Goals:

1. ~~Evaluate the consistency of the app compared to other apps in the market from the user's viewpoint.~~
2. ~~Evaluate the effectiveness of metaphor in helping user judging their work-life balance~~
3. Evaluate the effectiveness of adding and scheduling the task from the user's viewpoint.
4. Evaluate how interesting the challenge is from the user's viewpoint.
5. ~~Evaluate the effectiveness of task breakdown into attainable goals from the user's viewpoint.~~
6. Evaluate how intuitive work-life balance graphs are to understand from the user's viewpoint.

Questions:

Goal 3:

(a) How clear is the difference between adding and scheduling tasks?
(b) What is the current task load?
(c) What is the total time taken to complete both actions (adding and scheduling tasks)?
(d) What is the reaction of the user after receiving the pop-up alert?
(e) How would task breakdown be performed?

Goal 4:

(a) Would the challenge be actually performed in reality?
(b) How often would users want to join a challenge?

~~Goal 5:~~
~~(a) What is the reaction of the user after receiving the pop-up alert?~~
~~(b) How would task breakdown be performed?~~

Goal 6:
(a) Could the users clearly understand their work-life balance from the graphs?

Metrics:

Goal 3:

1. Likert scale: self-evaluate how clear it is
2. NASA task load scale
3. Time taken to perform tasks
4. % of times users perform task breakdown after receiving the pop-up alert?
5. Qualitative response from interviewing how users would perform or feel about task breakdown.

Goal 4:

1. Qualitative response from interviewing users their reasons for joining the challenge.
2. How often users want to join a challenge (scale: daily, few times per week, weekly, few times per month, monthly, once in a while, never).

Goal 6:

1. Qualitative response from interviewing users by asking them to list 2-3 conclusions from the work-life balance graph.

Persona:

Methods:
1. Think aloud
2. Interviews
3. Survey questions (Likert scale and NASA task load scale)

Task List

We ordered the task based on frequency and criticality, with 1 being our top priority and 4 being our lowest priority. We decided that it is necessary to test all tasks.

No.	Task	Conditions
1	Assess current balance between work and life	Prerequisite: At home page Completion: Verbal answer given to the facilitator that shows their understanding of the meaning of the figure and graph. Time limit: 1:30 min Error: 1. Cannot navigate to balance page 2. Misunderstood or does not understand meaning of figure and graph
2	Add a [work-related task](#)	Prerequisite: At home page Completion: Task appears on task list Time limit: 2:30 min Error: 1. Cannot navigate to task creation 2. Task not added even after hints are given
3	Schedule an event named "HCI Meeting Preparation" in the calendar to perform the tasks "Find palette HCI" and "HCI prototypes"	Prerequisite: At home page Completion: Event appear in calendar Time limit: 1:00 min Error: 1. Cannot navigate to calendar page 2. Event not scheduled even after hints are given
4	Join and complete the "Try dessert" challenge, and review history of challenges	Prerequisite: At home page Completion: Complete the dessert challenge Time limit: 1:00 min Error: 1. Cannot navigate to challenge page 2. Did not join and complete the challenge

Test Environment:
In an empty classroom

Data to be collected:
1. Likert scale score:
    (a) 5-point likert scale for users to self-evaluate how clear is the difference between adding and scheduling tasks

    (b) 7-point likert scale for users to rate how often they would want to join a challenge
2. NASA task load scale score
3. Time taken to perform task 2 and 3
4. Frequency of users performing task breakdown after receiving the pop-up alert
5. Qualitative response:
    (a) Reasons why users join a challenge
    (b) How users would perform task breakdown
    (c) 2-3 conclusions of work-life balance graph

Content of Report:
1. Bar chart for Likert scale
2. Bar chart for NASA task load scale
3. Average time taken to perform task 2 and 3
4. % of times users perform task breakdown after receiving the pop-up alert
5. Users' responses from interview
6. Users' reactions (think-aloud) while performing tasks

Script

Step	Content
Preparation script	Environment: In an empty classroom
	Equipment: Laptop
	Roles:
	User 1: Punn (Facilitator) Zhi Lin (Observer)
	User 2: Zhi Lin (Facilitator) Punn(Observer)
	User 3: Punn (Facilitator) Zhi Lin (Observer)
	Users: We recruited participants that fit the target user description.
Orientation script / Informed consent	"Hi, my name is [Facilitator's name]. I'll be working with you in today's session. [Observer's name] here will be observing. We're here to test Honey, an application tailored specifically to support undergraduate students managing the demanding balance between academics and

internships while striving to regain equilibrium in their work-life dynamic, and we'd like your help.

