Evaluation Report for a Human-Computer Interaction based Gamified Multimodal Energy-Saving Prototype:

A Coursework Evaluation Report for

Advanced Human Computer Interaction (F21AD)

BY: SYED ARIF ALI – sa4001@ hw.ac.uk

Table of Contents

1 Introduction ... 3
1.1 Goals And Objectives ... 3
2 Evaluation Protocol ... 4
2.1 Participants ... 4
2.2 Protocol Design ... 4
2.3 Evaluation Methods: ... 5
2.3.1 Qualitative ... 5
2.3.2 Quantitative ... 7
2.3.3 Rejected Evaluation Methods ... 10
2.4 Procedure ... 13
2.4.1 Task Pathway 1: Parent Remotely Managing Energy Usage ... 13
2.4.2 Task Pathway 2: Child Engaging In Gamified Task ... 14
2.5 Data Collection Instruments ... 15
2.6 Results And Analysis ... 15
2.7 Conclusions & Design Changes ... 20
3 Ethical Considerations: ... 22
4 Figma Prototype Link ... 22
5 Participants (Evaluators) ... 22
References ... 23
Appendix 1 ... 25
Appendix 2 ... 30

1 Introduction

This evaluation protocol outlines methods, procedures, participants, tasks, data collection, and analysis methods in a Human-Computer Interaction (HCI) study using the Eco-Watch Gamified Energy-Saving Prototype, that gamifies energy conservation through multimodal interaction within an (Internet of Things) IoT-integrated household to determine the consistency, reproducibility, and validity in assessing this interactive system (Lazar et al., 2017; Nielsen, 1993).

1.1 Goals And Objectives

The goal is to

1. Design an evaluation protocol.
2. evaluate the Human-Computer Interaction (HCI) in the Eco-Watch Gamified Energy-Saving Prototype.

The primary objectives therefore are to

1. Define the approach to evaluating the prototype.
2. Assess usability (effectiveness, efficiency, learnability) and user engagement of the prototype for its target audience, including comparisons between children and adults.
3. Uncover any design issues that hinder optimal workflow of the app.

2 Evaluation Protocol

2.1 Participants

The evaluation targeted a mix of users reflecting the intended demographic (families with tech-savvy adults and children, including users with normal and colour-blind vision).

Five users (N=5, ages 19–50) were recruited as participants, drawn from the target demographic of Eco-Watch (family members who might use a home energy app). This sample included two younger adults (19–30) and three middle-aged adults (31–50). Participants had varying familiarity with energy-saving or smart-home apps (two “very familiar”, one “somewhat familiar”, two “not familiar”), as recorded in a pre-survey.

Due to practical constraints, we were unable to include actual child users or any colour-blind users in this small-sample evaluation. Instead, we instructed the adult participants to consider the app’s appeal to children and to comment on visual elements to gain some insight into what might be confusing or problematic for children or colour-blind users.

2.2 Protocol Design

This protocol was designed following HCI best practices (e.g. combining qualitative and quantitative measures) to thoroughly evaluate how well Eco-Watch works in practice. A mixed-methods approach using qualitative and quantitative methods was selected and how each addresses the evaluation goals : three qualitative methods were considered to capture in-depth user feedback, and three quantitative methods to measure performance and perceived usability. Participants were asked to perform tasks one by one, using the prototype.

Two of the three considered methods each for the qualitative and two quantitative methods, were finally employed for the evaluation. The qualitative methods used were “Think aloud” engagement with the prototype and a semi-structured 10-minute interview, while the quantitative evaluation was carried out using the time taken for task completion and the standard ten-item System Usability Scale (SUS) questionnaire.

The next sections detail the execution of this protocol and the rationale behind the chosen methods.

2.3 Evaluation Methods:

The final choice of the two methods employed for each are justified with advanced HCI research below.

2.3.1 Qualitative

2.3.1.1 Think-Aloud Protocol

We employed a concurrent think-aloud protocol during task performance. In this well-established usability method, users verbalize their thoughts, feelings, and reasoning as they interact with the system. The facilitator prompted participants with gentle reminders to "keep talking" whenever they fell silent, without otherwise influencing their actions. The think-aloud method was chosen because it provides direct insight into users' cognitive engagement and immediate reactions to the interface to understand why they take certain actions or where they get confused, rather than only seeing what they do. This yields a richer qualitative data than a passive observation alone.

Justification: The think-aloud technique has its roots in cognitive psychology (Ericsson & Simon's work on verbal protocols) and was introduced to HCI by Clayton Lewis in 1982. It has since become standard in usability testing for uncovering practical usability problems related to task performance. In an international survey of usability professionals, 98% reported using think-aloud in their testing, ranking it as the most frequently used evaluation method. This widespread adoption in both industry and research attests to its effectiveness at revealing issues that matter to users (e.g. points of confusion, hesitations, misunderstandings). Think-aloud is particularly valuable for an advanced HCI prototype like Eco-Watch, which involves novel multimodal interactions (gesture and voice control): it helps the evaluator catch where users might be puzzled by these new interaction styles and hear their reasoning in real time. Studies have shown that think-aloud testing with real users tends to reveal more task-specific usability obstacles than expert inspection methods, for example.

