Designing ValoTheko: A Formative Needs Assessment and Low-Fidelity Design Concept Study of a Conversational User Interface for Bangla Multimodal Mental Health Survey

Deawan Rakin Ahamed Remal
CCDS
Independent University, Bangladesh
Dhaka, Bangladesh
2511929@iub.edu.bd

Sinthia Chowdhury
CCDS
Independent University, Bangladesh
Dhaka, Bangladesh
2511930@iub.edu.bd

Ashraful Islam
CCDS
Independent University, Bangladesh
Dhaka, Bangladesh
ashraful@iub.edu.bd

AKM Mahbubur Rahman
CCDS
Independent University, Bangladesh
Dhaka, Bangladesh
akmmrahman@iub.edu.bd

Moinul Zaber
Institute of Development Studies
University of Sussex
Brighton, United Kingdom
m.zaber@ids.ac.uk

M. Ashraful Amin[1]
CCDS
Independent University, Bangladesh
Dhaka, Bangladesh
aminmdashraful@iub.edu.bd

Accurate data is critical for research, and surveys are the link connecting researchers with public perception. However, conducting surveys is challenging in Low and Middle Income Countries (LMICs) such as Bangladesh due to limited digital experience, literacy issues, and discomfort in disclosing sensitive personal information. These factors exclude underrepresented populations and create biases, especially in mental health surveys, where stigma and privacy concerns affect participation in digital interventions. To mitigate these issues, we propose ValoTheko, a low-fidelity design concept for a Bangla multimodal Conversational User Interface (CUI) for mental health surveys. ValoTheko is examined through a two-fold approach comprising a formative needs assessment followed by a low-fidelity prototype concept evaluation. In the first phase, snowball sampling was used to recruit 137 participants (male = 72, female = 65, age = 18-65) for a needs-assessment survey. Based on the findings, we derived design requirements and developed a low-fidelity prototype. The second phase involved a formative prototype concept evaluation conducted with 17 newly recruited (random sampling) participants (male = 8, female = 9, age = 18-65), where participants reviewed a non-interactive screen-by-screen walkthrough and provided their anticipated reactions rather than interacting with a working system. The findings indicate that mixed-mode interaction represents a promising approach for sensitive mental health data collection, as it accommodates both expressive voice input and click-based response mechanisms.

Mental Health Survey, CUI, Multimodal Interaction, Sensitive Data Collection, User-Centered Design

1. INTRODUCTION

Mental health conditions remain highly underreported in Bangladesh: although reluctance to disclose sensitive information is broadly documented across populations, in many Bangladeshi communities mental distress is socially stigmatized and treated as taboo. In addition, mental illness stigma is widespread across cultures and a significant barrier to global mental-health goals (Faruk et al. 2023); cultural stigma, fear of judgment, and concerns about family reputation discourage disclosure of psychological distress in surveys (Khan et al. 2025; Patel 2007), which biases mental-health research through incomplete disclosure. Human-mediated surveys can amplify social desirability bias when questions involve mental health or other sensitive topics, as participants may provide socially acceptable responses rather than fully disclosing their experiences (Bispo Júnior 2022; Tourangeau and Yan 2007). Prior work in HCI and computer-mediated interviewing suggests that reducing interviewer presence may lower perceived judgment and increase willingness to disclose sensitive information (Slack and Van Cura

[1]Also at Department of Information Sciences and Technology, George Mason University, VA, USA (Email: mamin29@gmu.edu).

© Remal et al. Published by BCS Learning and Development Ltd.

1968; Lucas et al. 2014; Pickard and Roster 2020; Butler et al. 2009; Krumpal 2013).

The shift from paper-and-pencil to computer-assisted personal interviewing (CAPI) reduced data-entry errors and improved survey management (Caeyers et al. 2012; Schräpler et al. 2006), and platforms such as KoBoCollect are widely used in low-resource settings (Nampa et al. 2020). However, many CAPI tools are designed to support trained enumerators rather than independent participation by low-literacy respondents. Because an interviewer is still present, they do not address respondent-side barriers such as literacy, typing difficulty, and privacy concerns (De Leeuw et al. 1995). Limited digital literacy in Bangladesh (Shadat et al. 2020) further means low-literacy respondents may be under-represented or give shorter, less expressive responses in text-heavy formats. We treat disclosure as an interaction-design problem as well as a public-health one: the modality, language, and social setting of a survey shape what respondents are willing and able to report, motivating a respondent-facing, mixed-mode approach for Bangla mental-health data collection. We prioritise voice as the expressive modality because reading and typing burden low-literacy respondents most, retaining click-based input in parallel so respondents can preserve privacy when speaking aloud is inappropriate. We present ValoTheko as a Bangla multimodal conversational survey concept for sensitive mental health data collection.

