PREPRINT

Not Peer Reviewed

RESEARCH ARTICLE

Voice-Based Chatbots for English Speaking Practice in Multilingual Low-Resource Schools: A Multi-Stakeholder Mixed-Methods Study

[version 1]

Sneha Shashidhara[1], Lyle Ungar[2], Sharath Chandra Guntuku[2,3]

[1]Center for Social and Behaviour Change, Ashoka University, New Delhi, New Delhi, 110057, India
[2]Department of Computer and Information Science, University of Pennsylvania, Philadelphia, Pennsylvania, USA
[3]Leonard Davis Institute of Health Economics, Philadelphia, Pennsylvania, USA

v1 **First published:** 11 Jul 2025, **2**:179
https://doi.org/10.12688/verixiv.1351.1
Latest published: 11 Jul 2025, **2**:179
https://doi.org/10.12688/verixiv.1351.1

Abstract

This study investigates the deployment of a voice-based English-learning chatbot across four low-fee schools in Delhi, India, capturing perspectives from 23 students, six teachers, and five principals. Using an interpretivist multiple-case design, we combined real-time classroom observations, stakeholder interviews, and a six-day extended use period. Results reveal high motivation across all stakeholder groups and notable increases in student speaking confidence after 2–3 sessions. However, these gains were tempered by persistent technical barriers, namely Automatic Speech Recognition (ASR) errors (especially for proper nouns and Hindi loanwords), microphone usability issues, and network latency. Stakeholders converged on the need for more intelligible audio output (slower speech in an Indian accent), a simpler user interface, and actionable analytics; however, they differed on the preferred usage setting (classroom vs. home) and the extent of bilingual support. We offer design recommendations for voice-enabled chatbots in low-resource multilingual contexts, emphasising accent-matched speech output, one-tap interaction, and role-specific (teacher vs. student) dashboards.

Keywords

Education, Chatbot, English-speaking-skills, Low-resource schools, EdTech

This article is included in the Gates Foundation gateway.

Corresponding author: Sneha Shashidhara (sneha.shashidhara@ashoka.edu.in)

Author roles: Shashidhara S: Conceptualization, Data Curation, Formal Analysis, Investigation, Methodology, Project Administration, Supervision, Validation, Visualization, Writing – Original Draft Preparation, Writing – Review & Editing; **Ungar L**: Conceptualization, Data Curation, Funding Acquisition, Investigation, Methodology, Project Administration, Resources, Software, Supervision, Writing – Review & Editing; **Guntuku SC**: Conceptualization, Investigation, Methodology, Supervision

Competing interests: No competing interests were disclosed.

Grant information: The University of Pennsylvania provided partial funding through the Penn Global Research and Engagement Grant Program. Part of the work was supported by the Bill and Melinda Gates Foundation.
The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

How to cite this article: Shashidhara S, Ungar L and Guntuku SC. **Voice-Based Chatbots for English Speaking Practice in Multilingual Low-Resource Schools: A Multi-Stakeholder Mixed-Methods Study [version 1]** VeriXiv 2025, **2**:179 https://doi.org/10.12688/verixiv.1351.1

First published: 11 Jul 2025, **2**:179 https://doi.org/10.12688/verixiv.1351.1

1. Introduction

Conversational agents have become one of educational technology's most rapidly diffusing strands. A recent systematic review found that the research on educational chatbots has grown steadily, with language learning emerging as the single largest application domain (Hwang & Chang, 2023; Fu et al., 2024; Labadze et al., 2023; Laun & Wolff, 2025). Parallel meta-analyses similarly report robust but heterogeneous learning benefits (Okonkwo & Ade-Ibijola, 2021; Wang et al., 2024).

The lineage of today's chatbots is long. Early rule-based programs such as **ELIZA** (Weizenbaum, 1966) and **PARRY** (Colby, 1981) demonstrated the potential for computer-driven dialogue, highlighting the brittleness of purely rule-based interactions. A second generation of intelligent tutors, most notably **AutoTutor** (Graesser et al., 1999), combined limited natural language understanding (e.g. semantic parsing) with pedagogical dialogue to create a more interactive, mixed-initiative experience. Recent Large-language-model (LLM) based chatbots (e.g., ChatGPT) can generate context-sensitive responses with near-human fluency. Yet, their use in education has introduced new concerns about reliability, bias, and assessment integrity (Kasneci et al., 2023). Notably, scoping reviews note a surge in LLM-specific educational chatbot studies in the past few years (Fu et al., 2024).

Within educational applications, language learning remains the dominant testbed for chatbots. Conceptual reviews (Huang, Hew & Fryer, 2022; Jeon, Lee & Choe, 2023) and three independent meta-analyses (Zhang et al., 2023; Wang et al., 2024; Wu & Li, 2024) converge on a moderate mean effect ($g \approx 0.50$) for chatbot-assisted language learning. These effects are moderated by factors such as the type of language task and interface modality used. Voice-based agents merit special attention. A recent meta-synthesis of 57 voice-chatbot studies (Li et al., 2025) found that using voice can make interactions feel more authentic for learners. Still, it also introduces higher speech-recognition difficulties, especially for young and non-native speakers.

However, the existing research is geographically skewed toward high-resource settings. Large-scale deployments in low-resource contexts are rarely documented in academic literature. For instance, the **SwiftChat platform** reportedly serves over 100 million Indian learners across 10 languages, yet it has received little attention in peer-reviewed studies (Amazon Web Services, 2023).

From a theoretical standpoint, chatbots can be framed within established learning theories. In sociocultural terms, they function as tools in the learner's **Zone of Proximal Development** (ZPD) (Vygotsky, 1978) by offering contingent scaffolds (i.e., support tailored to the learner's current level). **Activity Theory** provides another lens, focusing on how the chatbot mediates language learning activity as part of a broader system (Cheng et al., 2024), where the chatbot serves as a mediator in the learning process. Chatbots operationalise ZPD by providing contingent scaffolding (Wood et al., 1976), while Activity Theory frames human-chatbot-task mediations (Engeström, 2016).

Triangulating the perspectives of administrators, educators, and learners is critical for developing scalable educational technology (edtech) in resource-constrained settings (Tondeur et al., 2017). Yet, few empirical studies triangulate multiple school-level stakeholders or track learners' voice interactions longitudinally under bandwidth constraints. To address these gaps, the present study engaged principals, teachers, and students in four low-fee private schools in Delhi, India, as they engaged with a spoken-English chatbot over a six-day period. Specifically, we asked the following research questions:

RQ1: How do students, teachers, and principals experience a voice-based English-learning chatbot in a low-resource, multilingual school environment?

RQ2: What technical and pedagogical factors enable or hinder student engagement with the chatbot?

RQ3: What design adaptations could optimise the chatbot's usability and educational value in such contexts?

The following section reviews the literature in greater depth.

1.1 Literature review

Conversational agents evolved from keyword-based chatbots like ELIZA and PARRY (Weizenbaum, 1966; Colby, 1981) to sophisticated mixed-initiative tutors exemplified by AutoTutor (Graesser et al., 1999). Modern LLM chatbots generate fluent dialogue but raise concerns over reliability, socio-cultural biases, and opaque reasoning (Kasneci et al., 2023).

Educational chatbots designed to support learning through tutoring, feedback, or language practice are showing promise. Systematic reviews and meta-analyses consistently report moderate learning gains and motivational benefits from chatbot use, tempered by heterogeneity and study design limitations (Okonkwo & Ade-Ibijola, 2021; Zhang et al., 2023; Wang et al., 2024; Wu & Li, 2024; Laun & Wolff, 2025). A randomised trial in Nigeria found that AI-powered tutoring using large language models significantly improved English learning outcomes among Secondary school students (De Simone et al., 2025). Another review focusing on English language speaking practice chatbots reports significant effectiveness on learner interaction, motivation, and oral language skills (Du & Daniel, 2024). However, review fatigue and brief interventions caution against over-generalisation (Fryer et al., 2017).

