# Local Language Adaptations of Mental Health Self-Help Apps: A Usability Study

**Dr. Tushar Mehrotra**
DCSE, Galgotias University , Greater Noida, UP, India
tushar.mehrotra@galgotiasuniversity.edu.in

## ABSTRACT

**Digital self-help applications for mental health are proliferating, yet many are designed primarily in English and exported with minimal adaptation to users' linguistic and cultural contexts. This study evaluates whether local language adaptations—beyond direct translation—improve usability, comprehension, and perceived trust in mental health self-help apps among adult smartphone users in India. We conducted a mixed-methods, within-subjects usability study (N = 208; speakers of Hindi, Bengali, Marathi, and Tamil) using a counterbalanced design to compare an English baseline app with a linguistically and culturally adapted version. The adapted version incorporated plain-language microcopy, dialect-aware wording, culturally resonant examples and metaphors for mood skills, script-specific typography, text-to-speech for low-literacy support, and localization of crisis resources and privacy statements. Quantitative outcomes included System Usability Scale (SUS), task success, time-on-task, error rates, and comprehension checks for consent and safety content. Qualitative outcomes were drawn from think-aloud and semi-structured interviews analyzed thematically. The adapted app increased SUS by 12.4 points on average (95% CI: 10.7–14.1), reduced median time-on-task by 22%, and improved correct comprehension of privacy statements by 27 percentage points (all p < .001). Gains were largest for participants reporting lower English proficiency and for tasks involving sensitive disclosures (mood logging, safety planning). Participants described the adapted wording as "clearer," "less clinical," and "more humane," and reported higher willingness to continue use and to recommend the app to family members.**

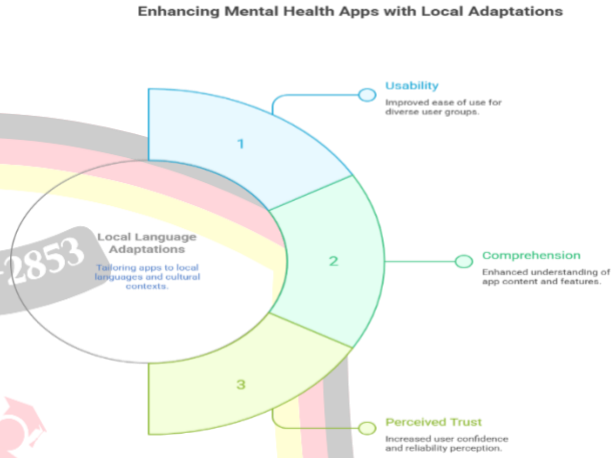


*Figure-1.Enhancing Mental Health Apps with Local Adaptations*

## KEYWORDS

***Mental Health Apps, Localization, Usability, System Usability Scale, Cultural Adaptation, Digital Health, India, Plain Language***

## INTRODUCTION

Smartphone-delivered self-help tools—mood trackers, cognitive restructuring exercises, mindfulness prompts, peer support, and safety planning—are increasingly recommended to augment mental health care and expand access where clinical resources are scarce. Yet the widely observed implementation gap is not only about internet connectivity or clinician endorsement; it is also about language. Millions of potential users encounter interfaces, consent forms, and therapeutic microcopy that are either untranslated or translated literally, without regard for dialect, register, or cultural resonance. In such conditions, even well-validated evidence-based techniques can become opaque, intimidating, or mistrusted.

Comparison of Mental Health App Versions

| Characteristic | English Baseline App | Adapted App |
|---|---|---|
| Usability (SUS) | Lower | Higher by 12.4 points |
| Time-on-task | Higher | Reduced by 22% |
| Comprehension of Privacy | Lower | Improved by 27% |
| User Perception | Clinical, less humane | Clearer, less clinical, more humane |
| Willingness to Use | Lower | Higher; recommend to family |

*Figure-2.Comparison of Mental Health App Versions*

Human–computer interaction (HCI) research distinguishes between translation (rendering words from one language to another) and localization (adapting content, symbols, measurements, and interactions to a target context). In mental health, adaptation further includes cultural and semantic fit—e.g., whether the metaphor for a "thought record" or "urge surfing" makes sense within users' lived experience, whether crisis pathways reference local helplines, and whether typography supports scripts with adequate legibility and line breaking. Despite strong theory and standards for cross-cultural adaptation of patient-reported outcomes, many consumer mental health apps still treat localization as an afterthought, often via machine translation or vendor glossaries detached from user research.