Limitations: We acknowledged a potential limitation as some user groups (especially young children) may have difficulty articulating their thoughts continuously. Prior research notes that children under ~10 years old can find the think-aloud process challenging, as their verbalization and self-reflection skills are still developing. However, our study involved only adults with all participants being able to vocalize their thoughts throughout the tasks.

Nevertheless, this consideration reinforces why actual child testing would be needed in future work.

Overall, the think-aloud protocol gave us invaluable qualitative observations: for example, one participant muttered “Hmm, I’m not sure what to do next…” when facing an unlabelled icon during a task, immediately flagged the icon as confusing. Such real-time feedback is extremely actionable. By using think-aloud, we adhere to Nielsen’s classic recommendation that observing users “while they talk through their experience” is one of the most effective ways to identify usability problems. Indeed, think-aloud remains a cornerstone of usability evaluation in HCI, as it captures the “voice of the user” during interaction, something quantitative methods cannot provide.

2.3.1.2 Semi-Structured Interviews (Qualitative)

These were conducted individually with each participant using a list of open-ended questions focusing on key topics (usability of the interface, the multimodal interaction experience, engagement and motivation, likes/dislikes, etc.). However, the interviews were flexible with follow-up questions for clarification of responses. This semi-structured format was chosen because it balances consistency (ensuring all important topics are covered) with flexibility to explore unexpected insights. Such interviews are widely used in HCI to gain *in-depth understanding* of users’ perceptions, feelings, and the contextual reasons behind their behaviors. They allow users to elaborate on issues in their own words, often uncovering design feedback that would not surface from observations or questionnaires alone.

Justification: Semi-structured interviews are a core qualitative method in HCI research, valued for their ability to reveal users’ *mental models, preferences, and rationale* in detail. According to Blandford (2022), interviews enable researchers to probe participants’ values and experiences, yielding rich insights that purely observational methods might miss. In our evaluation, this was crucial for understanding *why* certain aspects of Eco-Watch worked or didn’t work. For example, during the interview one participant explained that the gamified *Zaps* points “might be fun for kids but maybe not for adults,” providing context that the same feature can be motivating to one group but not another. Such nuanced feedback comes out readily in a conversational interview setting.

Semi-structured interviews are particularly appropriate here because our prototype involves novel interactions and a motivational design – areas where users’ subjective impressions and emotional responses are as important as task performance. Through the interviews, participants felt comfortable discussing both positive impressions and pain points in a narrative form. This method is supported by HCI research as a means to capture aspects like enjoyment, engagement, and perceived usefulness. Blandford et al. (2016) emphasize that interviewing users after a usability test can uncover issues that did not manifest as observable errors but still affect user satisfaction. In our instance, the absence of a "Back" button did not prevent task completion but was mentioned by multiple participants as a missing convenience – a finding we might have gotten overlooked without interviews.

In summary, the interviews provided a wealth of qualitative data on *why* users felt a certain way about Eco-Watch. They allowed us to explore the **adult vs. child appeal** in more depth by asking participants to reflect on how a child might use the app, and to discuss any accessibility concerns (like color usage) they noticed. This method was a critical complement to the think-aloud protocol: where think-aloud captured immediate reactions during tasks, the interviews enabled reflection and discussion, leading to concrete suggestions for improvement. Combining these two qualitative methods follows best practice in usability evaluation – capturing both real-time usage data and reflective feedback yields a holistic understanding of the user experience.

2.3.2 Quantitative

2.3.2.1 System Usability Scale (Quantitative)

We included the **System Usability Scale (SUS)** questionnaire as a standardized quantitative measure of perceived usability. SUS is a ten-item Likert-scale survey that yields a single score from 0 to 100, where higher scores indicate better overall usability . It has participants rate statements such as "I thought the system was easy to use" on a 5-point agreement scale; the mix of positive and negative phrasing in SUS provides a robust gauge of usability when computed. We chose SUS for several reasons: it is quick for users to complete, it's technology-agnostic (applicable to any system), and it has been extensively validated in HCI research. Brooke (1996) originally introduced SUS as a crude usability scale that still produces reliable results even with small sample sizes. Subsequent studies have confirmed its high reliability (often Cronbach's alpha > 0.9) and wide adoption as it is considered a de facto standard for summative usability evaluation.

Justification: Including a SUS questionnaire in our evaluation is supported by its strong track record in literature. Bangor et al. (2008) demonstrated that SUS is highly reliable and correlates well with other usability measures; they also established industry benchmarks for SUS scores across many products. For example, an average SUS around 68 is considered “okay” usability, while scores above 80 indicate “excellent” usability. Using SUS in our study provides an **objective metric** that we can compare against these known benchmarks and grading scales. This is particularly useful in an academic context to quantify how our prototype ranks. By analyzing the SUS scores, we can quantitatively confirm or refute what we see in observations. For instance, if users verbally said the app was easy but SUS scores were low, that discrepancy would be a red flag. Conversely, consistently high SUS scores would reinforce qualitative positive feedback. Including SUS is in line with advanced HCI practice of mixing subjective rating scales with behavioral data. It allowed us to benchmark Eco-Watch: if the average SUS was well below 68, we’d know usability needs serious improvement; if around 80+, it indicates we achieved a high level of usability even for first-time users.