The scope of this study is limited to evaluating a conceptually described Conversational User Interface (CUI), without implementing or testing conversational functionality, automatic speech recognition, or live interaction; a fully functional system remains future work. The study contributes (i) empirical evidence on respondent-side barriers to mental-health survey participation in Bangladesh, (ii) a set of design requirements for Bangla multimodal conversational survey systems, and (iii) early formative feedback on a low-fidelity design concept that identifies usability, privacy, and modality-specific concerns to be addressed before high-fidelity development.

2. BACKGROUND STUDY

The treatment gap for mental health in LMICs is well documented (Chau et al. 2025), but the corresponding data gap is often overlooked. In Bangladesh, social norms constrain disclosure: societal expectations often equate mental resilience with silence, particularly for men expected to be stoic providers, while concerns about family reputation shape women's willingness to discuss emotional distress (Mokhwelepa and Sumbane 2025; Sangar and Howe 2021). The presence of an interviewer can trigger impression management, in which respondents give socially desirable answers (Tourangeau and Yan 2007).

2.1. Voice Based Surveys and Interactive Voice Response in LMICs

In response to literacy barriers, researchers have long experimented with voice-based systems (Sharma Grover et al. 2009), and Interactive Voice Response (IVR) has been used for health surveillance as a way to reach populations with low literacy. Sherwani et al. showed that low-literate community health workers could access health information more effectively via speech interfaces than via text-based systems (2007). VideoKheti, a multimodal interface (voice + touch + graphics) in India, similarly found that combining modalities reduced error rates versus voice-only or text-only systems (Medhi Thies 2015). However, traditional IVR has three limitations for mental health. First, rigidity: DTMF key-press inputs force complex emotions into pre-defined categories, and short questionnaires limit in-depth, rich data collection (Rahmanian and Basil 2024). Second, cognitive load: complex wording (Brosnan et al. 2021) and non-native language (Rocabado et al. 2025) slow understanding, especially for users with depressive or anxious symptoms. Third, lack of empathy: often-robotic voices can fail to convey the empathy needed for sensitive disclosure (Schreibelmayr and Mara 2022), prompting frustration and abandonment.

2.2. The Shift to CUIs

The limitations of IVR have driven a shift towards CUIs and agents powered by Natural Language Processing (NLP). Unlike IVR, a CUI can understand natural speech, handle open-ended responses, infer emotional state, and provide empathetic feedback (Wang et al. 2025; Chin et al. 2025). Existing IVR deployments in Bangladesh depend heavily on digital literacy and mobile-network availability (Khan et al. 2023), and IVR's demand for sustained attention is difficult for elderly users (Miller et al. 2013), whose participation matters because depressive symptoms are common among older adults in Bangladesh (Rahman et al. 2020).

Several global mental-health support chatbots such as Woebot and Wysa exist, but they are predominantly English-centric and text-based, which excludes low-literacy Bangla speakers (Haque and Rubya 2023). Bangla mental-health chatbots have recently begun to emerge (Hussna et al. 2020; Ahamed et al. 2025), but Bangla CUIs remain comparatively under-developed relative to high-resource languages. A range of Bangla chatbot systems have

been developed across different domains. George et al. (2021) reviewed chatbot work bridging Indian languages, while Bangla-specific systems such as Golpo, Doly, Bengali Informative Chatbot (BIC), and Disha illustrate early efforts to enable conversational interaction in Bangla using retrieval-based, machine-learning, and NLP-driven approaches. Nahar et al. (2025) designed AmarDoctor, a Bangla voice-interactive health triage assistant that handles three major dialects to help overcome literacy barriers.

Two prior Bangladesh-focused studies sit closest to the present work and merit explicit positioning. Bhattacharjee et al. (2021) conducted an interview study on the social determinants of mental health among Bangladeshi undergraduate students; their work surfaces context-level factors influencing mental-health experience but does not address the interaction design of survey instruments themselves. Khan et al. (2023) examined community perception and acceptability of IVR survey methods in Bangladesh; their findings highlight digital-literacy and network limitations of touch-tone IVR but do not propose a multimodal, respondent-facing alternative. This work extends these efforts by examining respondent-side barriers to mental-health survey participation specifically as an interaction-design problem, and by translating empirical findings into a low-fidelity multi-modal prototype rather than retaining the enumerator- or IVR-centred model.

3. RESEARCH DESIGN AND METHODOLOGY

We followed a multi-stage, User-Centered Design (UCD) process to develop and examine a low-fidelity prototype concept for ValoTheko. The workflow comprised five stages: a background study, a 137-participant needs assessment, mixed quantitative and qualitative analysis, low-fidelity prototyping of a non-interactive interaction flow, and a formative evaluation with 17 participants. Figure 1 illustrates the workflow of the entire study.