This paper reports a usability study evaluating whether local language adaptations meaningfully improve interaction quality for mental health self-help apps. We ask: (1) Do adapted interfaces yield higher usability and efficiency than an English baseline among multilingual users? (2) Are improvements concentrated in safety-critical and comprehension-heavy tasks? (3) How do users describe the experiential differences? By combining quantitative and qualitative evidence, we aim to move the localization conversation from "nice to have" to "non-negotiable" in digital mental health.

## LITERATURE REVIEW

**Digital mental health effectiveness and adoption:** Systematic reviews show that smartphone apps can modestly reduce symptoms of depression and anxiety and improve wellbeing for many users, with heterogeneity driven by engagement and design quality. Yet attrition remains high, and trust is often undermined by confusing language or opaque data practices. Models such as the Technology Acceptance Model (TAM) and UTAUT emphasize perceived usefulness, ease of use, and facilitating conditions; language clarity and cultural fit are direct levers on these constructs.

**Usability standards and scales:** Usability is commonly operationalized per ISO 9241-11 as effectiveness, efficiency, and satisfaction in a specified context. The System Usability Scale (SUS) offers a robust, technology-agnostic measure with established benchmarks, while instruments like PSSUQ and UMUX-Lite provide complementary perspectives. Heuristic frameworks (Nielsen's heuristics; ISO 9241-110) and design rules (e.g., consistency, match with real-world language, error prevention) offer practical checklists for pre-release evaluation.

**Localization and cross-cultural adaptation:** Cross-cultural adaptation guidelines for health measures emphasize forward–back translation, expert review, cognitive debriefing, and equivalence assessment (semantic, idiomatic, experiential, conceptual). Within HCI, internationalization/usability research highlights that direct translation often fails when domain terms are culturally bound or when scripts require different typography, truncation rules, or input methods. For mental health, additional sensitivity is needed: literal translations of clinical labels can stigmatize, metaphors may misfire, and safety content must reference local resources.

**Privacy comprehension:** Consent documents and privacy microcopy are frequent failure points. In low-literacy and multilingual settings, plain-language rewriting and layered disclosure (short summary + expandable details) improve comprehension and autonomy. For mental health, clarity

about what is shared, when, and with whom is central to trust and ongoing use.

**Gaps:** Most localization studies in health tech report process and linguistic outcomes rather than empirical usability impacts at the interface level. Few quantify changes in time-on-task, error rates, or comprehension for safety-critical content across multiple Indian languages and scripts. This study addresses that gap with a controlled, within-subjects design and a mixed-methods analysis.

## METHODOLOGY

### Design

We conducted a within-subjects, counterbalanced usability study comparing an English baseline app to a locally adapted version in each participant's primary language (Hindi, Bengali, Marathi, or Tamil). Each participant completed identical tasks in both versions, with order randomized to control for learning effects. The adapted version was not a mere translation: it included plain-language rewrites, dialect-aware choices validated via cognitive debriefing, culturally resonant examples (e.g., everyday stressors and idioms), script-appropriate typography (e.g., Devanagari and Tamil letter spacing and line height), text-to-speech prompts, localized helplines and clinic directories, and a layered privacy disclosure.

### Participants

A convenience sample of 208 adults (52 per language; 57% women; mean age 31.4 years, SD 8.9) was recruited from community centers and college campuses in four metropolitan regions. Inclusion criteria: age 18–60, daily smartphone use (>1 hour/day), self-reported proficiency in one of the four target languages, and no current suicidal intent (screened with a brief risk protocol and referral pathway). We stratified by self-rated English proficiency (low vs. moderate/high) to explore moderation effects.

