution: it demonstrates that the DBR evidence-to-design translation process can produce a set of non-negotiable commitments that are both empirically grounded and sufficiently specific to function as a quality standard against which the final module can be evaluated, which is precisely the function the EVS fulfils. Similarly, the pilot finding that students attribute their anxiety reduction specifically to the private practice feature, rather than to the module as a whole, provides preliminary qualitative support for the theoretical rationale of the privacy-as-therapeutic-affordance design commitment, which Phase 3 subsequently confirms through controlled experimental evidence.

8. Conclusions

This paper has reported the design, development, and expert validation of a MAELT module for Pakistani undergraduate healthcare students, constituting the first validated, culturally embedded, AI-assisted mobile English language teaching resource for this population. The module is grounded in a 200-student, five-instrument needs analysis; validated by ten specialists achieving CVI=0.90 through a two-round iterative process; and confirmed as feasible through 16-week pilot testing generating six targeted revisions. Consequently, the validated module constitutes a deployable educational resource whose

evidence architecture, empirically grounded design, expert validation, and pilot confirmation, provides the basis for controlled experimental evaluation. Furthermore, the DBR methodology demonstrated in this study provides a replicable template for ESP module development in comparable multilingual professional education contexts, whereby needs-analysis evidence is systematically translated into design commitments, expert validation identifies and remediates weaknesses before experimental deployment, and the resulting module carries multi-layer quality evidence rather than the developer's theoretical assurance alone.

References

- Brown, A. L. (1992). Design experiments: Theoretical and methodological challenges in creating complex interventions in classroom settings. *Journal of the Learning Sciences*, 2(2), 141-178.
- Design-Based Research Collective. (2003). Design-based research: An emerging paradigm for educational inquiry. *Educational Researcher*, 32(1), 5-8.
- García, O., & Wei, L. (2014). *Translanguaging: Language, bilingualism and education*. Palgrave Macmillan.
- Hutchinson, T., & Waters, A. (1987). *English for specific purposes: A learning-centred approach*. Cambridge University Press.
- Krashen, S. D. (1985). *The input hypothesis: Issues and implications*. Longman.
- Ladson-Billings, G. (1995). Toward a theory of culturally relevant pedagogy. *American Educational Research Journal*, 32(3), 465-491.
- Nomaan, S. (2022). Teaching English for healthcare communication in Pakistani medical colleges: Challenges and prospects. *Pakistan Journal of Applied Linguistics*, 3(1), 12-28.
- Rochmawati, E., Sudartini, S., & Wijaya, A. (2024). Development and validation of E-SPAI: A mobile application for nursing ESP at an Indonesian university. *Asian ESP Journal*, 20(1), 45-78.
- 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.
- Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science*, 12(2), 257-285.
- Thiagarajan, S., Semmel, D. S., & Semmel, M. I. (1974). *Instructional development for training teachers of exceptional children*. Indiana University.
- Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Harvard University Press.
- Zimmerman, B. J. (2000). Self-efficacy: An essential motive to learn. *Contemporary Educational Psychology*, 25(1), 82-91.