I will ask you to perform some typical tasks with the system. Do your best, but don't be overly concerned with results – the system is being tested, and not your performance.

Since the system is a prototype for a class project, there are certainly numerous rough edges and bugs and things may not work exactly as you expect.

My only role here today is to discover the flaws and advantages of this new system from your perspective. Don't act or say things based on what you think I might want to see or hear, I need to know what you really think.

Please do ask questions at any time, but I may only answer them at the end of the session. While you are working, I will be taking some notes. We will also be videotaping the session for the benefit of people who couldn't be here today.

If you feel uncomfortable, you may stop the test at any time. Do you have any questions? If not, then let's begin by getting to know more about you and having you sign the nondisclosure agreement and consent to tape form.

Consent form: https://forms.gle/R6nQxC3hD2hnkwGy8

Training script	In this test, we want you to use a method called "think aloud". To perform this, I want you to imagine that I am a blind person and you are describing what you see. We're going to show you a demo video right now. Please keep in mind that this is intended to be a smartphone interface and that we are not testing *you;* we are testing the design of the interface. It is normal for users to not solve the task on the first try. We encourage you to not feel pressure to give the right answer but rather do what you would normally do when you download a new app.
Task scenarios	1. Assess current balance between work and life 2. Add a work-related task 3. Schedule an event named "HCI Meeting Preparation" in the calendar

	to perform the tasks "Find palette HCI" and "HCI prototypes" 4. Join and complete the "Try dessert" challenge, and review history of challenges
Data collection form	https://docs.google.com/document/d/1-ui-mYJN41bXqC6Y44iSvrtcxtg__uuEeGAGtMVU2nQ/edit?usp=sharing
Debriefing	Thank you, you have now finished all the tasks? How was it? Now that you have tried our user interface, I would love to collect some information about your feelings, suggestions and opinions. Would that be okay for you? ● Interface organization matches real-world tasks? ● Too much or too little information on screens? ● Similar information consistently placed? ● Problems with navigation? ● Computer jargon? ● Appropriate use of color? Please fill in this survey too. Thank you again for participating in our user study. Here's some chocolate for you.

Checklist
1. Preparation:
   - ❏ Reset interface for new user
   - ❏ Check that everything is ready in test room
   - ❏ Establish protocol for any observers
2. Opening:
   - ❏ Greet the participant
   - ❏ Go through orientation script and set the stage
   - ❏ Ask user to read and sign consent and non-disclosure forms
3. Test Session:
   - ❏ Move over to testing area (computer)
   - ❏ Start session recording
   - ❏ Provide any prior training
   - ❏ Provide training of thinking aloud
   - ❏ User begins with tasks
   - ❏ During task 1, ask them "Can you list 2-3 conclusions based on what you understood from the graphs?"
   - ❏ When the user finishes task 2, ask them: "What do you feel about the task breakdown? Does the alert help? What do you think makes you breakdown task?"
   - ❏ When the user finishes task 4, ask them: "In real-life situations, would you use this feature? Why? Do you think this method works to encourage workaholic to break down the task? Do you think challenges should be recommended such that it is specific to you or random?"
   - ❏ User finishes last task
4. Closing:
   - ❏ Interview: how was it?
   - ❏ Structured interview questions
   - ❏ Individual interview questions arising from test
   - ❏ Thank participant, provide any remuneration, show participant out
5. Clean-Up:
   - ❏ Summarize thoughts about this test
   - ❏ Organize data sheets and notes
   - ❏ Check and archive session recordings

Usability Issues
Task 1:
- Struggle to search for work and life balance [HIGH] (U1, U5)
- Confused because the division line of the two colors is in the middle but the scale is tilted [MEDIUM] (U2) ZL
- The smiley face is not a clear metaphor [HIGH] (U3, U4, U5) ZL
- Did not understand what each bar represents [HIGH] (U1) Punn ZL
- Similar to home page so would access home page more often than balance page. [HIGH] (U3) ZL