Moreover, SUS provides a single composite score which is easy to communicate to stakeholders and to track improvements over iterations. It complements the diagnostic insights from think-aloud and interviews with a bottom-line number for usability.

In our protocol, each participant completed the SUS immediately after the task session (before the interview) to capture their impression without too much reflection or moderator influence. The SUS covers various facets of usability (complexity, ease of use, need for support, consistency, etc.), so it acts as a broad summary of the user’s experience.

2.3.2.2 Task Completion Time and Error Rate (Quantitative)

We quantitatively assessed usability by measuring Task Completion Time and Error Rate, aligning with the ISO 9241-11 standards for efficiency and effectiveness.

Task Completion Time: We consistently measured the duration from task initiation to goal completion, capturing objective insights into interface efficiency (ISO 9241-11:2018). Faster task completion indicates intuitiveness, while prolonged task durations highlight potential inefficiencies or confusing interface elements.

Justification: Timing tasks is an established practice in Human-Computer Interaction (HCI) research, facilitating comparisons between different interface designs and identifying subtle usability issues users might not explicitly articulate (MacKenzie, 2013; Nielsen, 1993). Our study hypothesized that users familiar with smart-home applications would complete tasks faster and anticipated variations based on interaction modes (voice vs. touch), especially during initial interactions. These hypotheses are consistent with established HCI literature regarding expertise and modality effects.

Error Rate: We recorded errors defined as failures to complete tasks correctly without assistance, including incorrect actions that required corrections or external hints. Examples included clicking incorrect buttons or navigation mistakes requiring moderator intervention.

Justification: Quantifying errors is a core usability practice, essential for pinpointing specific interface challenges (Sauro & Lewis, 2016; Nielsen, 1993). We maintained consistent error criteria by qualitatively noting minor self-corrected slips unless indicative of significant confusion. Higher error rates signified problematic interface components needing redesign, while lower rates indicated user-friendly and forgiving designs (ISO 9241-11:2018). Error pattern analysis provided empirical evidence of usability issues, such as misleading iconography or unclear multimodal interaction steps, directly informing targeted improvements and validating the prototype's effectiveness (Sauro & Lewis, 2016; Nielsen, 1993).

2.3.3 Rejected Evaluation Methods

In designing the evaluation, we also considered other possible methods but rejected them in favor of those chosen above. The two notable methods considered but not used were **focus group** and **eye-tracking;** the justification for not using them based on HCI research, to demonstrate a critical approach to method selection.

2.3.3.1 Focus Groups (Qualitative)

A **focus group** involves gathering multiple users together to discuss their opinions and attitudes about a product, guided by a moderator. We initially considered running a focus group with potential users (for instance, a mixed group of adults and children) to get collective feedback on Eco-Watch. However, we rejected this idea for several reasons:

- **Not suitable for usability testing:** Focus groups are an attitudinal method better suited for early-stage idea generation or requirement gathering, not for observing how people use an interface. In a focus group, users *talk about* the product, but they are not performing tasks individually. As a result, a focus group cannot provide objective data on task performance or uncover detailed usability problems. (Nielsen, 1993).

- **Risk of group biases:** In a group setting, participants might influence each other's responses. Dominant individuals can steer the discussion, or quieter participants may hold back their true opinions (classic issues of "groupthink"). This is especially problematic given our target demographics: if we had a mixed adult-and-child focus group, adults might dominate and children might not speak freely, or children's presence might cause adults to adjust what they say (and vice versa). The dynamic could distort feedback. In a usability test or interview, each person's feedback is independent, avoiding this problem.

- **Limited insight on design usage:** While focus groups can yield general feelings or ideas, they often produce surface-level comments regarding likes and dislikes, without the context of actual use. Participants in focus groups might speculate about features in the abstract, which is less reliable than comments made while actually using the prototype. As Preece et al. (2015) notes, focus groups are useful for *collecting users' initial perceptions or expectations* about a concept, but not for evaluating a user interface's concrete usability in action, as in our Eco-Watch

- **Practical constraints:** Organizing a focus group with both children and adults together would be logistically and methodologically challenging. It's often better to separate such different age groups to get honest feedback from each; but our resources were limited, and we prioritized individual sessions with adults (which we could manage) over attempting a potentially flawed focus group with mixed ages.

We, thus, concluded that a focus group would **not meet our evaluation goals**. It might provide some collective opinions but would likely miss the detailed usability issues and varied individual experiences that one-on-one sessions captured. Our choice of individual think-aloud tests and interviews ensured more authentic and actionable usability findings.

2.3.3.2 Eye-Tracking (Quantitative)

We also considered integrating **eye-tracking** into the usability sessions. Eye-tracking technology can record where on the screen users are looking (their gaze points and fixations) during tasks. In theory, this could have given us additional data on what interface elements attracted attention or if users overlooked certain features. However, we decided against using it for several reasons:

- **Special equipment and setup:** Eye-tracking requires hardware (an eye-tracker device) and careful calibration for each user to ensure accurate data. Our project resources were limited and did not include access to an eye-tracking lab. Given the short 30–40 minute session time, doing a calibration and dealing with any technical issues would eat into the actual evaluation time. The cost and complexity were not justifiable for a small-sample study.