3.1. Use of AI

During the preparation of this manuscript, the authors used ChatGPT-4 (`https://chatgpt.com/`) and QuillBot (`https://quillbot.com/`) for language assistance and to improve the clarity and readability of the text. These tools were not used for idea generation, literature review, data collection, or data analysis, which were carried out entirely by the authors. After using these tools, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

4. NEEDS ASSESSMENT STUDY

We conducted a formative mixed-method needs assessment focusing on respondent barriers rather than system performance. A total of 137 participants (age=18-65) from diverse demographic, socioeconomic, and educational backgrounds were recruited via snowball sampling through in-person visits to selected rural and urban areas across Bangladesh, and sessions were conducted in person by the research team. Eligible respondents were 18 years or older, able to communicate in Bangla, and willing to provide informed consent. The final sample included 72 male and 65 female participants, 55 urban and 82 rural participants, and substantial representation from high, moderate, and low or no literacy groups. An illustration of the demographic information of the needs-assessment participants is shown in Appendix A, Table A6.

Each session began with a brief explanation of the study purpose and voluntary participation; all participants gave verbal and written consent and could withdraw or skip questions at any time. The study was approved by our home institution's IRB, responses were anonymised by removing direct identifiers, and standard data-protection measures were followed. We did not collect diagnostic or self-suspected mental health status because asking participants to identify themselves as having a mental health condition could have conflicted with the anonymity promised during the session. The structured survey used seven closed-ended items covering comfort with human-mediated disclosure, including two items that distinguished between an enumerator in general and a male enumerator, comfort with an anonymous self-administered digital mode, preference among text/click-, voice-, and mixed-mode interaction, difficulty reading Bangla text on small screens, difficulty typing on smartphones, and concern about being overheard when speaking aloud in a shared space. All items used a five-point agreement scale (1 = strongly disagree, 5 = strongly agree) and were written in simple, everyday Bangla without clinical terminology. Each finding in Appendix A, Table A5 is based on a single item; the reported percentages give the proportion of the stated denominator selecting agree or strongly agree. Open-ended responses explored interaction preferences, privacy expectations, and disclosure concerns, providing contextual insights that informed the design requirements of the low-fidelity design concept.

We summarised the quantitative responses using descriptive statistics to identify patterns in gendered



Figure 1: *Research workflow: from background study to formative prototype evaluation.*

disclosure discomfort, anonymity preferences, voice-input convenience, typing barriers, and survey burden, which shaped the design requirements. Qualitative data were analysed using an inductive thematic analysis approach. Two members of the research team independently read the full set of Bangla responses, generated initial codes close to participants' wording, and grouped codes into candidate themes (e.g., "fear of being overheard," "discomfort typing feelings," "preference for anonymity"). The team then met to reconcile codes, resolve disagreements through discussion, and consolidate the final theme set. Because the study is formative, we report the resulting themes as design-informing patterns rather than as saturated qualitative findings, and illustrative translated participant statements are included alongside the corresponding findings in Section 5.

5. FINDINGS FROM THE NEEDS ASSESSMENT

The needs assessment revealed that mental-health survey participation in Bangladesh is shaped by the intersection of privacy, literacy, digital access, gendered disclosure discomfort, and interaction modality, with participants facing different barriers depending on literacy, social context, and access to a private space. Appendix A, Table A5 reports each finding with its relevant denominator; these values are interpreted as formative design evidence rather than population-level estimates.

5.1. Privacy and Human Mediated Disclosure

Women reported discomfort discussing sensitive mental-health topics with human interviewers due to concerns about judgment and privacy, with most preferring confidential responses. One participant's translated remark was representative: "I could answer honestly if no one knew it was me." Removing the direct human presence can therefore reduce the barrier to disclosure, so the prototype focuses on making confidentiality clear. We asked separately about male enumerators because gender-of-interviewer effects are consequential here: norms around family reputation shape women's willingness to voice distress (Sangar and Howe 2021; Khan et al. 2025), while expectations of male stoicism discourage men from disclosing it (Mokhwelepa and Sumbane 2025). A randomised trial of IVR surveys in Bangladesh found higher contact, response, and cooperation rates with a female rather than a male survey voice, which the authors attribute to mixed-gender interaction being viewed as less appropriate in the local context (Labrique et al. 2023). Comfort with an enumerator in the abstract may therefore overstate comfort with the enumerator actually assigned: 78.5% (51/65) of female participants reported discomfort specifically with a male enumerator. Digitising a questionnaire while retaining a face-to-face administrator does not remove this barrier, since the barrier is the assigned interlocutor rather than the medium. It instead motivates self-administration and the click-based response path.