### Materials and Tasks

Both app versions implemented common self-help components: onboarding and consent, baseline mood check-in, psychoeducation on cognitive distortions, a three-minute breathing exercise, a five-step cognitive restructuring worksheet, a safety planning template, and settings for data sharing. The task set involved:

1. **Onboarding & Consent:** Read a short privacy summary and answer three comprehension questions (purpose of data collection, data sharing with third parties, and how to delete data).
2. **Mood Log:** Record a mood with one-word label and intensity slider; add a brief note.
3. **Psychoeducation:** Open an article on "unhelpful thinking styles" and correctly identify an example of catastrophizing.
4. **Exercise:** Start and complete the breathing timer; locate the pause function.
5. **Cognitive Worksheet:** Reframe an automatic thought using guided prompts.
6. **Safety Plan:** Add two local contacts and the national/state helpline number; preview the "call now" flow.
7. **Settings:** Opt in/out of weekly reminders and export a data summary.

### Adaptation Process

We followed established cross-cultural adaptation steps: dual forward translation, reconciliation, back translation, expert committee review (including clinicians, translators, and lived-experience advisors), and cognitive debriefing interviews (n = 8–10 per language). We prioritized naturalness over literalness, replaced clinical jargon with plain alternatives, and introduced culturally familiar metaphors (e.g., "thought loop" analogized to "a song stuck on repeat"). Typography and interface spacing were tuned per script to avoid clipping and awkward line breaks. Consent language used layered disclosure, bullets, and icon-supported summaries.

**Measures**

- **Usability:** System Usability Scale (SUS; 0–100).
- **Effectiveness:** Task success (binary) and comprehension accuracy (three consent items).
- **Efficiency:** Time-on-task (seconds) and error counts (mis-taps, backtracks).
- **Perceived Burden:** Single-item mental effort rating (0–10).
- **Acceptability:** Willingness to continue using (5-point Likert) and to recommend to a family member (Yes/No).
- **Qualitative:** Think-aloud notes and post-session semi-structured interviews (~20 minutes), thematically coded.

**Procedure**

Participants attended a 60–75 minute lab session. After consent and risk screening, they completed the task set in Version A (English or Local, randomized), then Version B. Short breaks were provided to reduce fatigue. Sessions were screen-recorded. At the end, participants completed SUS for each version and a preference question. Compensation was a modest gift card and a printed resource sheet of local mental health services.

**Analysis**

We used paired t-tests for within-participant comparisons (SUS, time, errors), McNemar's test for binary outcomes, and mixed-effects regressions to examine moderation by English proficiency and language group (random intercepts for participant). Effect sizes were reported as Cohen's d for continuous outcomes and risk differences for proportions. Interview data were analyzed via reflexive thematic analysis with double-coding on 25% of transcripts to enhance credibility.

## RESULTS

**Quantitative Outcomes**

**Usability (SUS):** The adapted app achieved a mean SUS of 81.6 (SD 9.2) versus 69.2 (SD 10.7) for the English baseline; mean difference = +12.4 points (95% CI: 10.7–14.1), $t(207) = 15.41$, $p < .001$, $d = 0.77$. Improvements moved the experience from "OK–Good" to firmly "Excellent" on common SUS adjective ratings.

**Effectiveness:** Overall task success improved from 88.1% to 95.7% (risk difference +7.6%, 95% CI: 5.0–10.2, $p < .001$). Consent comprehension (three-item quiz) rose from 58.3% to 85.5% correct (Δ +27.2 percentage points, 95% CI: 22.4–32.0, $p < .001$). Safety plan completion (with correct local helpline) increased from 72.6% to 91.3% (Δ +18.7 pp, $p < .001$).

**Efficiency:** Median time-on-task decreased by 22% across tasks (IQR-adjusted), with the largest reductions in consent (−31%) and cognitive worksheet (−25%). Error counts (mis-taps/backtracks) fell by 34% ($p < .001$).

**Acceptability:** Willingness to continue using shifted upward (mean +0.8 on a 5-point scale, $p < .001$). Intention to recommend to a family member increased from 41% to 68% (McNemar $p < .001$).

**Moderation by English proficiency:** Participants with low English proficiency exhibited larger gains (SUS +16.9) than those with moderate/high proficiency (SUS +8.2), interaction $p < .01$. Gains were significant across all four languages; effect sizes were largest in Bengali and Tamil groups for consent comprehension and in Hindi and Marathi for safety planning.

