Usability Test Plan

Testing Overview

This document describes how we will evaluate the prototype of our public transport application built in Figma. Our goals are to observe how first time users perform 5 key tasks. This is to uncover any points of confusion or error, and to measure overall satisfaction. We will gather quantitative data such as task completion times and error counts alongside qualitative feedback recorded during testing and rating from a workload assessment form.

Methodology

Methodologies used are usability testing and a Survey. Testers will be given 5 tasks to perform on the prototype, one by one, while moderators record data. Usability tests are done on the prototype, having users think out their process of completing the task and explaining to the facilitator, while the time keeper keeps track of how long it takes, and note takers write down how the testers perform the task. Once the usability test is complete, a NASA TLX survey is then used to record feedback on the prototype.

We will run a moderated in-person session with six members of another group from SWEN303 at Victoria University of Wellington. The session will take place in a quiet room on campus. The participants will be tested individually, bringing each participant one by one, after each usability test is done.

At the start of the session, the facilitators will welcome the participants, explain the purpose of the test, and provide a short introduction to the prototype. They will clearly communicate the first task, ensuring the participants understand what to do.

Each task will be performed on Figma with hints disabled, so users don't get shown the location of all buttons on the screen when they misclick. This setting is available from the settings button in the top-right corner and then going to Advanced settings → Show hints on click. Additionally, the Figma UI will be hidden (by pressing Ctrl\) and the web browser will be put into fullscreen mode to minimise distractions.

While participants work through the task. Timekeepers will start their timers and track how long each participant takes to complete the task. Meanwhile, the note takers will closely observe the participants' behavior. Documenting key moments such as where users hesitate, express confusion, or show signs of frustration or satisfaction. If a participant becomes clearly stuck or expresses that they do not know how to proceed. The facilitator will intervene—but only to help them continue, not to give away solutions. This ensures that usability problems are identified while still allowing participants to complete the task.

Once a participant finishes a task, the timekeeper will record the completion time, and the group will move on to the next task. After all tasks have been completed. The facilitators may conduct a short debrief or ask follow-up questions to gather participants' impressions, feedback, and suggestions.

Throughout the test, the observers and note takers will work together to capture both quantitative data (task times) and qualitative data (behaviors, comments, errors), ensuring a well-rounded set of findings for later analysis.

Script

Facilitator's Opening:

"Hi, thank you for joining us today! We're going to test a prototype of a transport app we've been working on. We're not testing you — we're testing the design, so don't worry about doing anything 'right' or 'wrong.' Your honest reactions, thoughts, and feedback will help us improve the design."

Setting Expectations:

"We'll give you a series of tasks to complete using the prototype. Please think out loud as you go — tell us what you're doing, what you're looking for, and what you're thinking, even if it feels obvious. This will help us understand how you interact with the design."

Explaining Moderator Roles:

"During the session, some of us will act as facilitators, some will take notes, and some will be measuring time. We may ask you questions about your decisions, and we'll be writing down how long tasks take and what kinds of actions you make."

Intervention Rules:

"If you get stuck, you're welcome to ask for help. We'll step in to help if you ask or if you go significantly over time on a task — but we'll generally let you try to work through things on your own so we can see how the design supports you."

Tasks Overview:

"We'll go through 5 tasks today, such as:

1. Find next departure time table for the Hutt Valley Line train service
2. Making a route from 200 Cuba St. to Wellington Station
3. Read a notice about Hutt Valley Line
4. Open a saved route in the home page
5. Change the colour theme of the app to dark mode/contrast mode.

We'll move on to the next task when you complete the current one or if you feel you can't finish it. After all tasks are done, we may ask you a few quick follow-up questions."

Closing the Intro:

"Do you have any questions before we begin? If not, we'll get started with the first task."

The tester will then proceed to ask the user to complete the tasks detailed below.