While empirical findings highlight promising trends, theoretical frameworks can offer deeper insight into how and why chatbots may support learning and the importance of sociocultural and cognitive paradigms. Theoretical perspectives emphasise chatbots as mediational tools within social and cultural frameworks (Vygotsky, 1978). Cognitive Load Theory (Sweller, 1988; Paas et al., 2003) highlights the importance of dialogue complexity and multimodal design in managing learner cognitive burden.

Sociolinguistic and affective factors shape learner engagement and accessibility. Accent and intelligibility critically affect learner motivation and comprehension; sociolinguistic research supports tailoring chatbot accents to culturally familiar varieties, such as Indian English, to enhance engagement (Derwing & Munro, 2005). Multimodal interfaces integrating voice, text, and visual elements (avatars, emojis) improve comprehension and emotional connection, especially for young or multilingual learners (Rigas & Alseid, 2008). In parallel, affective computing techniques, embedding empathetic tone and emotional feedback, sustain engagement and lower speaking anxiety (Picard, 1997; Arroyo et al., 2009; Zhai & Wibowo, 2022). Learner agency in dialogic interactions further supports scaffolded language development (Satar & Akcan, 2018).

However, persistent challenges remain. These challenges include technical frictions such as ASR errors and device constraints, pedagogical alignment including curricular fit and data overload, and ethical concerns spanning bias, plagiarism, and data privacy (Okonkwo & Ade-Ibijola, 2021; Kasneci et al., 2023; Williamson & Eynon, 2020; Zhang et al., 2022). Data security and informed consent are crucial in low-resource settings with weak regulatory environments. ASR systems often underperform for non-native accents and multilingual code-switching (Sullivan et al., 2022; Chi & Bell, 2022), exacerbating equity gaps.

Mobile-optimised learning promises to leverage widespread smartphone availability to support grammar and language acquisition (Li & Hegelheimer, 2013). Yet global reach demands sensitivity to local linguistic diversity and infrastructural realities (Godwin-Jones, 2014).

In summary, prior work has largely not explored three critical aspects that our study targets: a) longitudinal use of voice-based chatbot interactions (most studies are short-term), b) inclusion of multiple school stakeholders (teachers, students, and principals), and c) settings that are both multilingual and bandwidth constrained. We address these gaps by conducting a mixed-methods study in four affordable private schools in Delhi, India, where both Hindi and English are the media of instruction. We deployed a prototype spoken-English chatbot in these schools and collected rich qualitative and quantitative data over six days. During this period, students interacted with the chatbot in and outside class, and we gathered detailed feedback from students, their English teachers, and their principals about the chatbot's usability, design, and contextual appropriateness. This approach allowed us to identify key technical and pedagogical challenges in situ and derive design recommendations, making voice-based educational chatbots more effective in low-resource, multilingual settings, specifically to support spoken English.

2. Methods

2.1 Study design and philosophical stance

We adopted an **interpretivist, multiple-case qualitative design** (Kamal, 2006) to generate a granular understanding of how middle-school students, their teachers, and principals experience a spoken English practice chatbot in a real-world school setting. The study combined *two sequential field phases*:

1. **Real-time observation with immediate feedback** (Day 1): Students were introduced to the chatbot and engaged in an initial supervised session, followed by an immediate feedback discussion.

2. **Extended use with delayed feedback** (Days 2–6): Students continued to use the chatbot daily for five more days, with periodic observations, and we collected their input at the end of this period.

The two phases allowed us to contrast first-impression use against longer-term engagement, capturing more reflective participant feedback.

About the chatbot

The prototype English-speaking chatbot (nicknamed 'ChatFriend') was designed to help students practice conversational English. Its dialogue content drew from familiar contexts from the students' lives and school materials, such as talking about their best friend, favourite sport or ordering food in a restaurant. The chatbot personalised its prompts to each learner's profile, specifically age, interests, vocabulary, native language, and topics from their class curriculum. All conversations with the bot were conducted via voice in English, but the interface also displayed an English transcript of the conversation in real time. If students got confused, they could view transcriptions and ask clarifying questions in local languages like Hindi. The bot would attempt to understand and respond (either by translating or providing clarification in English). This bilingual support was included to make the chatbot more accessible in our multilingual context.

Chatfriend is a web-based chatbot app built with React and hosted as a static site on AWS S3, delivered globally via CloudFront for fast performance. Our Django backend runs on Elastic Beanstalk and automatically scales up or down based on demand. We use Redis for fast data caching and WebSocket messaging, MySQL (RDS) for structured data, and S3 for storing uploaded files. When a user types or speaks, the input is sent over a WebSocket connection: text goes directly to GPT-4o-mini for a streaming response, and audio is transcribed live using Whisper-1. Replies stream back in real-time, and users can read them, listen via Google Text-to-Speech, or both. We also use OpenAI's Moderation API to automatically detect and block unsafe content like hate speech or self-harm.

2.2 Study setting and participants

We conducted the study in four low-fee (affordable) private schools located in the Badarpur and Shakti Vihar neighbourhoods of Delhi, India. We also included a fifth school for a limited perspective (only the principal of that school was interviewed to broaden the administrative insights). We used purposive maximum-variation sampling to select schools that differed in management type and digital infrastructure to ensure diverse contexts. All schools served middle-income communities and had both Hindi and English as languages of instruction.

Participants: We engaged three groups of stakeholders at these schools, with the following composition and roles (Table 1).

To strengthen the *credibility* of our findings, we triangulated data across these three stakeholder groups and across the two phases of observation described above. We also held peer debriefing sessions within the research team throughout the study to cross-check interpretations and emerging findings.

2.3 Sampling and recruitment

We began by securing institutional buy-in. Each school's principal was briefed about the study (via an in-person meeting, followed by email/phone communication) and formally agreed to participate. After this, English teachers sent opt-in consent forms to all Grade 7 and 8 students. From the pool of students who returned signed parental consent, we randomly selected our 23 student participants. (Due to the tight timeline, this random selection was not stratified by grade or gender; however, our final sample still included a mix of boys and girls and representation from both grades.) The participating teachers and the principals at each school were recruited after their schools joined the study (we invited the Grade 6–8 English teachers and the head principal at each site to take part, and all agreed). No students, teachers, or principals who consented to the study dropped out before it was completed. Table 2 describes the overall data collection exercise.

Table 1. Describes the different stakeholders engaged in the study.

Stakeholder	n	Inclusion criteria	Role in study
Students	23 (13 female, 10 male), 17 were in Grade 8, and 6 were in Grade 7	Regularly enrolled, parental consent, no prior chatbot exposure	Primary users of the bot
English teachers	6 (5 female, 1 male)	Taught Grades 6-8 in participating schools	Classroom observers, interviewees, and potential facilitators
Principals	5 (3 female, 2 male)	Head of participating schools	Key informants on feasibility and policy fit

Table 2. Describes the overall data collection and data management exercise.