- **Limited benefit for our research questions:** Our focus was on multimodal interaction usability and user engagement. Eye-tracking primarily provides information about visual attention – *where* users look and for how long. However, it does not tell us *why* they are looking there or whether they understood what they saw. For instance, if a user stares at an icon for 5 seconds, eye-tracking can show that fixation, but only through think-aloud or interview would we learn if that was because they found the icon confusing or interesting. Researchers caution that **fixations do not equal comprehension** – users might look at something and not grasp it, or conversely perceive something in peripheral vision without gazing directly at it. In other words, eye-tracking alone cannot capture the user's cognitive state. Since we already had a think-aloud protocol revealing the *why* behind user actions, adding eye-tracking would offer relatively little new insight into the issues we were investigating.

- **Not well-suited for multimodal interactions:** Eco-Watch is not a purely visual interface; it involves gestures and voice. During a voice command, for example, the user might be looking at a microphone icon or perhaps at no particular visual target. Eye-tracking in such moments (when the user is concentrating on speaking or gesturing) might not yield meaningful data – a user could even look away from the screen while speaking, which doesn't necessarily indicate a problem. Similarly, if a user performs a hand gesture, their gaze might flicker to some visual feedback, but the critical measure of success is whether the gesture was recognized, not where they looked. Eye-tracking is most useful in scenarios like web page navigation or reading, where the primary channel is visual, but not in a multimodal context, like ours. Our evaluation questions (e.g. "Do users understand how to invoke voice mode?") are better answered by task success and user comments than by gaze patterns.
- **Data analysis overhead:** Eye-tracking produces complex data (heatmaps, gaze plots, fixation counts) that require careful analysis and interpretation and would be inappropriate for a small study. Often, it involves significant post-processing to correlate gaze data with interface areas and user actions. Besides, no substantial difference in results could be expected and would require a lot of additional effort. Prior research has noted that interpreting eye-tracking results needs complementary methods like verbal reports, which were already available with us. S

Considering these points, we decided that eye-tracking was **included** for this evaluation. Jacob & Karn (2003) noted that while eye-tracking can enrich usability studies, its benefits must be weighed against practicality and the particular study goals. Eye-tracking would have added technical complexity without substantially altering our conclusions.

Selecting the above qualitative (think-aloud, interviews, thematic analysis) and quantitative (task metrics, SUS, etc.) methods, addresses both *usability* and *engagement*. Although we did not have actual child participants or color-blind users, the methods chosen are appropriate for evaluating those dimensions by proxy.

2.4 Procedure

Step-by-step instructions for conducting the study, including task scenarios, session flow, and briefing/debriefing. Each evaluation session lasted around 30–40 minutes in a quiet room, using a high-fidelity Eco-Watch prototype on a laptop (via Figma). All sessions were conducted online , and each participant received a brief orientation and gave informed consent before beginning.

Task Scenarios

A set of representative **task scenarios** was defined to cover Eco-Watch's core interactions and features. These tasks were derived from the app's key use cases and included:

- **Task 1:** Navigate the dashboard to find a specific device's current energy usage.
- **Task 2:** Use the gesture-based control (a "wave" gesture) to turn off a device.
- **Task 3:** Issue a voice command to activate an energy-saving mode (e.g. "eco mode").
- **Task 4:** View the gamified rewards – check the "Zaps" points balance and the leaderboard of energy savers.

2.4.1 Task Pathway 1: Parent Remotely Managing Energy Usage

The flow of the app use case in the parent's scenario goes as follows :-

- Parent logs into the Eco-Watch app.
- Views real-time household energy consumption on the Dashboard.
- Notices that an appliance (e.g., Air Conditioner) is consuming excess energy.
- Navigates to the Device Management section.
- Turns off the appliance remotely using the app.
- Notices updated energy statistics on the Parent Home Screen.

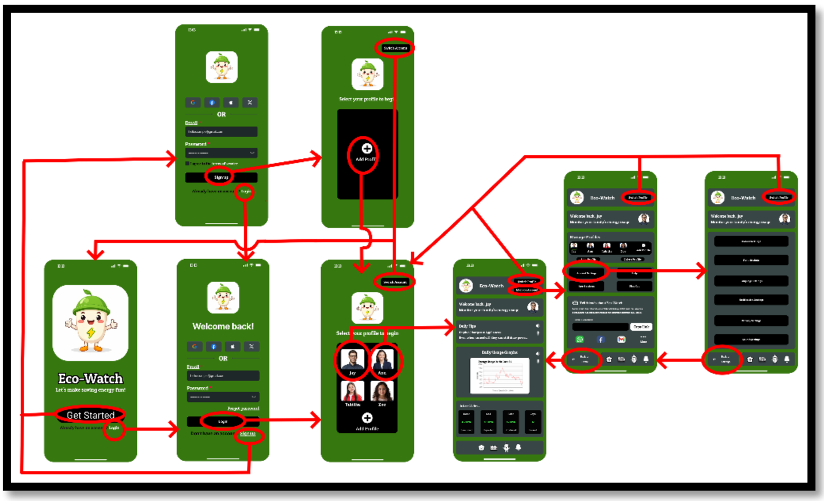

Fig.1 - User Flow for Parent through the App (New User)