5.2. Literacy, Reading, and Typing Barriers

The needs assessment showed that text-heavy digital surveys can create barriers for participants with low or moderate literacy, with 99 of 137 participants reporting low/no or moderate literacy. However, not all would necessarily be unable to complete a text-based survey. The participants described two related barriers. The first was a reading and comprehension barrier, where some participants found Bangla text-based surveys difficult to read on small screens, especially when the questions used formal, clinical, or unfamiliar terminology. The second was an input barrier among participants with low or no literacy. Among these participants, 77.1% (37 of 48) reported difficulty typing answers on smartphones. This was particularly relevant for open-ended mental health questions, where respondents are expected to describe feelings, symptoms, or personal experiences in their own words. One participant described this directly: "My education is limited. It takes me several minutes to find Bangla letters on the keyboard. It is much easier to simply speak if no one is listening." Accessibility therefore involves more than device availability. This pattern informed the low-fidelity ValoTheko prototype's requirement for voice-supported input, audio support, and simple Bangla instructions.

5.3. Voice Preference and the Limits of Voice-Only Interaction

Many participants preferred voice-supported interaction for open-ended questions; 83.2% (114/137) reported that voice input could reduce typing effort and help them express thoughts more easily. However, they also raised privacy concerns, as speaking

about mental health in shared homes, workplaces, or public spaces could be overheard. One participant expressed this directly: "Voice mode may let others hear the conversation. Button tapping makes it easier to take part in mental health surveys from anywhere." Voice input should therefore not fully replace text or click-based interaction; the prototype adopts a mixed-mode design allowing users to switch between voice and click depending on comfort, privacy, and preference.

6. DESIGN REQUIREMENTS AND LOW-FIDELITY PROTOTYPE

From the needs-assessment findings, we synthesised design requirements, translating them into a low-fidelity prototype meant to make the interaction concept visible and evaluable rather than to validate a functional system. Because the barriers were interdependent, the prototype centres on a mixed-mode interaction concept giving respondents greater control over how they answer. Appendix A, Table A1 maps each finding to a design requirement and prototype implication. Figure 2 shows the proposed low-fidelity prototype, whose six core components are a consent and privacy screen, mode selection, a structured-question screen, an open-ended response screen, privacy reminders, and a distress-support closing screen. A concrete interaction flow (Appendix A, Table A2) shows how ValoTheko departs from rigid IVR-style systems (Rojas-Galeano 2025): respondents use click-based responses for private or structured items and voice for open-ended expression, with skip and pause options available throughout.

7. FORMATIVE EVALUATION OF THE LOW-FIDELITY PROTOTYPE

We conducted a formative evaluation to examine whether the interaction flow, privacy features, and mixed-mode input concept were understandable and whether the design addressed privacy and accessibility before high-fidelity development. Because participants never interacted with a working system or experienced Bangla speech recognition, the Likert ratings should be read as anticipated reactions to a described concept and formative design signals, not as measured usability or validation of a system.

7.1. Participants and Procedure

We recruited 17 participants (age=18-65) using random sampling across diverse literacy levels, gender, location, technology access, and privacy concerns; this sample size is consistent with established guidance on formative evaluations with low-fidelity prototypes. Participants were able to communicate in Bangla and provided informed consent, and were newly recruited rather than drawn from the original 137 needs-assessment participants, to obtain fresh reactions to the concept. Table A4 (Appendix A) summarises the participant profile.

Each session described the study goal and the low-fidelity state of the prototype; we walked participants through the screens and intended flow, asking them to imagine using the system at home, in clinics, workplaces, and shared spaces. Participants then completed a 5-point Likert assessment (1 = strongly disagree, 5 = strongly agree), analysed with descriptive statistics (mean, SD, agreement percentage); no inferential tests were run due to the small sample size. Each Likert item is stated explicitly and reported in Appendix A, Table A3. Open-ended feedback was analysed using the same thematic procedure described in Section 4.

7.2. Evaluation Findings

Results are shown in Appendix A, Table A3. The most prominent outcome was the importance of clearly explaining privacy and anonymity, rated highest among all items (M = 4.59, SD = 0.62; 94.1% agreement), reinforcing the needs-assessment finding that privacy must be made visible before sensitive questions. Participant control also mattered: skip and pause options rated M = 4.24 (SD = 0.83; 76.5% agreement). The mixed-mode approach received moderate agreement, with click/tap input for private questions and modality switching both rated M = 3.82 (64.7% agreement). This supports the decision to avoid both text-only and voice-only formats. Voice input received mixed agreement (M = 3.71, SD = 1.10), and discomfort using voice in shared spaces showed high variation (M = 3.47, SD = 1.46), reinforcing that voice should remain optional and context-sensitive. The lowest-rated item was the intended Bangla wording (M = 3.12, SD = 1.32; 41.2% agreement). Because the prototype did not present a complete Bangla interface, this is feedback on the explained language direction and signals that future iterations should use simpler Bangla, reduce clinical wording, and add audio or icon supported guidance. The open-ended feedback pointed to corresponding refinements: simpler Bangla with audio guidance, clearer privacy communication before sensitive questions, optional voice with click-based alternatives, and visible modality switching.