Data source	Instrument & logistics	Timing/frequency	Artefacts generated
Student–bot interactions	Custom **observer checklist** (15 closed + open items) completed in situ by trained moderators; screen-and-audio capture	23 students across 5 days (≈4–5 sessions per student)	Structured field spreadsheet + audio files
Immediate student feedback	Short guided discussion (5–8 min) post Day 1 session	Once	Audio recordings, moderator memos
Extended-use reflections	Circle-time discussion (12–15 min) on Day 6	Once	Audio, moderator memos
Teacher interviews	Semi-structured guide (≈20 questions)	30–40 min each, face-to-face	Audio, verbatim transcripts
Principal interviews	Semi-structured guide (≈25 questions)	35–45 min each, face-to-face	Audio, verbatim transcripts

2.4 Data-collection instruments and procedures

Phase 1 – Real-Time Observation with Immediate Feedback (Day 1): On the first day at each school, we conducted two back-to-back sessions with the students:

- **Initial chatbot session (15–20 minutes, supervised):** Moderators first briefly demonstrated the chatbot's features and usage. Each student then used the chatbot on a school-provided smartphone or tablet for 15–20 minutes while a moderator closely observed. We focused our observations on how students navigated the interface (e.g., any confusion with buttons or features), their engagement level (interest or frustration), and any difficulties they encountered, such as trouble understanding the chatbot, mishearings by the ASR or other technical glitches. Moderators took handwritten or checklist-based notes in real-time during this period.

- **Immediate feedback discussion (5–10 minutes):** We facilitated a short group discussion with the students after the usage session. We asked open-ended questions about their first impressions: what they liked or disliked, any problems they faced, and suggestions for improvement. This guided debrief allowed students to reflect on the experience while it was fresh and for us to capture their initial reactions in their own words.

Additionally, on Day 1, we took the opportunity to interview the English teacher(s) who observed the session and the school principal. The teacher interview (approximately 30–40 minutes, semi-structured) focused on the teacher's perspective of the chatbot's classroom integration, student engagement, and any support or training needs. The principal interview (around 35–45 minutes) explored higher-level issues like the chatbot's fit with school priorities, infrastructure concerns, and perceived value for student learning. The teacher and principal interviews were conducted one-on-one in a quiet room on campus.

All student discussions and the teacher/principal interviews on Day 1 were audio-recorded (with consent). These recordings and the moderators' observation notes were later transcribed to ensure we accurately captured details and quotes for analysis.

Phase 2 – Extended Use with Delayed Feedback (Days 2–6): In the subsequent five days, students continued to use the chatbot in short daily sessions at school. Each day, a moderator was on-site to ensure the sessions ran smoothly, to encourage students to participate (reminding them to use the bot for the allotted time), and to observe their interactions quietly. Each student typically spent about 10–15 minutes per day with the chatbot during this phase. Moderators noted any notable interaction patterns over time, for example, if a student became more confident or conversed longer as days went by or if specific technical issues (like ASR errors or slow responses) persisted or resolved. Students were also encouraged to ask questions or report any problems to the moderator after each session, which the moderator documented. This extended use period gave students time to become more familiar with the chatbot and allowed them to generate more considered feedback beyond the novelty of the first day.

We asked the English teachers to observe at least one of these mid-week sessions if their schedule allowed. In practice, most teachers could only briefly drop in or could not participate during Phase 2 due to their class duties. As a result, the moderator was the primary observer throughout these days.

At the end of Day 6, we conducted a longer "delayed feedback" focus group with the students (lasting ~15 minutes in a circle-time format). In this discussion, students reflected on their cumulative experience: what improvements they noticed in themselves, which features or challenges persisted, and any new suggestions after using the chatbot multiple times. This delayed feedback session was also audio-recorded and later transcribed. By comparing the insights from Day 1 and Day 6 discussions, we could see how perceptions evolved over the week of usage.

Training of moderators

All field moderators (the researchers who facilitated the sessions) underwent comprehensive training before data collection began. Our training program had two main components: (1) Observation training – learning how to systematically observe and note student–chatbot interactions; and (2) Interview facilitation training – learning how to conduct productive feedback discussions and interviews with students, teachers, and principals.

- *Observation training:* We prepared the moderators to capture detailed, structured notes during student chatbot sessions. They were provided with an observation checklist highlighting specific things to watch for: e.g., instances of miscommunication between student and bot (such as the bot misunderstanding a student's speech), particular English words or sentences students struggled with, any system errors or crashes, whether students were speaking or resorting to reading the on-screen text, and when/if students switched from English to Hindi during the interaction. During training, we ran a live demo of a student using the chatbot while the moderators practised taking notes. Afterwards, we collectively reviewed their notes, discussing what they noticed, what they might have missed, and how to format them clearly. This exercise taught moderators to focus on key behaviours/events and remain open to unexpected observations (beyond the checklist). The research team gave feedback and tips, which helped calibrate the observers on what to look for and how to record it consistently.

- *Discussion/interview training:* Moderators were also trained to facilitate semi-structured interviews and group discussions. We provided them with a discussion guide (a list of questions and follow-ups for students, teachers, and principals) and went through each question, explaining its intent. Moderators practised techniques for probing deeper into responses (without leading the participant) and creating a comfortable, non-judgmental atmosphere so that participants—especially students—would speak freely. The training included mock sessions: team members role-played as students, teachers, or principals while a moderator practised asking questions. After each role-play, the group reflected on what went well and what could be improved (for instance, rephrasing a confusing question or giving a hesitant student more time to answer). These mock interviews and constructive feedback from the research led to improved moderators' confidence and skill in guiding real conversations during the study.

During the field deployment, we also held regular debriefing meetings with the moderators (daily in the first few days, then a bit less frequently). In these debriefs, moderators could share any difficulties they encountered, and we collaboratively identified solutions or adjustments. For example, if a moderator noted that students seemed shy in the group discussion, we might adjust our approach the next day (such as using an ice-breaker question). This ongoing support ensured that any challenges in data collection were caught and addressed early and that our team remained aligned with data collection techniques throughout the study.

2.5 Data management

All audio recordings (from student discussions and interviews) were transcribed verbatim. Any Hindi or code-switched speech in the recordings was translated into English during transcription, so the analysis could be done primarily in English. We removed personal identifiers from the transcripts and observation notes to protect participant anonymity (e.g., replacing student names with codes). All cleaned transcripts, observation checklists, and related data files were stored securely on an encrypted institutional drive accessible only to the research team. We assigned each participant a unique pseudonymous ID (for example, "4-4-1S-4" for a particular student) to label their data. This coding scheme allowed us to track each individual's inputs across multiple days and data sources—facilitating longitudinal analysis—without using their real names or any identifiable information.

2.6 Analytic approach

We followed a multi-step qualitative analysis process involving three members of the research team working together:

1. **Familiarisation** – The research team read all transcripts and field notes twice and wrote margin memos.

2. **Open coding** – Two analysts independently applied inductive, line-by-line codes and the intercoder agreement after codebook reconciliation was above 90%.

3. **Axial & temporal coding** – Codes were clustered into categories (e.g., *mic-UX*, *ASR mishearing*, *confidence shift*). For students with ≥3 sessions, matrices plotted codes against session numbers to visualise the trajectory.

4. **Theme construction & triangulation** – Themes were refined through constant comparison across data sources and stakeholder groups, and negative cases were checked against.

2.7 Ethical considerations

This study was approved by the Institutional Ethics Committee of Ashoka University (Approval No. 24-F-10019-Sharma). We obtained informed consent at multiple levels: each school's administration provided institutional consent, each participating teacher and principal gave individual written consent, and all student participants gave written assent with their parent's written consent obtained via the school. As we approached the school, they gave us permission and were responsible for providing parents with the information regarding this school-sanctioned exercise and obtaining written consent. We provided a written document explaining the exercise and its intent to the schools so they could use it to get permission from parents. We emphasised to all participants that participation was voluntary and that they could skip any interview question or withdraw from the study at any time without any penalty. To ensure no undue influence on student academics, we clarified that the chatbot usage and feedback did not affect grades or class standing. As a token of appreciation, we gave each school a set of books for their library after the study; however, we did not provide any individual compensation, nor were any incentives tied to individual performance. Throughout the project, we maintained strict confidentiality of participant data. All transcripts and records were anonymised, and we used ID codes instead of real names. The audio recordings (and their transcripts) were encrypted, and only the core research team could access them. Participants (and for minors, their parents) had the option to decline being audio-recorded if they wished (though, in practice, all consented). We will use the data solely for research purposes and publication, and this was communicated to participants. Any potentially identifying details have been altered in our reporting to protect privacy.