Fig .2 - User Flow for Parent through the App (Existing User)

2.4.2 Task Pathway 2: Child Engaging in Gamified Task

The flow of the app use case in the child's scenario goes as follows :-

- Child logs into the Eco-Watch app.
- Receives an alert: "Task Available: Help Mr. Zappy Save Energy!"
- Presses "Begin Task".
- Waves at Mr. Zappy (gesture-based interaction) to turn off a virtual light.
- Earns Zaps (reward points) and moves up the leaderboard.
- Receives an energy-saving tip as a reward.

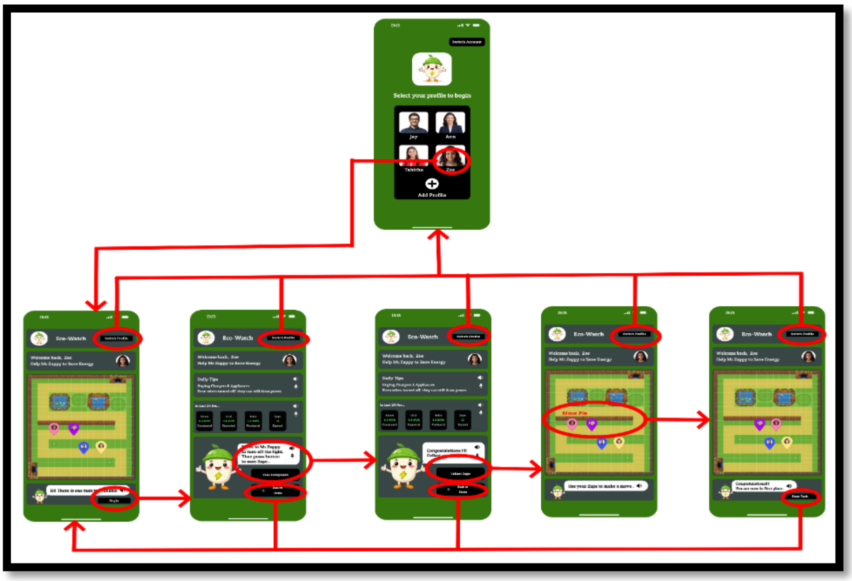

Fig.3 - User Flow for Child through the App

2.5 Data Collection Instruments

Tools and measures used

Qualitative

1. Think-aloud protocol
2. Semi-structured interviews

Quantitative

1. System usability scale
2. Task completion time and error rate

2.6 Results And Analysis

Qualitative Data:

The technique employed for analysing qualitative data was thematic separation. All notes and transcripts from the sessions were reviewed and coded for key themes using the approach of Braun & Clarke (2006). We looked for patterns in user comments, recurring usability issues, and common suggestions. For example, comments related to *navigation* were grouped together, as were comments about *gamification*. Through iterative coding, we identified a set of themes that summarize the user feedback (such as "Intuitive Touch Navigation," "Voice Interaction Challenges," "Engagement for Adults vs Children," and "Desire for Better Navigation Controls"). This qualitative analysis method was chosen to rigorously interpret user feedback rather than relying on anecdotal impressions.

Qualitative data highlighted positive reactions to gamified elements, with users finding the reward system engaging. However, there was recognition that adults and children might differ in gamification preferences. Multimodal interactions (voice, gesture, touch) were positively received, though users identified friction in switching between modes.

Navigation challenges, especially the absence of a Back button, were frequently noted, particularly from a child's usability perspective.

Feedback also emphasized the need for clearer instructions for multimodal actions and suggested potential improvements like customizable gamification features or a global voice activation button.

Overall, the evaluation demonstrated that Eco-Watch was user-friendly, effective, and engaging, with targeted areas for improvement clearly identified through quantitative and qualitative analyses.

Minor errors included navigation issues, such as misunderstanding menu icons or confusion over the absence of a Back button. Another common issue was the requirement to tap a microphone icon repeatedly to use voice commands, causing temporary confusion. Errors were typically corrected quickly, with participants suggesting specific UI improvements (e.g., clearer navigation icons or a global voice activation button). No irrecoverable errors occurred, indicating the interface was forgiving and safe.

