

## User Evaluation Report

Cohort 1, Group 9

Chris Sewell, Fedor Kurochkin, Matt Durham, Max Peterson, Oladapo Olaniran, Wojtek Tomaszewski, Yuqi Fu.

A task based 1:1 user evaluation was conducted in a lab environment involving live game test sessions as a means to identify and confirm user problems with the maze game prototype.

The purpose of task-based usability testing is to assess whether intended users are able to complete the system's designed tasks effectively and whether their experience using the prototype is satisfactory (Sharp et al., 2023).

This task-based method was employed as it best serves the project size and goals. The prototype's events and known problems could be compartmentalised into specific timed tasks and tested individually by users without the need for counterbalancing. The evaluations were conducted 1:1 in a structured interview format by an observer in a lab based distraction free environment. This was to avoid external influences on the participants and to prevent biases such as the order effect.

six participants were recruited satisfying the minimum number of users necessary to attain results according to Dumas and Redish (1999).

Participants were primarily recruited from within the Engineering module for convenience and accessibility but a few 'hallway participants' were also recruited due to time constraints and availability close to the holiday period.

Prior to each user test, participants were sent links to consent forms and information sheets outlining the evaluation including confirmation that no audio or video recording would take place.

[See evaluation section for these documents provided.](#)

The majority of evaluations were conducted in person, with one user test done remotely via Discord due to availability.

Prior to user testing, evaluation scripts were made based on team player tests to predetermine problems with the prototype and predict user behaviour in advance. To be conducted alongside these scripts, a structured usability testing template for consistent data recording was adapted from an industry resource (Doyle, 2020).

[See User Evaluation Data page.](#)

Each user test lasted 20-30 minutes long in total to maintain participant concentration and avoid frustration or boredom. The prototype game was tested on a laptop with minimal hardware requirements (users could use their own if they preferred remotely).

To optimise the data gained from the limited recruitment pool, the evaluation employed both qualitative and quantitative data collection methods:

Observations record the number of times participants missed key information or encountered a known problem in a certain amount of time etc.

Participants were encouraged to 'think aloud' in order to track user behaviour whilst engaging with tasks. The lab setting combined with the distraction of the task means users provide an accurate real time thought process via this technique.

The scripts and severity report conducted also include post task feedback discussion sections for qualitative user data gathering.

After the tests were completed and the data compiled in the evaluation template, findings were ascertained and problems identified. The severity ratings for these problems were compiled into the table below.

*Figure 1: Table summarising usability problems identified via the task based user evaluation along with the number of participants that assigned each severity rating. Severity ratings were gathered via a post evaluation [severity report](#).*

Usability Problem	Severity Rating				
	Not a problem or (N/A)	Cosmetic	Minor	Major	Catastrophic
Attendance system balancing and ambiguity	1	1	3	1	0
Geese event ambiguity and lack of challenge	0	0	4	2	0
Coffee points ambiguity	4	1	1	0	0
Hidden event pressure plate lack of visibility	0	3	2	1	0
Hidden event Bob ambiguity	3	0	2	1	0

Severity rating is categorised as:

**N/A** (the user didn't consider it a problem or was indifferent)

**Cosmetic** (not thematic or unclear as to what it is, vague or misleading etc)

**Minor** (minor inconvenience or annoyance)

**Major** (persistent problem that reduces engagement/ causes frustration)

**Catastrophic** (major bug or annoyance, prevents user from completing a task)

Following the iterative design process, the findings obtained from the user evaluations and severity reports allowed the team to identify new user, functional and non-functional requirements and to develop updated architectural diagrams to address existing problem areas and support the implementation of new features satisfying user feedback.

## References:

Doyle, S. (2020) '*A template for Usability Testing*', Shane Doyle [Blog], 12 March. Available at:  
<https://web.archive.org/web/20250326045926/https://www.shanedoyle.io/post/a-template-for-usability-testing> (Accessed: 20 December 2025).

Dumas, J.S. and Redish, J.C. (1999) *A practical guide to usability testing*. Revised edn. Exeter: Intellect.

Sharp, H., Rogers, Y. and Preece, J. (2023) *Interaction design: beyond human-computer interaction*. 6th edn. Hoboken, NJ: John Wiley.