ChatGPT model O3 was used for language improvement.

3. Results

3.1 School-leader perspectives (principals, *n* = 5)

3.1.1 Digital-infrastructure constraints

Across the five schools, only three (60%) possessed even a handful of functioning desktop or laptop computers, and just one permitted students to go online during class time. Principals consequently framed **mobile phone compatibility as the sole scalable path** for any technology intervention.

> *"Without phones, the plan dies—our lab has two working PCs and forty students."* —Principal P2

> *"Connectivity drops at least twice a period; any tool must survive offline for a bit."* —Principal P4

3.1.2 Pedagogical alignment and staff workload

Four out of five principals (80%) noted that their **teachers could only spare about "30 minutes, three times a week" for any new activity**—anything beyond that would overwhelm staff capacity unless additional personnel were hired. Consequently, the principals argued that the chatbot should be used in short, structured sessions rather than as an open-ended daily practice. In their view, it would be more feasible to integrate brief, scheduled chatbot exercises a few times a week (e.g., during existing English periods) than to expect teachers to supervise lengthy or frequent chatbot use on top of their regular teaching duties. This reflects an apparent concern for teacher workload: to gain buy-in, the tool must complement the curriculum in manageable doses instead of adding substantial new burdens.

3.1.3 Accent, speech rate, and scaffolding concerns

A majority (3/5) predicted that **30–50% of learners would struggle** unless the bot spoke more slowly and with an Indian accent. Principals linked intelligibility directly to equity, noting that weaker students "switch off the moment they cannot parse the audio."

> *"The speed should be slower; an Indian accent should be used."* —Principal P1

3.1.4 Preferred deployment model

Three schools (60%) wished to **launch the chatbot inside the classroom for orientation, then migrate use to homework** once routines solidified. The prospect of visible learning gains tempered concerns about unsupervised phone use at home.

Three principals expected parents to appreciate the chatbot, while two anticipated concerns around unsupervised internet access that may fade with visible learning gains. For at-home practice, parental support was generally seen as limited, with two principals reporting a lack of digital familiarity or a lack of time as constraints to parental engagement.

3.1.5 Appetite for formative assessment data

All principals (5/5) valued the prospect of **transcripts or dashboards** that would surface individual learner progress for teachers and parents.

3.2 English-teacher perspectives (teachers, *n* = 6)

3.2.1 Learner proficiency and the classroom timetable

Teachers estimated their pupils' spoken English competence at roughly 3/5, citing hesitation, lack of confidence, tense errors, and restricted vocabulary as dominant barriers. They also noted that students often lack exposure and opportunities to converse or practice English at home. Each English period lasts 35–45 minutes, leaving little space for supplementary technology unless it **replaces, rather than adds to, existing activities**.

3.2.2 Technology readiness and training needs

Half the schools already run "smart-class" video sessions, yet none employ digital assessment. Four of the six teachers (67%) have more than six years of experience and regard **curriculum alignment and hands-on training** as prerequisites for adoption.

3.2.3 Perceived affordances of the chatbot

All six teachers endorsed the bot's capacity to provide **"non-judgmental" practice space** for shy students. Post-session satisfaction scores averaged 4.0–4.4 on a five-point scale for task design, conversational flow, and overall format. One of the teachers reported the bot's ability to register Hindi as English text as a helpful feature for learning correct English expressions. However, teachers catalogued four persistent pain points:

1. Microphone/"hold-to-talk" confusion.
2. Automatic speech recognition (ASR) mishearings, especially of proper nouns.
3. Long or multi-clause bot prompts that overwhelm learners.
4. Speech delivered "too fast, with an unfamiliar accent."

> *"Struggled with the mic button; shorter sentences would help hesitant users."* —Teacher T4

3.2.4 Implementation preferences

When asked how the tool should be scheduled, three distinct camps emerged: homework use (50%), a dedicated in-class period (33%), and hybrid models (17%). Differences hinged on device availability in class or potential classroom disturbances versus fears of excessive phone use at home.

> *"Not as homework; students might become too dependent on phones."* —Teacher T5

3.2.5 Current evaluation methods & Value of learner-level analytics

Teachers employ several informal methods to assess students' spoken English, most commonly through classroom conversations, presentations, and reading activities. They also noted that students' fear of making mistakes and being

judged negatively often affects their willingness to speak. Every teacher (6/6) anticipated instructional benefit from reviewing transcripts and concise dashboards. They cautioned that data **must be filterable to highlight growth or persistent errors** and not to burden already crowded preparation time.

3.2.6 Utility for formal assessments and content integration

While 3/6 teachers expressed openness to using the chatbot as an assessment tool, concerns were raised regarding its alignment with curricula and the lack of digital infrastructure. Regarding uploading learning/practice content, most teachers indicated ease in sourcing and uploading digital content to the chatbot from online platforms and traditional resources such as newspapers, magazines, and storybooks. Some emphasised the need for a brief orientation to do this.

3.2.7 Views on sustaining student engagement

To counteract students' decreasing attention spans and boredom, teachers suggested incorporating videos, visuals, cartoons, games, jokes, and interactive scoring systems to help maintain their interest.

3.3 Interim synthesis

Both principals and teachers highlighted that students struggle with speaking fluency, often hesitating due to a lack of confidence, limited vocabulary, shyness, and minimal exposure to English outside the classroom.

Taken together, principals and teachers **converge on three critical adoption conditions**: (a) intelligible audio (accent + pace), (b) ultra-low-friction voice UX, and (c) analytics that illuminate—not complicate—formative assessment. At the same time, they diverge on *where* practice should occur: principals favour an initial, classroom-anchored pilot, whereas half the teachers prefer homework deployment to protect lesson time.

While both groups noted general parental acceptance, they expressed concerns about limited engagement due to time constraints or low digital familiarity.

These convergences and tensions are revisited in Section 3.5 after introducing student data.

3.4 Learner perspectives (students, *n = 23*)

3.4.1 Student background and general profile

The student sample comprised 23 learners from Grades 7 and 8 across four schools. At baseline, roughly one-fifth described themselves as comfortable speaking English, while the majority reported hesitation linked to vocabulary limitations or fear of errors. Opportunities for spoken English are confined mainly to school settings, with limited or no practice at home. Many emphasised English's importance for future education, career opportunities, and travel. Most learners primarily accessed digital content via low-end smartphones, with varying degrees of familiarity with app login procedures and microphone controls (See Table 3 for a full description).

These background data reinforce two points already noted in later subsections: (i) learners' **strong instrumental motivation** for English and (ii) the **critical role of interface simplicity** given varied digital literacy levels.

3.4.2 Motivation and baseline affect

Many learners expressed a strong instrumental motivation for improving spoken English, frequently linking proficiency to career and travel aspirations. While some (approximately 20–50%) self-identified as relatively confident, a majority conveyed anxiety and nervousness about speaking, citing fear of mistakes and limited vocabulary. These findings emerge primarily from qualitative interview data, and therefore, percentage estimates should be interpreted cautiously rather than as precise prevalence measures.

> *"I hesitate when speaking English because I fear getting it wrong."* —Grade-7 student (ID S-7-12)

3.4.3 Digital access context

Most students rely on a **parent-owned Android phone** at home; only a minority possess laptops or stable Wi-Fi. Consequently, 65% say they would prefer to practise at home, 25% in class, and 10% are flexible.