FEEDBACK GRAPHS AND STATISTICS

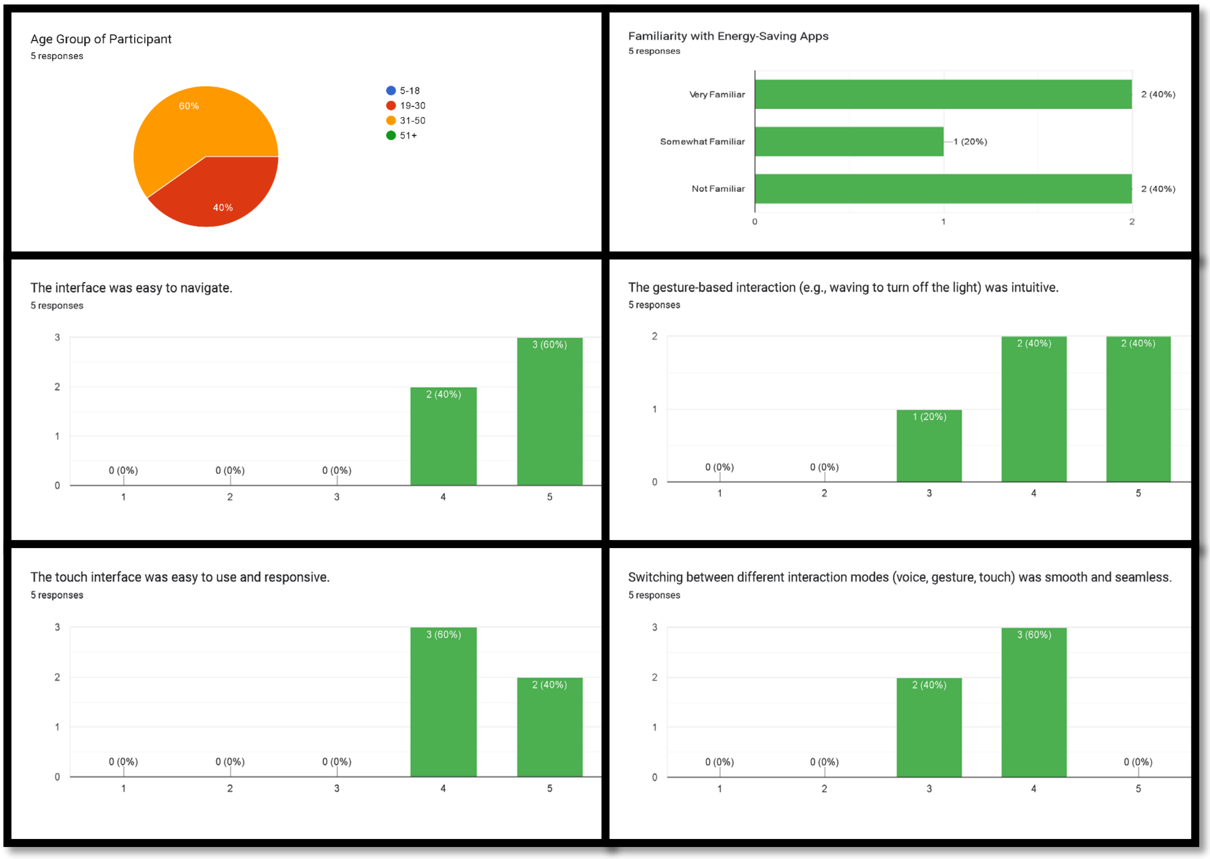

Fig.4 – Frequency Graphs for Think out Loud Protocol

FEEDBACK GRAPHS AND STATISTICS

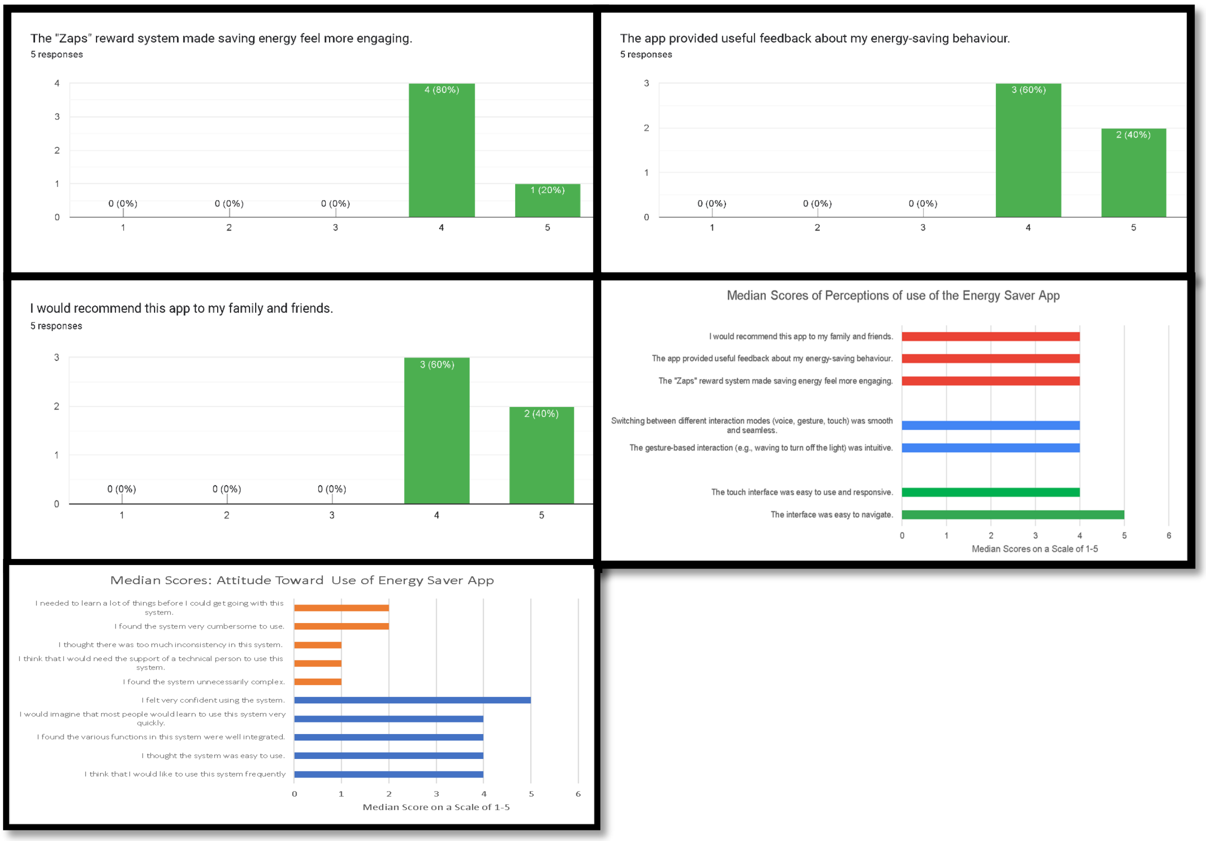

Fig.5 – Frequency Graphs for User Feedback

QUANTITATIVE DATA

Standard descriptive statistical analysis was employed.

1 User Satisfaction and Engagement:

Participants rated usability highly, particularly for ease of navigation (average rating 4.6/5), touch responsiveness (4.4/5), and gesture intuitiveness (4.2/5). However, the smoothness of switching interaction modes received a lower rating (3.6/5), indicating a need for improvement. Gamification features were generally well-received, particularly for motivating engagement, though preferences varied among adults versus children. Users also appreciated feedback about their energy-saving behaviours, rating this aspect highly (4.4/5). The majority indicated they would recommend the app (4.4/5).

2 System Usability Scale (SUS):

Eco-Watch achieved high SUS scores, ranging from 72.5 to 92.5, with an average of 84.5, categorizing it as "Excellent." High SUS scores confirmed positive qualitative feedback and suggested strong perceived usability. Users reported high confidence in system use and low perceived complexity, indicating that the core interface was robust and accessible even to novices.

