<u>Methods used for prototype evaluation:</u>

1.) System Usability Scale (SUS) - General Standard
2.) Mixed Methods Survey - System Specific Questions

Both evaluations are conducted in a controlled setting.

<u>System Usability Scale (SUS)</u>

Overall Post Usage Usability Evaluation was conducted using the standard System Usability Scale (SUS) Questionnaire:

All items were answered on a 1 (Strongly disagree) -5 (Strongly agree) Likert scale.

The items on the questionnaire are as follows:

1. I think that I would like to use this system frequently

2. I found the system unnecessarily complex

3. I thought the system was easy to use

4. I think that I would need the support of a technical person to be able to use this system

5. I found the various functions in the system were well integrated

6. I thought there was too much inconsistency in this system

7. I would imagine that most people would learn to use this system very quickly

8. I found the system very cumbersome to use

9. I felt very confident using the system

10. I needed to learn a lot of things before I could get going with this system

<u>Scoring SUS</u>

SUS yields a single number representing a composite measure of the overall usability of the system being studied. To calculate the SUS score, we first sum the score contributions from each item. Each item's score contribution will range from 0 to 4. For items 1,3,5,7,and 9 the score contribution is the scale position minus 1. For items 2,4,6,8 and 10, the contribution is 5 minus the scale position. Then we multiply the sum of the scores by 2.5 to obtain the overall value of SU. SUS scores have a range of 0 to 100. Goal is to get 80+

Mixed Methods Survey - System Specific Questions

The next set of questions evaluates more system-specific requirements for our prototype. This includes some quantitative questions as well as some open-ended interview questions to gauge maximum user experience and feedback regarding the high-fidelity prototype. The high-fidelity prototype is based on features that we extracted based on participant responses from our user beliefs and requirements surveys during Project 2. The features correspond to 4 key themes:

Theme	Corresponding Features
The Privacy Concern	-Clear control over the ability to delete recordings, opt out of cloud storage, and understand exactly what data is collected.
Distrust of AI Accuracy	-Transparency demonstration of how the system interprets their voice. - Pattern Visualization Over Time: Emotional trends across weeks or months, not single point assessments. Integration with Campus Support: User-initiated, optional connection to university mental health services.
Fear of Stigma and Labeling	-Non-judgmental Tone: careful framing of feedback, wellness language rather than diagnostic terminology. -Low-Stigma Framing: Alternative framings (e.g., "mood journaling" vs. "depression screening") significantly impact willingness to engage.
Need for Different Engagement Styles	-Multilingual support -Flexible Engagement Options: Some prefer scheduled assessments, others want daily lightweight check-ins, and some need as-needed access only. -Various clear options to choose from to visualise their progress/past status. -Ability to set their own notification and reminder frequencies. No consistent reminders from the app.

Table 1- Summary of themes and corresponding features that were used to develop the high fidelity prototype.

The prototype-specific questions gauge feedback on these four themes and corresponding features. The questionnaire for the same is as follows:

Section 1: Basic Demographics
1. Age Group

- 18–20
- 21–24
- 25–30
- Prefer not to say

2. Gender
   - Woman
   - Man
   - Non-binary / Gender diverse
   - Prefer not to say

3. Primary Language(s) Spoken at Home
   (Select all that apply)

   - English
   - Hindi
   - Other (please specify): _____
   - Prefer not to say

Section 2: Theme 1- Privacy & Data Control
Likert Questions
1. The app clearly communicated what data it was collecting.

2. I understood how my speech data would be stored, deleted, or used.

3. I felt confident that I could control my privacy settings (e.g., delete recordings, opt out of cloud storage).

4. The app's privacy explanations reduced my concerns about data misuse.

Open-ended
5. What aspects of the app made you feel more (or less) in control of your data?

6. Is there anything the app should explain more clearly about privacy?

Section 3: Theme 2- AI Transparency & Accuracy Concerns
Likert Questions
1. The app explained how AI analyses my speech in a way I could understand.

2. The feedback visualizations (patterns over time) helped me trust the system.

3. I felt the AI's interpretations were reasonable based on my interaction.

4. I understood that the app is not diagnosing me but offering wellness insights.

Open-ended
5. Was there anything about the AI explanation that you found confusing or unclear?

6. What would increase your trust in the app's AI analysis?

Section 4: Theme 3- Fear of Stigma & Tone Sensitivity
Likert Questions
1. The language used by the app felt non-judgmental and supportive.

2. The app framed mood and mental health in a way that reduced stigma.

3. I did not feel labeled or judged by the feedback provided.

Open-ended
4. How did the tone or wording of the app affect your willingness to use it regularly?

5. Describe any points where the app felt stigmatizing, uncomfortable, or overly clinical.

Section 5: Theme 4- Personalization & Engagement Styles
Likert Questions
1. The app's language options made it accessible to me.

2. I found the different engagement styles (scheduled, daily check-ins, on-demand) useful.

3. The app allowed me to tailor notifications/reminders to my preferences.

4. The visualization options (charts, summaries) were helpful in understanding my emotional progress.

Open-ended
5. Which engagement style did you prefer and why?

6. How could the app better adapt to your personal needs or usage habits?

Section 6: Additional Usability & Missing Elements (Beyond SUS)
(These address aspects not covered by SUS, such as emotion, comfort, trust, relevance, and feature completeness.)
Likert Questions
1. I would feel comfortable using this app in a real-world setting (e.g., daily life).

2. I would be willing to use this app over a longer period (weeks or months).

Open-ended
3. What important features or options do you think are missing from the prototype?

4. What was the most helpful and least helpful part of the app for you?

5. If you could change one thing in the app, what would it be?

Section 7: Overall Evaluation
Likert Questions
1. Overall, the app met my expectations.

2. Overall, I would recommend this app to someone looking to track their mental wellness.

Open-ended
3. Please share any final feedback, concerns, or suggestions you have.