

Evaluation

Cohort 3 - Group 4

AJAJARA

Abbie Spears

Adam Fuller

Alexander Gibson

Artem Frolov

James Goude

James Sheeley

Rosie Sherwood

User Evaluation Report

Part A: Method for User Evaluation

Participants recruited for the user evaluation were peers within the cohort - external to the development team to avoid bias. This group was chosen for three main reasons: the convenience of 'Hallway Recruiting', their familiarity with the context of the project, and their close resemblance to the user group (young people attending one of the University of York's Open Days and interested in pursuing a degree in Computer Science or related subject).

Each member of the team recruited one participant, resulting in a total of 6 users. This number was deemed sufficient following Jacob Nielsen's guideline of testing five users to reach saturation in qualitative research[1]. The decision for each member of the team to engage one participant was chosen to maximise time efficiency and ensure fairness.

The team created a [User Evaluation Protocol](#), which was followed during each User Test. The aim was to evaluate the software's usability and gather actionable feedback for improvement.

After participants read the information sheet and gave their informed consent - users were asked to complete a set of tasks pre-defined by the evaluating team. These tasks were chosen to simulate the intended experience of a real user playing the game and were designed to align with the User Requirements. For example, tasks such as placing & moving buildings (*UR_LOCATIONS_MOVE*, *UR_LOCATIONS*) and muting & unmuting the music (*FR_AUDIO_MUTE*) were included. Each session involved both task completion and feedback collection, focusing on usability challenges and areas for refinement.

The team used Google Docs for collaborative creation of the User Evaluation Protocol and decided to record findings under the "Interviewer Notes" subheading within this centralised document. This ensured all findings were kept in one place and maintained data consistency due to the real-time editing capabilities of Google Docs.

Data was collected by:

- Observations: Interviewers documented any usability problems the users encountered whilst interacting with the game. These were classified using a 4-point scale (cosmetic to catastrophic) to identify their severity in terms of user experience. The participants were encouraged to describe their experiences using the 'Think-Aloud protocol' to aid in the interviewer's observations.
- Post-Task Questionnaire: The participants completed the System Usability Scale[2] questionnaire for the system after they completed all the tasks. We used this standardised method to gather consistent, comparable results. We could then calculate the overall SUS score for each participant: all findings can be found in the [User Evaluation Protocol](#). The average SUS score for the 6 participants tested was 80.42. The team extended the questionnaire to include two final qualitative questions: '*Was there anything you thought was missing?*' and '*Do you have any other comments on the game?*' to gain more qualitative data on the user's personal experiences and to gather additional suggestions for improvement the team may not have considered - allowing the participants to provide their fresh, unbiased perspective.

Ethics

All users provided informed consent before participating in the User Evaluation. They were fully briefed about the process and purpose of the test, and informed that any responses/information they

gave would be used only for the purposes of the user evaluation. Anonymity was ensured by assigning each participant a unique number for reference to comply with ethical standards.

Part B: Usability Problems in Prototype System

The table below lists the usability problems found by our users. It is structured with 6 columns. Firstly the problem ID allows quick identification of each usability issue. The severity rating ranks each problem on a 4-point scale from cosmetic (1) to catastrophic (4), allowing the team to easily prioritise which issue to resolve first. Frequency identifies the proportion of the users where this problem occurred. Notes/ Recommendations details the evaluating team's recommendations on how to resolve each problem. The final column is the resolution status - which has been updated throughout the project as problems have been systematically resolved or partially resolved based on their priority level. Any problems remaining unresolved were done so due to the time constraints of the project combined with their low severity level.

Problem ID	Description	Severity Rating	Frequency	Notes/Recommendations	Resolution Status?
P1	Satisfaction score not visible to player when paused.	Catastrophic (4)	2/6	Ensure satisfaction score does not disappear when the game is paused. Reported to the development team and fixed between second and third user tests - hence 2/6 frequency.	Resolved
P2	The use of shapes as placeholders for building art is unintuitive so difficult for users to identify building types.	Minor (2)	3/6	Integrate art assets for each building type so users can identify the types more intuitively.	Resolved
P3	Found red lines surrounding buttons on menu offputting.	Cosmetic (1)	3/6	Remove lines bordering buttons. These were table cell borders created by a debug mode that had been left enabled.	Resolved
P4	Game crashed upon reaching Year 3.	Catastrophic (4)	1/6	Used the error message generated when the game was run from the terminal to identify the problem in code. Occurred in the first user test, reported to the development team and immediately fixed - hence 1/6 frequency.	Resolved
P5	Users felt the spacing of building icons in UI was unaesthetic.	Cosmetic (1)	2/6	Evenly space the UI buttons to select buildings to place.	Unresolved

P6	Events did not feedback their effect on satisfaction to the user.	Minor (2)	5/6	Implement + or - symbols to indicate if response to even positively or negatively impacted student satisfaction.	Unresolved
P7	Could not view the leaderboard to compare their own score to other users.	Major (3)	4/6	Ensure the past leaderboard entries are visible when the game is over.	Resolved
P8	Event text overlaps 'Currency' value.	Minor (2)	5/6	Alter UI to allow both to be visible at the same time.	Unresolved

References

[1] Nielsen, J., "Why You Only Need to Test With 5 Users", Alertbox March 19, 2000, at <http://www.useit.com/alertbox/20000319.html>

[2] Lewis, J. R. (2018). "The system usability scale: past, present, and future." *International Journal of Human-Computer Interaction*, 34(7), 577-590. March 30, 2018, at <https://www.tandfonline.com/doi/full/10.1080/10447318.2018.1455307>