8. DISCUSSION

The needs-assessment findings and prototype evaluation converge on three design implications: (i) mixed-mode interaction, (ii) privacy by design, and (iii) self-administration to address gendered discomfort.



Figure 2: *An overview of the proposed low-fidelity ValoTheko prototype.*

8.1. Multimodality as a Response to Uneven Access and Privacy

Text-only surveys challenge low-literacy users, while voice is easier but risks privacy in shared spaces. The evaluation ratings map onto this tension directly, so supporting multiple modalities is a fundamental design necessity rather than an optional feature. This extends prior conversational-survey findings (Celino and Calegari 2020) to a low-literacy, multilingual setting where modality choice is a privacy mechanism, not only a preference.

8.2. Privacy as an Interaction Feature

Interfaces should provide contextual privacy assurances and reinforce user control before sensitive disclosures or voice capture. Because voice processing limits complete anonymity, systems should communicate transparently about data collection, storage, and access rather than promise it.

8.3. Gendered Disclosure Discomfort and the Need for Self-Administered Interaction

Self-administered interfaces may reduce the social pressure of in-person interviews and can encourage more candid self-disclosure (Pickard and Roster 2020; Lapidot-Lefler and Barak 2015), provided they do not introduce new fears around hidden recordings or being overheard. In this study we foreground private, self-administered disclosure because the needs assessment identified social observation and family reputation as primary barriers. Integrating multi-informant assessment, while valuable clinically, would sit in tension with the confidentiality needs our participants prioritized. This remains an important design trade-off to be addressed in future work.

9. LIMITATIONS AND FUTURE WORK

As an early-stage study using a non-interactive low-fidelity prototype, our results highlight anticipated usefulness and privacy concerns rather than actual interaction performance, survey completion, or real-world disclosure. The needs-assessment sessions were themselves administered in person by researchers, so they may have been subject to interviewer presence and social desirability. We treat the self-reported data as formative signals rather than population-representative measurements. We also did not record mental-health status; disclosure comfort and modality preference may differ between respondents who experience mental distress and those who do not, and capturing this under an anonymised protocol is a goal for the high-fidelity study. Future work will develop a high-fidelity ValoTheko addressing these refinements, followed by large-scale field testing across rural, semi-urban, and urban settings. It must also address Bangla speech recognition, dialect handling, noise robustness, secure voice-data processing and retention, explicit recording consent, and distress-support protocols.

10. CONCLUSION

This study presents ValoTheko, a Bangla multimodal conversational interface design concept for mental-health surveys in Bangladesh, where neither text only nor voice only formats are sufficient and the evidence supports a flexible, privacy-aware mixed-mode approach letting respondents switch between click and voice based on question sensitivity, literacy needs, and immediate environment. The study contributes empirical evidence on these barriers, a low-fidelity prototype with design requirements, and an HCI design process linking needs assessment, prototyping, and formative evaluation for sensitive data collection in low-resource settings. A fully functional Bangla CUI remains future work.

ACKNOWLEDGEMENT

This research was supported by Independent University, Bangladesh (IUB).

REFERENCES

Ahamed, S., Chowdhury, M. H., Zaman, M., Agarwala, A., Al Mashud, M. A., Al Muntasir, T., Rahman, R., Shahriyar, R., Sarker, F., Mamun, K. A., et al. (2025). SuSastho.AI: A multimodal medical copilot for adolescents using evidence-based medicine and large language models. *Informatics in Medicine Unlocked*, page 101720.

Bhattacharjee, A., Haque, S. M. T., Hady, M. A., Alam, S. M. R., Rabbi, M., Kabir, M. A., and Ahmed, S. I. (2021). Understanding the social determinants of mental health of undergraduate students in Bangladesh: Interview study. *JMIR Formative Research*, 5(11):e27114.

Bispo Júnior, J. P. (2022). Social desirability bias in qualitative health research. *Revista de Saúde Pública*, 56:101.

Brosnan, K., Grün, B., and Dolnicar, S. (2021). Cognitive load reduction strategies in questionnaire design. *International Journal of Market Research*, 63(2):125–133.