> *"I'd like to use it at home in my free time on my dad's phone in the evening."* —Grade-8 student (ID S-8-04)

Table 3. The table describes the background data collected from students through the qualitative interviews.

Thematic area	Main takeaways
Baseline comfort in spoken English	Roughly **22%** of learners described themselves as comfortable or "enjoying" English conversation, whereas ≈**78%** reported difficulty linked to vocabulary gaps, struggle with long sentences or fear of mistakes.
Perceived importance of English	Virtually all respondents (≈ 95%) framed English as essential for communication, job interviews, or future mobility; **60%** explicitly linked it to international travel.
Career aspirations	Students most often aspired to medical careers (**35%**), teaching (**15%**), or business; each group stressed English as a prerequisite for success.
Favourite and challenging school subjects	Mathematics (8 mentions), Science (6), and English (6) topped the favourite list; Social Studies (≈ 38%) and Mathematics (≈ 21%) were most often labelled difficult.
Exposure to English-language media	Only a minority regularly consumed English media; most preferred dubbed Hindi cartoons or regional content, citing comprehension difficulty with English shows.
Digital and login experience	About **60%** had prior experience logging into apps such as Instagram or ChatGPT; 20% had never logged in anywhere and expected difficulty.

3.4.4 First-encounter experience

Observations of over 100 student-chatbot interactions show an overwhelmingly **cautious first use**: short, hesitant replies dominate, with 78% of interactions noting at least one ASR misrecognition, proper nouns (friends' names, games, sports figures) being most error-prone. Ratings captured immediately after Session 1 are nevertheless positive (Table 4).

3.4.5 Technical friction and coping strategies

Learners encountered several recurrent technical challenges, which moderators consistently recorded across sessions:

- **Microphone/button confusion:** About 35% of learners required prompts to operate the mic button correctly, often requesting a larger, single-tap interface to replace the hold-to-talk mechanic.
- **Voice-recognition errors:** Nearly all students experienced misrecognition by the chatbot's ASR system, particularly for proper nouns and Hindi loanwords, which impeded smooth dialogue.
- **Latency and system delays:** Approximately 48% faced significant network latency or system errors, including "WebSocket not open" messages, sometimes causing premature session termination.

To compensate, many learners relied on reading the on-screen transcript before replying and frequently code-switched into Hindi when confronted with unfamiliar English vocabulary, leveraging the bot's bilingual support. While the chatbot attempts to process and translate Hindi and Hinglish inputs, our observations revealed inconsistent performance. Several instances were documented where the bot misinterpreted or blocked Hindi phrases, resulting in conversational breakdowns and learner frustration. These interruptions hindered smooth bilingual communication and sometimes negatively impacted learner confidence and engagement. Enhancing the chatbot's multilingual robustness remains a critical area for improvement to fully support users' linguistic diversity in these settings.

Table 4. Post-session-1 star ratings (n = 23).

Dimension			
Tasks	10 (43%)	9 (39%)	4 (17%)
Conversation flow	11 (48%)	7 (30%)	5 (22%)
Interface	10 (43%)	9 (39%)	4 (17%)

Table 5. The table describes the progression of the chatbot interactions through 4 example students.

Learner (pseudonym)	Day-1 affect	Day-3	Day-5	Persistent friction	Net arc
(ID 4-4-1S-4)	Nervous, very brief answers	"More comfortable"	"Felt confident today"	Mishearing of names, fast speech	↑ steady gain
(ID 4-1-1S-4)	Reads transcript, pace fast	"Replying confidently"	Asks meta-questions but is irritated by lag	Accent still hard	↗ then frustration
(ID 4-3-1S-4)	Bored	Enjoys Hindi→English translation	"Spoke freely"	Mic faults	↑ bloom
(ID 4-3-2S-4)	Nervous, did not ask any questions	"Responding with excitement"	"I wanted to end the conversation quickly because of misinterpretations"	Response delay, bot freezes, misinterpretations	(Excitement) ↗ then frustration

3.4.6 Confidence trajectory over repeated use

Longitudinal observation of 14 students completing three or more sessions revealed a clear pattern of growing confidence: the proportion of turns involving student-initiated questions increased from near zero on first use to approximately 25% by the third session. It is important to note that this metric may underestimate true learner initiative due to coding constraints relying on explicit rubric ticks.

By Day 5, moderator notes indicated that around 40% of learners predominantly spoke English, 20% primarily used Hindi, and 40% code-switched fluidly. These patterns suggest an evolving linguistic negotiation as learners gain comfort, though persistent technical friction sometimes disrupts confidence gains.

Although many learners demonstrated increased confidence and proactive engagement over repeated chatbot interactions, this progression was not uniform. Several students experienced setbacks characterised by nervousness, disengagement, or reliance on moderator assistance. Such fluctuations are often correlated with technical frictions—namely, speech recognition errors, delays in response, and usability challenges with the microphone interface. These findings underscore the fragile nature of confidence development and highlight the need for consistently low-friction, supportive user experiences to sustain learner motivation over time (Table 5).

3.4.7 Feature desires and improvement requests

Learners' wish-list aligns with teacher observations but adds a richer engagement layer:

- Wider topic breadth (*food, gaming, careers, science, mathematics*).
- **One-tap microphone** activation and a larger, more prominent mic icon.
- An optional **Hindi translation toggle** to manage "Hinglish" episodes.
- Visuals, emojis or avatars for a richer effect.
- Voice choice and adjustable speed control.

3.4.8 Preferred setting for use

Stakeholder preferences diverged on the deployment setting. Students strongly preferred home use, valuing privacy and reduced peer scrutiny. Principals predominantly advocated for classroom initiation to establish routines and provide supervision. Teachers were split: half preferred homework for extended practice, a third favoured dedicated in-class periods to ensure instructional alignment, and the remainder suggested hybrid models. These contrasts underscore the need for flexible deployment strategies accommodating varied contexts and concerns.

3.4.9 Affective responses to interacting with the chatbot

Student reflections reveal a complex emotional landscape when engaging with the chatbot. Several students described initial discomfort or feeling "weird" talking to a non-human interlocutor, especially during early sessions:

> *"At first, it felt strange talking to a robot instead of a person."* —Grade 8 student (ID S-8-06)

Others reported increased comfort over time, appreciating the bot's patience and non-judgmental nature, which helped reduce speaking anxiety:

> *"I was nervous, but the bot did not make me feel bad for mistakes."* —Grade 7 student (ID S-7-14)

Despite overall positive trends, a notable subset of students exhibited persistent hesitation, nervousness, or disengagement during interactions. Observer notes frequently captured signs of low enthusiasm, including short replies, avoidance of speaking, and requests for moderator assistance. These affective challenges often correlated with technical difficulties such as microphone operation and speech recognition errors, suggesting that while many students grew more comfortable, the experience was uneven and warrants tailored support.

3.4.10 Reported language learning gains

Several learners noted improvements in their English vocabulary and speaking confidence attributed to chatbot interaction. Examples include remembering new words or phrases encountered during dialogue and increased ease in forming sentences:

> *"I learned new words like 'interesting' and 'favourite' by using the bot."* —Grade 8 student (ID S-8-02)

> *"Talking with the bot made me more confident to speak in class."* —Grade 7 student (ID S-7-10)

These self-reports complement observation notes documenting longer, more fluent utterances over repeated sessions. While no formal pre/post testing was conducted, such subjective gains indicate potential learning benefits worthy of future empirical validation.

3.4.11 Synthesis of learner data

The findings from students underscore a **high latent motivation that is unlocked by a non-judgmental interlocutor** but constrained by three technical ceilings: ASR accuracy, mic ergonomics, and network latency. Confidence accrues quickly—typically after two or three low-friction encounters—but **regresses just as quickly when friction resurfaces**. Bilingual scaffolding and concise, single-clause prompts emerge as powerful design levers for sustaining engagement.