Task 2:
- Finds alert pop-up annoying [MEDIUM] (U1) Punn ZL
- Would only breakdown task if the task can be modularized [LOW] (U2)

Task 3:
- Did not understand and found it redundant to add an existing task to the event. [HIGH] (U5) ZL
- Inconsistent layout makes order of completing the task confused [HIGH] (U4, U5)

Task 4:
Would prefer randomized challenge (U2)
Would prefer challenges being recommended following YouTube's recommendation system (U1)
Would prefer a mixture of randomized challenges (when she feels like challenging herself) and challenges she enjoys (U3)
- => Every participant has a different preference on the type of challenges recommended to them. [HIGH] (U1, U2, U3)   ZL Punn
- Wants a reward system [LOW] (U3) ZL
- Misinterpreted that we can make our own challenge [HIGH] (U2)ZL Punn

Overall:
- Wants to customize color that represent work and life [HIGH] (U1, U2, U3) Punn ZL
- Was slightly confused when navigating to home because normally applications has their home button at the middle [LOW] (U1)
- Font should be smaller [LOW] (U1)

- Likert scale score:
    - (c) 5-point likert scale for users to self-evaluate how clear is the difference between adding and scheduling tasks

    - (d) 7-point likert scale for users to rate how often they would want to join a challenge
- NASA task load scale score
- Time taken to perform task 2 and 3
- Frequency of users performing task breakdown after receiving the pop-up alert
  ⅖ performed task breakdown; 2 did not;; 1 performed task breakdown even before alert
- Qualitative response:
    - (d) Reasons why users join a challenge
        - User wants to join the challenge at least once a week, because this will increase his life-activities time spending
        - Will not join in real life, since it is just suggesting something for you to do when you have nothing to do
        - She would join because it seems interesting.
        - She would join because she does not want to have an uncompleted challenge left. She mentioned that it'll be motivating for her to join if there is some sort of reward system.

    - (e) How users would perform task breakdown
        - User decided to create a challenge requiring a long time of completion rather than creating three separate challenges, but will not complete the task at once at will manage the duration of the working sessions by himself
        - Break it down because they want to gradually finish it before the deadline.
        - Break the task down only if it can be modularized.

    - (f) 2-3 conclusions of work-life balance graph
        Users want to see the analytics from the graph. E.g. best days, worst days.
        Graph is not fully intuitive as the help of facilitators is needed. We need to include more legend / description around the graph.

Q: Your application makes extensive use of Figma functionalities. Were there any constraints imposed by Figma that might limit certain features?
A:
It's difficult to let users break down the tasks to the ones they want because Figma does not allow that kind of flexibility. Instead, we have to let them imagine it. It might be difficult for the users to relate with our work-life balance graphs because it is not their real data. We couldn't use keyboard / emoji selection functionality as in the real app, since we hard-coded the interface.

Q: You have tested many users; What kind of differences among users did you notice along the experiment?
A: Everybody has different preferences regarding what kind of challenges they prefer (random/ specific). Furthermore, some understood the metaphors and some did not. People may like or dislike certain interactions / elements based on their personal preferences (e.g. emojis on the tasks, challenges category, display of balnaces).

Q: I'm curious what the criteria was for selecting participants. What questions did you ask in the interviews after the user test?
A: We selected undergraduate students who are overwhelmed with work, working (lab, internship, etc) while also striving to regain work-life balance. Some of our users were not working, but since they are overwhelmed with their academic life, we thought that it would be useful to test our applications on them too.

Q: It was impressive that you conducted offline testing for 5 people. Is there any method you recruited them? And was there any difficulty when conducting offline testing for many users?
A: No, we recruited our friends, which fits the characteristics of our persona and knows nothing about our application

Q: Uichin: clarification - you have two parts: challenges vs. tasks. Here, tasks are clear, but what are challenges here?
A: Since we target work-life balance, challenges greatly contribute to the enrichment of life metrics. They allow users to explore new hobbies, activities, and passions. Challenge is a task that is predefined by our team that the user needs to complete and provide evidence of completion.

Q: Uichin - I noticed that most students are selecting metrics that can be quantitatively measured. But you can do the interview and find out qualitative results (e.g., themes in interviews, errors found during observations)..
A: We performed a qualitative interview at the end of the session asking the user about some of their subjective personal preferences. The conclusion of the said interviews is in the report and also the slides!