Butler, S. F., Villapiano, A., and Malinow, A. (2009). The effect of computer-mediated administration on self-disclosure of problems on the Addiction Severity Index. *Journal of Addiction Medicine*, 3(4):194–203.

Caeyers, B., Chalmers, N., and De Weerdt, J. (2012). Improving consumption measurement and other survey data through CAPI: Evidence from a randomized experiment. *Journal of Development Economics*, 98(1):19–33.

Celino, I. and Calegari, G. R. (2020). Submitting surveys via a conversational interface: An evaluation of user acceptance and approach effectiveness. *International Journal of Human-Computer Studies*, 139:102410.

Chau, L. W., Lam, R. W., Minas, H., Hayashi, K., Nguyen, V. C., and O'Neil, J. (2025). Digital health interventions for depression and anxiety in low- and middle-income countries: Rapid scoping review. *JMIR Mental Health*, 12:e68296.

Chin, H., Baek, G., Cha, C., and Cha, M. (2025). Chatbots' empathetic conversations and responses: A qualitative study of help-seeking queries on depressive moods across 8 commercial conversational agents. *JMIR Formative Research*, 9:e71538.

De Leeuw, E. D., Hox, J. J., and Snijkers, G. (1995). The effect of computer-assisted interviewing on data quality. A review. *Market Research Society Journal*, 37(4):1–19.

Faruk, M. O., Khan, A. H., Chowdhury, K. U. A., Jahan, S., Sarker, D. C., Colucci, E., and Hasan, M. T. (2023). Mental illness stigma in Bangladesh: Findings from a cross-sectional survey. *Cambridge Prisms: Global Mental Health*, 10:e59.

George, A. S., Muralikrishnan, G., Ninan, L. R., Varrier, P. S., and Dhanya, L. K. (2021). Survey on the design and development of Indian language chatbots. In *2021 International Conference on Communication, Control and Information Sciences (ICCISc)*, volume 1, pages 1–6. IEEE.

Haque, M. D. R. and Rubya, S. (2023). An overview of chatbot-based mobile mental health apps: Insights from app description and user reviews. *JMIR mHealth and uHealth*, 11:e44838.

Hussna, A. U., Laz, A. N. K., Sikder, M. S., Uddin, J., Tinmaz, H., and Esfar-E-Alam, A. M. (2020). PRERONA: Mental health Bengali chatbot for digital counselling. In *International Conference on Intelligent Human Computer Interaction*, pages 274–286. Springer.

Khan, I. A., Shano, S., Islam, K., Shahnaz, S., Al Kibria, G. M., Pariyo, G., Ali, J., Gibson, D. G., and Rahman, M. (2023). Understanding community perception and acceptability of interactive voice response survey method: Findings from formative research in Bangladesh. *Oxford Open Digital Health*, 1:oqad012.

Khan, R., Akter, F., Rahman, M., Amin, N., Rahman, M., and Winch, P. J. (2025). Socio-demographic factors associated with mental health disorders among rural women in Mymensingh, Bangladesh. *Frontiers in Psychiatry*, 16:1446473.

Krumpal, I. (2013). Determinants of social desirability bias in sensitive surveys: A literature review. *Quality & Quantity*, 47(4):2025–2047.

Labrique, A., Nagarajan, M., Kibria, G. M. A., Vecino-Ortiz, A., Pariyo, G. W., Ali, J., et al. (2023). Improving success of non-communicable diseases mobile phone surveys: Results of two randomized trials testing interviewer gender and message valence in Bangladesh and Uganda. *PLOS ONE*, 18(5):e0285155.

Lapidot-Lefler, N. and Barak, A. (2015). The benign online disinhibition effect: Could situational factors induce self-disclosure and prosocial behaviors? *Cyberpsychology: Journal of Psychosocial Research on Cyberspace*, 9(2):Article 3.

Lucas, G. M., Gratch, J., King, A., and Morency, L.-P. (2014). It's only a computer: Virtual humans increase willingness to disclose. *Computers in Human Behavior*, 37:94–100.

Medhi Thies, I. (2015). User interface design for low-literate and novice users: Past, present and future. *Foundations and Trends in Human–Computer Interaction*, 8(1):1–72.

Miller, D., Gagnon, M., Talbot, V., and Messier, C. (2013). Predictors of successful communication with interactive voice response systems in older people. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, 68(4):495–503.

Mokhwelepa, L. W. and Sumbane, G. O. (2025). Men's mental health matters: The impact of traditional masculinity norms on men's willingness to seek mental health support; a systematic review of literature. *American Journal of Men's Health*, 19(3):15579883251321670.