3.5 Cross-stakeholder integration

This final results subsection juxtaposes the perspectives of principals, teachers, and learners to identify *where they agree, where they differ,* and what those patterns imply for subsequent design or evaluation.

3.5.1 Areas of convergence

This exercise distils the common ground across stakeholder viewpoints, underscoring the principal points of agreement that can guide and streamline chatbot creation and roll-out in multilingual, resource-constrained educational environments. Table 6 describes the areas of convergence.

3.5.2 Areas of divergence

This section synthesises contrasting perspectives among stakeholders, highlighting key areas of divergence that pose challenges for chatbot design and deployment in multilingual, low-resource educational settings (Table 7).

3.5.3 Integrative summary

Taken together, the data portray a **psychologically welcome but technically fragile** tool. Stakeholders have an appetite for a slower, Indian-accented audio channel and low-friction voice input. They also agree that concise learner analytics

Table 6. **The table describes the areas of convergence across the three stakeholders.**

Converging theme	Empirical footing	Interpretive note
Audio intelligibility is pivotal. All stakeholder groups identified fast speech and a non-Indian accent as significant barriers to comprehension: principals estimated 30–50% of learners would struggle without slower, Indian-accented speech; teachers and students confirmed these challenges during interaction sessions.	Observation notes documented frequent ASR misrecognition and interview summaries highlighted accent and pace as sources of confusion.	This convergence underscores that audio clarity is a fundamental prerequisite for effective chatbot adoption, especially in linguistically diverse, low-resource classrooms where learners vary widely in exposure to standard accents.
A "no-judgment" space motivates learners. Teachers described the chatbot as providing a psychologically safe environment that encourages hesitant or shy students; learners repeatedly emphasised the absence of peer pressure and ridicule during chatbot interaction.	Multiple teachers and students emphasised reduced anxiety, with quotes describing the chatbot as a "friend" or "safe space."	This shared valuation suggests that the chatbot's social affordances are as critical as technical functionality, confirming theoretical expectations from sociocultural learning frameworks that low-affective-filter contexts promote language practice.
Technical friction suppresses engagement. Persistent microphone usability issues, ASR mishearings, and latency errors were widely reported across observation data and stakeholder interviews.	Detailed logs noted mic/button confusion in approximately 65% of learners, ASR errors affecting nearly all, and network-related delays in almost half of all the sessions.	The ubiquity of these technical barriers highlights the need for robust, fault-tolerant interaction design before pedagogical or content expansion can be effective. Without resolving these foundational issues, learners' confidence gains remain vulnerable, and adoption risks stagnation.
Concise, actionable analytics are universally attractive. Both principals and teachers expressed strong interest in formative assessment data via learner transcripts or dashboards but stressed the importance of streamlined, exception-based reporting.	All principals favoured detailed learner transcripts; teachers emphasised usability, preferring alerts highlighting key learner difficulties rather than voluminous logs.	This alignment indicates that analytics tools must balance transparency with manageability, tailoring data presentation to different user roles to support instructional decision-making without overwhelming educators.
Visual appeal and an attractive interface increase engagement. Both students and teachers strongly prefer visually engaging designs to support comprehension and make interactions more playful and enjoyable.	Students suggested emojis, visuals, and avatars for richer effects. Teachers suggested adding interactive elements like videos, cartoons, games, jokes, and a scoring system for sustained interest.	Preference for richer audio-visual features aligns with evidence that **multimodal engagement lowers the cognitive load and boosts motivation.** It shows that, beyond accuracy, **emotional and sensory appeal** is valuable to users.

Table 7. The table describes the divergent opinions across the three stakeholders.

Divergent issue	Stakeholder positions	Practical consequence
Venue for practice	Learners predominantly preferred home-based, private use (~65%), whereas teachers were divided between homework (50%) and dedicated class periods (33%). Principals generally favoured classroom-led onboarding before permitting homework use.	A staged roll-out model with classroom orientation followed by optional home practice may accommodate these differences.
Device assumptions	Students rely chiefly on low-end Android phones; some principals still envision computer-lab setups.	Prioritising mobile-first design and offline capabilities is essential.
Bilingual scaffolding	Learners commonly code-switched to Hindi, while some teachers expressed concern that excessive Hindi use might dilute English immersion.	Configurable bilingual support, toggled per lesson or learner proficiency, could balance these competing priorities.

could raise instructional value. Yet *where* and *how* the chatbot should sit in the school–home ecology and how much bilingual support or data detail is advisable remain contested.

These convergences and divergences underpin the design implications articulated in the Discussion (Section 4) and shape the forthcoming Recommendations (Section 6).

4. Discussion

4.1 Summary of findings

This study explored the experiences of principals, teachers, and students interacting with a spoken-English chatbot across four low-fee Delhi schools. The chatbot was regarded as a psychologically safe and motivating tool, particularly valued for its non-judgmental environment and the potential to deliver formative learner analytics. Notably, roughly half the students demonstrated measurable confidence gains within two to three low-friction sessions, as evidenced by increased rates of student-initiated questions. However, these gains were fragile, with engagement noticeably impeded by persistent technical barriers—most prominently, automatic speech recognition (ASR) errors, microphone usability challenges, and network latency. Stakeholders broadly concurred on the need for slower, Indian-accented audio and simplified voice interaction while diverging on preferences regarding the practice venue (classroom versus home) and the granularity of analytic feedback.

4.2 Relation to prior research

The affective and motivational outcomes reported here echo findings from recent systematic reviews and meta-analyses. For instance, Okonkwo and Ade-Ibijola (2021) and Huang, Hew, and Fryer (2022) identify chatbot use as conducive to learner motivation and anxiety reduction. The emphasis placed by stakeholders on a "safe, non-judgmental space" aligns with sociocultural learning theories (Vygotsky, 1978; Guo et al., 2024) and is corroborated by empirical studies highlighting lowered affective filters in computer-mediated language learning environments (Satar & Akcan, 2018). Our study extends activity theory by illustrating how the chatbot mediates language learning activities in a low-resource setting but also introduces new contradictions (e.g., between the tool and technical infrastructure) that must be resolved for effective implementation.

Beyond motivation, students' affective responses to chatbot interaction were mixed. Some initially found talking to a machine 'weird' or awkward, a phenomenon documented in human-computer interaction literature (Brave et al., 2005). However, the chatbot's patient, non-judgmental dialogue was widely credited with reducing anxiety, echoing findings by Zhai and Wibowo (2022) on empathetic chatbot design. This suggests gradual familiarisation and affect-aware interface design are critical to sustaining engagement.

The salience of ASR accuracy and accent intelligibility resonates strongly with findings from Han and Lee (2024) and Jeon, Lee, and Choe (2023), who underscore these factors as major bottlenecks in voice-based chatbot implementations, particularly for non-native speakers. This study extends these findings by demonstrating the amplification of such challenges in bandwidth-constrained, multilingual school settings. The call for accent-matched speech engines further reflects the growing consensus in the field (Zhai & Wibowo, 2022; Kruk et al., 2020).

Beyond technical usability, our findings highlight the importance of affective and empathetic design in fostering learner persistence. As Zhai and Wibowo (2022) argue, chatbots that adopt socially supportive dialogue styles can meaningfully reduce learner anxiety and sustain engagement, a pattern reflected in our participants' hesitancy in conversing in English and their appreciation of the bot's non-judgmental space.