**Qualitative Themes**

**1) "It speaks the way we do:"** Participants praised colloquial, non-clinical wording: "It felt like a friend explaining," contrasted with the English version's "textbook" tone. Culturally resonant examples (e.g., exam stress, family obligations, commute frustrations) helped users grasp cognitive distortion labels more easily.

**2) Lower cognitive load:** Users described the adapted microcopy as "lighter," "less mental work," and "no need to reread." This aligned with reduced time-on-task and error rates.

**3) Trust through clarity:** Layered consent with bullet points and icons made data practices "less scary." Naming local helplines, showing call charges, and clarifying that entries are private unless shared built confidence.

**4) Accessibility matters:** Text-to-speech and script-tuned typography aided users with lower literacy or eyesight strain, particularly on budget phones. A few noted that voice prompts should be optional to preserve privacy in shared spaces.

**5) Nuance and dialect:** Some words carried unintended connotations in specific dialects (e.g., terms for "worry," "anger"). Iterations that swapped these for neutral alternatives improved acceptance. Participants recommended offering two or three synonym options for certain mood labels.

### Adverse Events and Safety

No acute risk events occurred. Two participants requested referrals after the session; both were provided with local resources per protocol.

## Discussion

This study demonstrates that local language adaptations—when conducted as rigorous, user-centered redesign rather than literal translation—yield substantial usability gains for mental health self-help apps. The adapted interface improved effectiveness, efficiency, and satisfaction, with especially large effects on comprehension of consent and safety planning—two areas where misunderstanding can have ethical and safety consequences. Qualitative feedback corroborated the quantitative results: users perceived the adapted app as friendlier, clearer, and more trustworthy.

Our findings align with foundational usability principles: match between system and the real world; recognition rather than recall; and error prevention. They also echo cross-cultural adaptation science, which prioritizes conceptual over literal equivalence. The magnitude of improvement—SUS +12.4, 22% faster tasks, +27 pp consent comprehension—suggests that language and cultural alignment is not superficial polish but core product quality. In multilingual settings like India, treating localization as a post-hoc engineering step is likely to depress engagement and exacerbate inequities; treating it as primary design work can expand access and impact.

## Conclusion

Practically, we offer four takeaways for teams building or scaling mental health apps:

1. **Adopt a formal adaptation workflow:** Use forward–back translation, expert panels with clinicians and lived-experience advisors, and cognitive debriefing. Validate not only words but also metaphors, examples, and tone.
2. **Invest in script-appropriate UI:** Tune typography, line height, truncation behavior, and input methods per script. Verify on low-end devices.
3. **Localize safety and trust touchpoints:** Provide layered, plain-language consent and local helplines; make emergency flows one-tap and place disclaimers responsibly.
4. **Measure what matters:** Don't stop at linguistic QA. Run usability tests with time-on-task, comprehension checks, and SUS/UMUX-Lite. Track where language changes reduce errors in sensitive flows.

In short, local language adaptation is an accessibility and equity intervention that measurably improves user experience and ethical quality in digital mental health.

## Scope and Limitations

**Scope**

The study evaluates short-term usability outcomes for a general-audience self-help app across four widely used Indian languages and scripts. It covers core self-help features (mood logs, psychoeducation, breathing, cognitive reframing, safety planning) and key trust touchpoints (consent, data settings). The adaptation employed a structured, replicable process with both linguistic and HCI adjustments.

**Limitations**

First, our sample, though diverse, was urban and smartphone-accustomed; results may differ in rural contexts, among older adults, or for populations with very low literacy. Second, we examined immediate task performance and perceived usability, not longitudinal engagement or clinical outcomes; whether improvements persist and translate into symptom change requires follow-up trials. Third, we compared a single English baseline to one adapted variant per language; alternative adaptation choices could yield different effects. Fourth, we studied four languages; findings may not generalize to others, especially those with distinct scripts (e.g., Urdu's Nastaliq layout) or right-to-left text. Fifth, because participants completed both versions in one session, residual learning effects are possible despite counterbalancing.