Nahar, N., Ruparel, R. H., Kabir, S., Khan, S. T., Saha, S., and Rashid, M. (2025). AmarDoctor: An AI-driven, multilingual, voice-interactive digital health application for primary care triage and patient management to bridge the digital health divide for Bengali speakers. *arXiv preprint arXiv:2510.24724*.

Nampa, I. W., Mudita, I. W., Riwu Kaho, N. P. L. B., Widinugraheni, S., and Lasarus Natonis, R. (2020). The KoBoCollect for research data collection and management (an experience in researching the socio-economic impact of blood disease in banana). *Jurnal Sosial Ekonomi Pertanian*, 14(3):545–556.

Patel, V. (2007). Mental health in low- and middle-income countries. *British Medical Bulletin*, 81(1):81–96.

Pickard, M. D. and Roster, C. A. (2020). Using computer automated systems to conduct personal interviews: Does the mere presence of a human face inhibit disclosure? *Computers in Human Behavior*, 105:106197.

Rahman, M. S., Rahman, M. A., Ali, M., Rahman, M. S., Maniruzzaman, M., Yeasmin, M. A., Ahmed, N. A. M. F., Abedin, M. M., and Islam, S. M. S. (2020). Determinants of depressive symptoms among older people in Bangladesh. *Journal of Affective Disorders*, 264:157–162.

Rahmanian, J. and Basil, L. (2024). Interactive voice response technology as a data collection tool compared to a household survey: What we learned. *Canadian Journal of Program Evaluation*, 39(1):62–82.

Rocabado, F., Schmitz, G., and Duñabeitia, J. A. (2025). You can stand under my umbrella: Cognitive load in second-language reading. *Behavioral Sciences*, 15(8):1051.

Rojas-Galeano, S. (2025). Beyond IVR touch-tones: Customer intent routing using LLMs. *arXiv preprint arXiv:2510.21715*.

Sangar, M. and Howe, J. (2021). How discourses of sharam (shame) and mental health influence the help-seeking behaviours of British-born girls of South Asian heritage. *Educational Psychology in Practice*, 37(4):343–361.

Schräpler, J.-P., Schupp, J., and Wagner, G. G. (2006). Changing from PAPI to CAPI: A longitudinal study of mode-effects based on an experimental design. Technical Report 593, DIW Discussion Papers, Deutsches Institut für Wirtschaftsforschung, Berlin.

Schreibelmayr, S. and Mara, M. (2022). Robot voices in daily life: Vocal human-likeness and application context as determinants of user acceptance. *Frontiers in Psychology*, 13:787499.

Shadat, M. W. B., Islam, M. S., Zahan, I., and Matin, M. (2020). Digital literacy of rural households in Bangladesh. Technical report, BRAC Institute of Governance and Development (BIGD), BRAC University, Dhaka, Bangladesh.

Sharma Grover, A., Stewart, O., and Lubensky, D. (2009). Designing interactive voice response (IVR) interfaces: Localisation for low literacy users. In *Proceedings of Computers and Advanced Technology in Education (CATE 2009)*, pages 328–335, St. Thomas, US Virgin Islands.

Sherwani, J., Ali, N., Mirza, S., Fatma, A., Memon, Y., Karim, M., Tongia, R., and Rosenfeld, R. (2007). Healthline: Speech-based access to health information by low-literate users. In *2007 International Conference on Information and Communication Technologies and Development*, pages 1–9. IEEE.

Slack, W. V. and Van Cura, L. J. (1968). Patient reaction to computer-based medical interviewing. *Computers and Biomedical Research*, 1(5):527–531.

Tourangeau, R. and Yan, T. (2007). Sensitive questions in surveys. *Psychological Bulletin*, 133(5):859–883.

Wang, C., Peng, T., Yang, W., Bai, Y., Wang, G., Lin, J., Jia, L., Wu, L., Wang, J., Zong, C., et al. (2025). OpenS2S: Advancing fully open-source end-to-end empathetic large speech language model. In *Proceedings of the 2025 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, pages 906–917.

A. SUPPLEMENTARY TABLES

The following tables provide supplementary details on the needs-assessment findings, prototype design rationale, interaction flow and formative evaluation results reported in the main text.

Table A1: *Translation of needs-assessment findings into ValoTheko design requirements*

Needs-assessment finding	Design requirement	Prototype implication
Low-literacy participants reported difficulty with reading and typing.	Reduce dependence on text-heavy interaction.	Provide simple Bangla prompts, audio support, and optional voice input.
Participants expressed discomfort disclosing mental-health information to human enumerators.	Reduce interviewer pressure and make anonymity visible.	Include self-administered interaction and visible privacy explanations.
Female participants reported discomfort discussing mental health with male enumerators.	Support gender-sensitive and private participation.	Allow respondents to answer without face-to-face questioning.
Voice input was preferred for ease of expression.	Support expressive open-ended responses.	Allow verbal responses for descriptive or open-ended questions.
Speaking aloud in public or shared spaces was perceived as risky.	Avoid dependence on voice-only interaction.	Provide click-based alternatives and easy modality switching.
Sensitive mental-health questions may create discomfort.	Preserve participant agency and reduce harm.	Include skip, pause, consent reminders, and distress-support information.