The observed pattern of rapid but fragile confidence improvements parallels the "novelty-decay" phenomenon documented by Huang et al. (2022) and the empirical plateau observed between the fourth and sixth sessions by Fryer et al. (2017). This study enriches this narrative by showing that technical friction can cause immediate reversals in learner confidence, thereby supporting the proposition that usability issues must be addressed before pedagogical enhancements (Kasneci et al., 2023). These fragile confidence gains also resonate with productive failure theory (Kapur, 2016), which argues that initial struggle and failure are not detrimental but instead lay the cognitive groundwork for deeper conceptual understanding when followed by structured guidance. In this context, the abrupt confidence dips triggered by usability breakdowns may serve as productive moments of dissonance, revealing overestimated mastery and inviting reflection. Such dynamics highlight the need to design learning technologies that carefully balance challenges with scaffolded support, ensuring that technical barriers do not derail learners before they can benefit from pedagogical gains.

Learner self-reports of vocabulary acquisition and increased speaking confidence align with meta-analytic findings on the efficacy of chatbot-assisted language learning (Zhang et al., 2023; Wang et al., 2024). While subjective, these reports correspond with observed longer utterances and more spontaneous questions during the interaction, supporting the chatbot's role as a low-stakes practice tool. Nonetheless, formal assessments remain necessary to substantiate learning outcomes rigorously.

Differences in stakeholder preferences for practice venues echo tensions reported in recent AI dashboard research (Leeuwen et al., 2023). Notably, the strong learner preference for private, home-based practice adds a novel dimension often absent from existing literature, foregrounding equity and contextual realities in low-resource settings.

Lastly, by triangulating multi-stakeholder perspectives in a low-resource, multilingual environment, this study bridges a significant gap left by extant research, which predominantly features high-resource, single-stakeholder samples (Zhang et al., 2023; Wang et al., 2024; Godwin-Jones, 2014).

4.3 Implications for research and practice

The findings highlight several critical imperatives for designing and deploying educational chatbots in low-resource contexts. First, achieving intelligibility through accent-matched, variable-speed speech and robust ASR, particularly for regional proper nouns, is foundational. Such adjustments are comparatively low-effort but have an outsized impact on learner affect and engagement, corroborating recommendations by Zhai and Wibowo (2022) and Munro and Derwing (1998). Simplifying the microphone interaction model is critical to reducing user error and frustration. We recommend replacing the current hold-to-talk mechanism with a single-tap activation and increasing the size and visibility of the microphone button to enhance accessibility. On the backend, improvements to the automatic speech recognition (ASR) system are necessary to handle proper nouns, Hindi loanwords, and code-switched phrases (Hinglish). Minimising recognition errors and reducing inappropriate message blocking will significantly improve the conversational flow and learner confidence.

Second, a dual-mode rollout featuring an initial, teacher-supervised classroom phase followed by optional, privacy-respecting home use could reconcile institutional oversight with learners' expressed preferences. This approach parallels the blended instructional models that Fryer et al. (2017) advocated and aligns with the mobile learning strategies that Li and Hegelheimer (2013) identified. Our findings concerning the tension between leadership's demand for detailed learner data and teachers' limited capacity resonate with Nouri et al.'s (2021) findings that dashboards must prioritise simplicity and actionable alerts to be feasible in classroom contexts. Balancing these needs is crucial for the sustainable adoption of AI-supported language assessment.

Third, analytics should be designed to surface actionable exceptions, such as persistent grammatical errors, rather than burden teachers with exhaustive transcripts—a conclusion supported by findings from Leeuwen et al. (2023). Finally, interface refinements addressing cognitive and motor load, including search functions, visual cues, and simplified microphone interaction, will be essential to support sustained use and scaffold initial adoption.

The reliance on low-end mobile devices and variable connectivity among our learners aligns with Kukulska-Hulme's (2020) observations that mobile-assisted learning in developing regions must carefully address infrastructural and digital literacy challenges. This underscores the imperative for chatbot designs that accommodate intermittent network access and simplified interfaces rather than transplanting high-resource models indiscriminately.

While adult supervision and support are essential during initial chatbot use, especially for younger or less digitally literate learners, limited teacher bandwidth and variable parental involvement pose significant challenges. Teachers reported availability restricted to approximately 30 minutes, three times per week, and parents' capacity to assist was inconsistent. These constraints suggest that scalable chatbot deployment requires a highly intuitive, low-maintenance interface and clear training resources for educators and families. Additionally, consideration of supplementary staffing or community-based facilitators may be necessary to sustain effective use.

Finally, the ethical dimensions of deploying conversational AI in multilingual, low-resource schools must remain central, especially regarding data privacy, bias, and equitable access. Policymakers must mandate localised ASR benchmarks as they rallied around data privacy safeguards (UNESCO, 2023). As Williamson and Eynon (2020) and Selwyn et al. (2020) caution, without careful governance and design for justice, such technologies risk reinforcing existing educational inequities.

4.4 Limitations and future directions

This study's findings are primarily from qualitative and observational data, subject to interpretive coding and technological variability inherent in real-world deployment. Percentage estimates presented reflect thematic trends rather than precise prevalence measures. While triangulation across multiple stakeholder groups strengthens credibility, these limitations underscore the exploratory nature of the research. The study's short intervention period (six days) limits our ability to observe long-term engagement and learning outcomes. Additionally, the absence of pre- and post-speaking assessments prevents definitive claims about learning gains. Future studies incorporating formal pre- and post-assessments and larger, more diverse samples with longitudinal designs and objective measures would help validate and extend these insights.

Future research should integrate standardised speaking assessments with qualitative tracking to better quantify learning outcomes. Controlled experiments manipulating accent and speech rate, as pioneered by Munro and Derwing (1998), could elucidate their specific impacts on comprehension. Comparative studies of LLM-based versus hybrid rule-based chatbots would also be valuable to clarify the cost-benefit profile of generative AI in bandwidth-limited settings (Kasneci et al., 2023).

4.5 Recommendations and next steps

1. **Technical refinement:** Prioritise tap-to-record interfaces, accent- and speed-adjusted speech output, ASR improvements for proper nouns, and buffering strategies to mitigate connectivity disruptions.

2. **Pedagogical scaffolding:** Develop toggleable bilingual prompts and curate content aligned with curricular and interest-based topics.

3. **Pilot evaluation:** Implement a blended six-week pilot incorporating automated logging and pre-/post-standardised speaking tests.

4. **Interface improvements:** Introduce features such as search capability, colour-coded activity screens, contextual emojis, and enlarged single-tap microphone controls to reduce user burden.

5. **Role-specific dashboards:** Design separate, streamlined interfaces for teachers and students to reconcile differing analytic needs.

6. **Stakeholder co-design:** Engage teachers and learners in iterative prototyping to optimise usability and information flow.

Together, these measures will advance the chatbot from a promising prototype to a sustainable, contextually adapted classroom companion, enriching the sparse evidence base on AI-mediated language learning in multilingual, low-resource educational environments.