Roles

Person	Role
	Facilitator
	Facilitator
	Note taker
	Note taker

	Timekeeper
	Timekeeper

Facilitator:

- Guides the participant through the tasks and explains the overall process.
- Ensures the participant understands the consent procedure before starting.
- Delivers tasks clearly, monitors participant progress, and provides hints when allowed (either when the participant asks or when time thresholds are reached).
- Handles any urgent technical issues that arise during the session.
- Conducts the debrief at the end of the session, including follow-up questions or surveys.
- Collects all completed materials (notes, timing sheets, survey responses) and ensures they are stored securely.

Note Taker:

- Carefully observes the participant's actions, comments, and body language throughout the session.
- Writes down key observations, such as where the participant hesitates, gets confused, or expresses opinions.
- Records participant quotes and specific behaviors that could highlight usability issues.
- Ensures no detail is missed by collaborating with the facilitator and timekeeper during breaks or after the session.

Timekeeper/Observer:

- Starts and stops the timer for each task, accurately recording how long each task takes.
- Monitors time thresholds and signals to the facilitator when it's time to intervene or move on.
- Acts as a secondary observer, helping note takers by watching for patterns, repeated mistakes, or behaviors that might be missed.
- Supports overall session flow by keeping track of time for the entire session, ensuring it stays on schedule.

Participant

- Starts each task together with other participants, organised by the facilitator. Completes the tasks individually.
- Fills out NASA TLX survey for qualitative data.

Tasks

1. Find next departure time for the Hutt Valley Line train service (less than 40 seconds ± 15 seconds)
2. Find a notice about Hutt Valley Line → return to home (less than 20 seconds ± 10 seconds)
3. Making a route from 200 Cuba St. to Wellington Station (less than 1 minute ± 30 seconds)
4. Open a saved route in the homepage (less than 30 seconds ± 15 seconds)

5. Change the colour theme of the app to dark mode. (less than 1 minute ± 30 seconds)

Metrics: NASA TLX

After researching various methods to use for our testing, we decided to use NASA TLX, though SUS was also an alternative. Taking NASA TLX as our choice was due to the reliable results it yields, it has been widely used for many sorts of industries, and due to its human perspective focus, it becomes especially useful as we can evaluate our system on how people perceive it.
We have additionally modified the NASA TLX slightly to use a scale from 0-10 rather than 1-21 for simplicity when testers fill out the survey.
https://humansystems.arc.nasa.gov/groups/tlx/downloads/TLXScale.pdf

Usability Goals

Our usability goals are going to be based on the NASA TLX. By incorporating a multi-dimensional rating procedure we can derive an overall workstore
Ideally, we overall aim for higher than average on all 7 points of measurement in metrics. Easier tasks such as finding timetables and notices should have lower tolerances to score than harder tasks such as planning a route.

Score we aim for per task:

Task	Mental Demand	Physical Demand	Temporal Demand	Performance	Effort	Frustration
Finding timetable	5 ±3	2 ±3	4 ±3	3 ±3	4 ±3	3 ±3
Finding notice	6 ±3	2 ±3	3 ±3	4 ±3	5 ±3	4 ±3
Making route	15 ±3	3 ±3	14 ±3	7 ±3	16 ±3	12 ±3
Opening a saved route	8 ±3	6 ±3	5 ±3	10 ±3	6 ±3	8 ±3
Changing Theme	4 ±3	1 ±3	3 ±3	2 ±3	3 ±3	2 ±3

Score we aim for on average:

Category	Average
Mental Demand	7.6
Physical Demand	2.8
Temporal Demand	5.8
Performance	5.2
Effort	6.8
Frustration	5.8

Problem Severity

Each issue will have a severity rating from one to four:

Severity 1: Cosmetic/Minimal

Being able to finish the task that has been assigned with ease.
Example: Navigate through the settings after carefully reading the description for each setting.

Severity 2: Minor

A discrepancy that causes slight hesitation or extra time but does not prevent users from finishing the task
Example: Navigate through the settings after making a few mistakes. By navigating to the wrong page once or twice before going to the correct one.

Severity 3: Major

A problem that interferes with task flow and requires the user to ask for help, seek hints or repeat steps.
Example: Not having a return button to go to a previous screen.

Severity 4: Catastrophic

A blocker that prevents task completion entirely and leads the user to give up or abandon the task.
Example: Task user flow is incomplete, and tasks can not be finished.