Table A2: *Example ValoTheko interaction flow*

Step	Example system prompt	User action
1. Consent	আপনি চাইলে যেকোনো প্রশ্ন বাদ দিতে পারেন। আপনার উত্তর নাম ছাড়া সংরক্ষণ করা হবে।	User taps Continue or Exit.
2. Mode selection	আপনি কীভাবে উত্তর দিতে চান: চাপ দিয়ে, মুখে বলে, নাকি দুইভাবেই?	User selects click, voice, or mixed mode.
3. Structured item	গত দুই সপ্তাহে আপনার মন খারাপ বা আগ্রহ কমে যাওয়ার অনুভূতি কতবার হয়েছে?	User selects a response option.
4. Optional open response	আপনি চাইলে নিজের ভাষায় আরও কিছু বলতে পারেন। না চাইলে Skip চাপুন।	User speaks, types, or skips.
5. Privacy reminder	আপনি যদি মানুষের মধ্যে থাকেন, মুখে উত্তর না দিয়ে চাপ দিয়ে উত্তর দিতে পারেন।	User switches to click mode if needed.
6. Closing support	আপনার অংশগ্রহণের জন্য ধন্যবাদ। আপনি যদি অস্বস্তি অনুভব করেন, নিচের সহায়তা তথ্য ব্যবহার করতে পারেন।	User views support information.

Table A3: *Formative evaluation results for the low-fidelity prototype ($n = 17$)*

Evaluation Item	Mean	SD	Agree/Strongly Agree
The prototype flow, as explained in the walkthrough, is easy to understand.	3.65	1.22	58.8%
The intended Bangla text/instruction style seems simple enough.	3.12	1.32	41.2%
Voice input would help answer open-ended mental-health questions in a future version.	3.71	1.10	58.8%
Click/tap input would feel safer for private or sensitive questions.	3.82	1.13	64.7%
Switching between voice and click input would give better control.	3.82	1.01	64.7%
I would feel more comfortable answering without a human enumerator present.	3.53	1.01	52.9%
The system must clearly explain anonymity and data privacy before I answer.	4.59	0.62	94.1%
I may feel uncomfortable using voice input in a shared or public space.	3.47	1.46	52.9%
A skip/pause option is important for sensitive mental-health questions.	4.24	0.83	76.5%
Overall, I would be willing to try this prototype after future implementation.	3.59	1.28	58.8%

Table A4: *Profile of formative prototype evaluation participants ($n = 17$)*

Variable	Category	Count (n)	Percentage
Gender	Female	9	52.9%
	Male	8	47.1%
Location	Rural	7	41.2%
	Urban	6	35.3%
	Semi-urban	4	23.5%
Age group	18–25	5	29.4%
	26–40	6	35.3%
	41–65	6	35.3%
Literacy	High	5	29.4%
	Moderate	6	35.3%
	Low/none	6	35.3%
Technology access	Personal smartphone owner	9	52.9%
	Shared household smartphone access	5	29.4%
	Feature phone or shared basic phone access	3	17.6%

Table A5: *Key needs-assessment findings used to inform ValoTheko design requirements ($N = 137$)*

Finding area	Observation	Value
Gendered disclosure discomfort	Female participants reporting discomfort discussing mental health with a male enumerator	78.5% (51/65)
Confidential survey mode	Participants reporting comfort with sharing sensitive information if confidentiality is assured	65.0% (89/137)
Voice-supported input	Participants preferring voice input for ease of expression or reduced typing burden	83.2%(114/137)
Typing barrier	Low- or no-literacy participants reporting difficulty typing answers on smartphones	77.1% (37/48)
Text-heavy survey burden	Participants in low/no- and moderate-literacy groups for whom text-heavy survey barriers were examined	72.3% (99/137)

Table A6: *Demographic characteristics of needs-assessment participants ($N = 137$)*

Variable	Category	Count (n)	Percentage
Gender	Male	72	52.6%
	Female	65	47.4%
Location	Urban	55	40.1%
	Rural	82	59.9%
Age group	18–25	45	32.8%
	26–40	58	42.3%
	41–65	34	24.8%
Literacy	High	38	27.7%
	Moderate	51	37.2%
	Low / none	48	35.0%
Technology access	Smartphone owner	94	68.6%
	Feature phone / shared	43	31.4%