References

Amazon Web Services: **ConveGenius builds a SwiftChat conversational AI platform on AWS to improve education for 100 million students.** 2023, September 21.
Reference Source

Arroyo I, Cooper DG, Burleson W, *et al.*: **Emotion sensors go to school.** *Proceedings of the International Conference on Artificial Intelligence in Education.* 2009; **200**(1): 17–24.
Publisher Full Text

Cheng L, Croteau E, Baral S, *et al.*: **Facilitating Student Learning With a Chatbot in an Online Math Learning Platform.** *J. Educ. Comput. Res.* 2024; **62**(4): 907–937.
Publisher Full Text

Chi J, Bell P: **Improving Code-switched ASR with Linguistic Information.** *Proceedings of the 29th International Conference on Computational Linguistics.* 2022; 7171–7176.
Reference Source

Colby KM: **Modeling a paranoid mind.** *Behav. Brain Sci.* 1981; **4**(4): 515–534.
Publisher Full Text

Derwing TM, Munro MJ: **Second Language Accent and Pronunciation Teaching: A Research-Based Approach.** *TESOL Q.* 2005; **39**(3): 379.
Publisher Full Text

de Simone M, Tiberti F, Barron Rodriguez M, *et al.*: *From Chalkboards to Chatbots: Evaluating the Impact of Generative AI on Learning Outcomes in Nigeria (WPS11125).* Washington, DC: World Bank; 2025.
Publisher Full Text

Du J, Daniel BK: **Transforming language education: A systematic review of AI-powered chatbots for English as a foreign language speaking practice.** *Computers and Education: Artificial Intelligence.* 2024; **6**: 100230.
Publisher Full Text

Engeström Y: **Studies in Expansive Learning: Learning What Is Not Yet There.** *Studies in Expansive Learning.* Cambridge University Press; 2016.
Publisher Full Text

Fryer LK, Ainley M, Thompson A, *et al.*: **Stimulating and sustaining interest in a language course: An experimental comparison of Chatbot and Human task partners.** *Comput. Hum. Behav.* 2017; **75**: 461–468.
Publisher Full Text

Fu Y, Weng Z, Wang J: **Examining AI Use in Educational Contexts: A Scoping Meta-Review and Bibliometric Analysis.** *Int. J. Artif. Intell. Educ.* 2024; **2024**: 1–57.
Publisher Full Text

Graesser AC, Wiemer-Hastings K, Wiemer-Hastings P, *et al.*: **AutoTutor: A simulation of a human tutor.** *Cogn. Syst. Res.* 1999; **1**(1): 35–51.
Publisher Full Text

Godwin-Jones R: **Global reach and local practice: The promise of MOOCs - Language Learning and Technology.** *Lang. Learn. Technol.* 2014; **18**(3): 5–15.
Reference Source

Guo K, Li Y, Li Y, *et al.*: **Understanding EFL students' chatbot-assisted argumentative writing: An activity theory perspective.** *Educ. Inf. Technol.* 2024; **29**(1): 1–20.
Publisher Full Text

Huang W, Hew KF, Fryer LK: **Chatbots for language learning—Are they really useful? A systematic review of chatbot-supported language learning.** *J. Comput. Assist. Learn.* 2022; **38**(1): 237–257.
Publisher Full Text

Hwang GJ, Chang CY: **A review of opportunities and challenges of chatbots in education.** *Interact. Learn. Environ.* 2023; **31**(7): 4099–4112.
Publisher Full Text

Jeon J, Lee S, Choi S: **A systematic review of research on speech-recognition chatbots for language learning: Implications for future directions in the era of large language models.** *Interact. Learn. Environ.* 2023.
Publisher Full Text

Kamal MM: **IT innovation adoption in the government sector: identifying the critical success factors.** *J. Enterp. Inf. Manag.* 2006; **19**: 192–222.

Kasneci E, Sessler K, Küchemann S, *et al.*: **ChatGPT for good? On opportunities and challenges of large language models for education.** *Learn. Individ. Differ.* 2023; **103**: 102274.
Publisher Full Text

Laun M, Wolff F: **Chatbots in education: Hype or help? A meta-analysis.** *Learn. Individ. Differ.* 2025; **119**: 102646.
Publisher Full Text

Labadze L, Grigolia M, Machaidze L: **Role of AI chatbots in education: systematic literature review.** *Int. J. Educ. Technol. High. Educ.* 2023; **20**(1): 1–17.
Publisher Full Text

van Leeuwen A, Strauß S, Rummel N: **Participatory design of teacher dashboards: navigating the tension between teacher input and theories on teacher professional vision.** *Front. Artif. Intell.* 2023; **6**: 1039739.
Publisher Full Text

Li Y, Zhou X, Yin HB, *et al.*: **Design language learning with artificial intelligence (AI) chatbots based on activity theory from a systematic review.** *Smart Learn. Environ.* 2025; **12**(1): 1–23.
Publisher Full Text

Kapur M: **Examining Productive Failure, Productive Success, Unproductive Failure, and Unproductive Success in Learning.** *Educ. Psychol.* 2016; **51**(2): 289–299.
Publisher Full Text

Li Z, Hegelheimer V: **Mobile-assisted grammar exercises: Effects on self-editing in L2 writing.** *Lang. Learn. Technol.* 2013; **17**(3): 135–156.
Reference Source

Munro MJ, Derwing TM: **The effects of speaking rate on listener evaluations of native and foreign-accented speech.** *Lang. Learn.* 1998; **48**(2): 159–182.
Publisher Full Text

Okonkwo CW, Ade-Ibijola A: **Chatbots applications in education: A systematic review.** *Computers and Education: Artificial Intelligence.* 2021; **2**: 100033.
Publisher Full Text

Picard RW: *Affective Computing.* The MIT Press; 1997.
Reference Source

Rigas D, Alseid M: **Multi-modal aided Presentation of Learning Information: A Usability Comparative Study.** *Proceedings of IADIS International Conference on Interfaces and Human-Computer Interaction.* 2008; 234–238.

Satar HM, Akcan S: *Pre-service EFL teachers' online participation, interaction, and social presence.* University of Hawaii National Foreign Language Resource Center; 2018.
Reference Source

Sullivan P, Shibano T, Abdul-Mageed M: **Improving Automatic Speech Recognition for Non-Native English with Transfer Learning and Language Model Decoding.** *Signals Commun. Technol.* 2022; 21–44.
Publisher Full Text

Sweller J: **Cognitive Load During Problem Solving: Effects on Learning.** *Cogn. Sci.* 1988; **12**(2): 257–285.
Publisher Full Text

Tondeur J, van Braak J, Ertmer PA, *et al.*: **Understanding the relationship between teachers' pedagogical beliefs and technology use in education: a systematic review of qualitative evidence.** *Educ. Technol. Res. Dev.* 2017; **65**(3): 555–575.
Publisher Full Text

UNESCO: *Guidelines for the governance of digital platforms: safeguarding freedom of expression and access to information through a multi-stakeholder approach.* UNESCO; 2023.
Reference Source

Vygotsky LS: **Mind in Society.** Cole M, Jolm-Steiner V, Scribner S, *et al.*, editors. *Mind in Society.* Harvard University Press; 1978.
Publisher Full Text

Wang F, Cheung ACK, Neitzel AJ, *et al.*: **Does Chatting with Chatbots Improve Language Learning Performance? A Meta-Analysis of Chatbot-Assisted Language Learning.** *Rev. Educ. Res.* 2024.
Publisher Full Text

Weizenbaum J: **ELIZA-A computer program for the study of natural language communication between man and machine.** *Commun. ACM.* 1966; **9**(1): 36–45.
Publisher Full Text

Williamson B, Eynon R: **Historical threads, missing links, and future directions in AI in education.** *Learn. Media Technol.* 2020; **45**(3): 223–235.
Publisher Full Text

Wood D, Bruner JS, Ross G: **THE ROLE OF TUTORING IN PROBLEM SOLVING.** *J. Child Psychol. Psychiatry.* 1976; **17**(2): 89–100.
Publisher Full Text

Wu X, Li R: **Unraveling Effects of AI Chatbots on EFL Learners' Language Skill Development: A Meta-analysis.** *Asia-Pac. Educ. Res.* 2024.
Publisher Full Text

Zhai C, Wibowo S: **A systematic review on cross-culture, humor and empathy dimensions in conversational chatbots: the case of second language acquisition.** *Heliyon.* 2022; **8**(12): e12056.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Zhang S, Shan C, Lee JSY, *et al.*: **Effect of chatbot-assisted language learning: A meta-analysis.** *Educ. Inf. Technol.* 2023; **28**(11): 15223–15243.
Publisher Full Text