Future work should include multi-month field deployments, adaptive personalization of dialect registers, A/B tests of consent and safety flows, and rigorous evaluation of clinical endpoints. Collaborations with public health agencies can help integrate localized apps with community services and helplines.

## REFERENCES


- *Baumel, A., Muench, F., Edan, S., & Kane, J. M. (2019). Objective user engagement with mental health apps: Systematic review. Internet Interventions, 18, 100286.*
- *Bangor, A., Kortum, P. T., & Miller, J. T. (2008). An empirical evaluation of the System Usability Scale. International Journal of Human–Computer Interaction, 24(6), 574–594.*
- *Beaton, D. E., Bombardier, C., Guillemin, F., & Ferraz, M. B. (2000). Guidelines for the process of cross-cultural adaptation of self-report measures. Spine, 25(24), 3186–3191.*
- *Brooke, J. (1996). SUS: A "quick and dirty" usability scale. In P. W. Jordan, B. Thomas, I. L. McClelland, & B. Weerdmeester (Eds.), Usability evaluation in industry (pp. 189–194). Taylor & Francis.*
- *Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS Quarterly, 13(3), 319–340.*
- *Donker, T., Petrie, K., Proudfoot, J., Clarke, J., Birch, M.-R., & Christensen, H. (2013). Smartphones for smarter delivery of mental health programs: A systematic review. Journal of Medical Internet Research, 15(11), e247.*
- *Eysenbach, G. (2011). CONSORT-EHEALTH: Improving and standardizing evaluation reports of Web-based and mobile health interventions. Journal of Medical Internet Research, 13(4), e126.*
- *Firth, J., Torous, J., Nicholas, J., Carney, R., Pratap, A., Rosenbaum, S., & Sarris, J. (2017). The efficacy of smartphone-based mental health interventions: A meta-analysis of randomized controlled trials. World Psychiatry, 16(3), 287–298.*
- *Hall, J. L., & McGraw, D. (2014). For telehealth to succeed, privacy and security are needed. Health Affairs, 33(2), 216–221.*
- *ISO 9241-11. (2018). Ergonomics of human–system interaction—Part 11: Usability: Definitions and concepts. International Organization for Standardization.*
- *ISO 9241-110. (2020). Ergonomics of human–system interaction—Part 110: Interaction principles. International Organization for Standardization.*
- *Lewis, J. R. (1995). IBM computer usability satisfaction questionnaires: Psychometric evaluation and instructions for use. International Journal of Human–Computer Interaction, 7(1), 57–78.*
- *Lewis, J. R., & Sauro, J. (2009). The factor structure of the System Usability Scale. International Conference on Human Centered Design, 94–103.*
- *Nielsen, J. (1994). Heuristic evaluation. In J. Nielsen & R. L. Mack (Eds.), Usability inspection methods (pp. 25–62). John Wiley & Sons.*
- *Norman, D. A. (2013). The design of everyday things (Revised and expanded ed.). Basic Books.*
- *Torous, J., & Wykes, T. (2019). Opportunities and challenges of mental health apps. World Psychiatry, 18(1), 76–82.*
- *van Gemert-Pijnen, J. E. W. C., Nijland, N., van Limburg, M., Ossebaard, H. C., Kelders, S. M., Eysenbach, G., & Seydel, E. R. (2011). A holistic framework to improve the uptake and impact of eHealth technologies. Journal of Medical Internet Research, 13(4), e111.*
- *Venkatesh, V., Morris, M. G., Davis, G. B., & Davis, F. D. (2003). User acceptance of information technology: Toward a unified view. MIS Quarterly, 27(3), 425–478.*
- *WHO. (2021). Comprehensive mental health action plan 2013–2030. World Health Organization.*

- *WHO. (2022). World mental health report: Transforming mental health for all. World Health Organization.*
- *Wild, D., Grove, A., Martin, M., Eremenco, S., McElroy, S., Verjee-Lorenz, A., & Erikson, P. (2005). Principles of good practice for the translation and cultural adaptation process for patient-reported outcomes measures. Value in Health, 8(2), 94–104.*