SUS GRAPHS AND STATISTICS

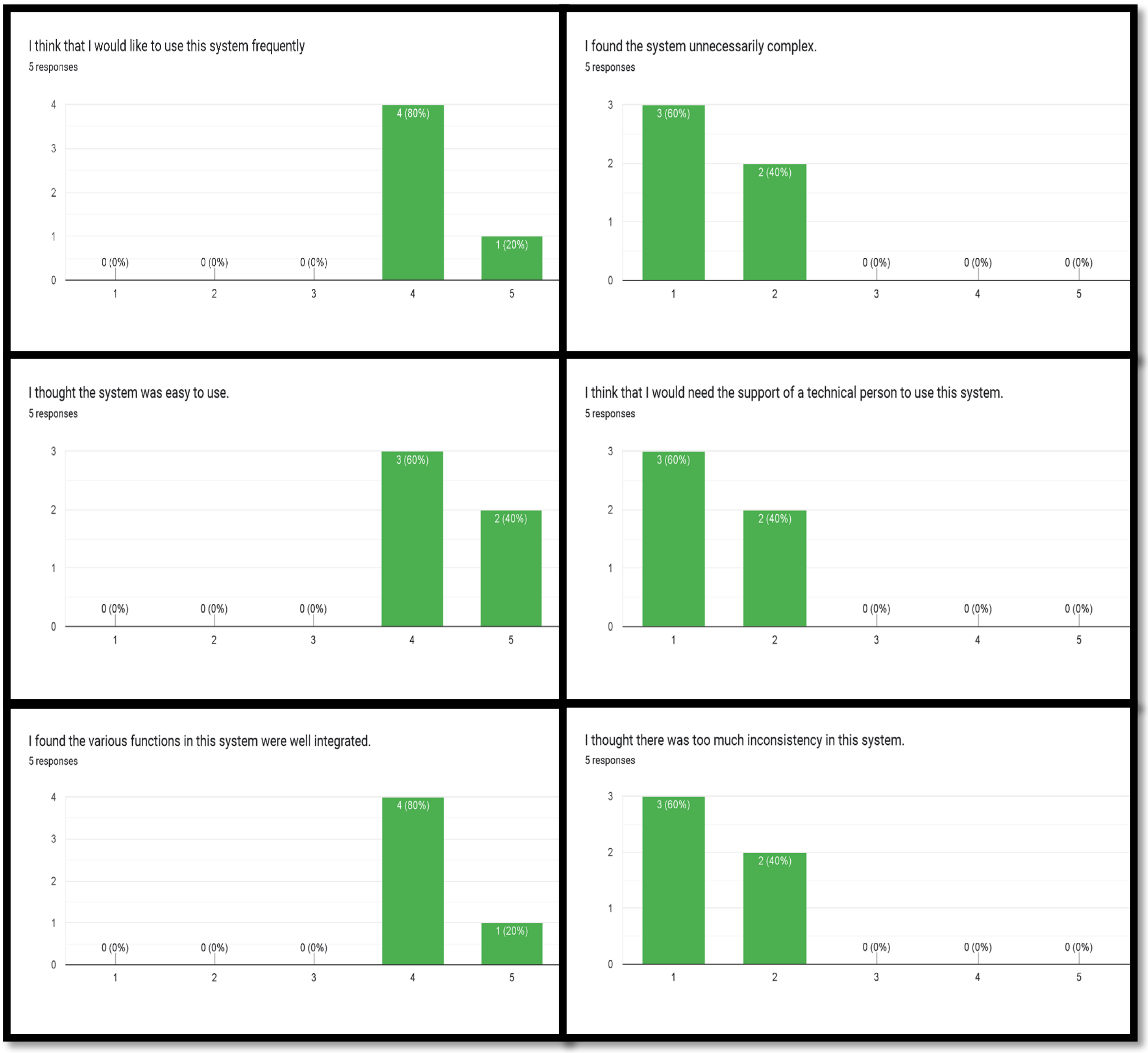

Fig.6 – Frequency Graphs for System Usability Survey

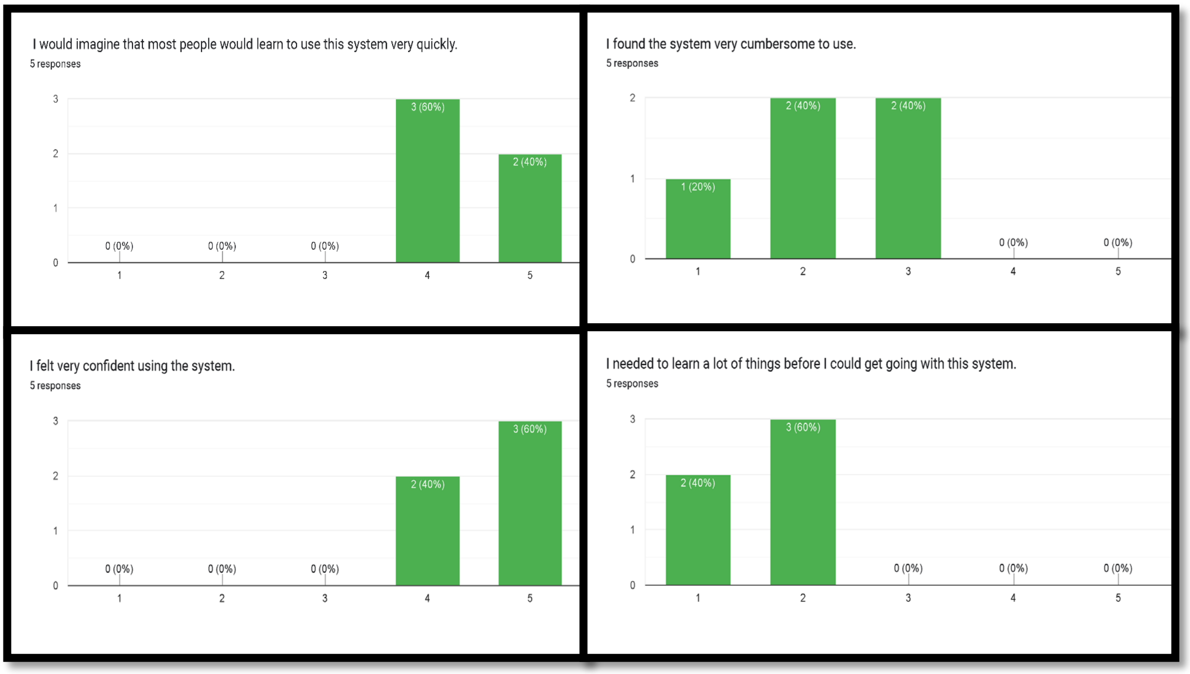

Fig.7 – Frequency Graphs for System Usability Survey

3 Research Basis and Justification: Quantifying errors is a core practice in usability studies, essential for identifying interface issues and prioritizing redesigns (Sauro & Lewis, 2016; Nielsen, 1993). Clear criteria for error classification ensured consistency. Error analysis highlighted specific usability problems (e.g., misleading iconography), directly informing targeted improvements and validating overall effectiveness (ISO 9241-11:2018).

2.7 Conclusions & Design Changes

The evaluation indicates that the Eco-Watch prototype is highly usable and engaging. Participants praised the intuitive interface, particularly the touch navigation and informative dashboard, successfully utilizing voice and gesture features despite minor issues. Gamified elements, such as earning Zaps points and challenges, increased motivation and enjoyment, aligning well with the app's objective of making energy conservation enjoyable. The SUS scores (average ~84.5) suggest excellent usability, and task performance was efficient, indicating a solid core design for user-friendly family engagement.

Nevertheless, several areas for improvement were identified. Key issues included smoother navigation, more seamless mode switching, and accommodating diverse user preferences and accessibility (children, adults, colour-blindness). The following targeted design changes are proposed based on evaluation findings:

Navigation & Icons: Introduce a dedicated Back button to enhance intuitive navigation, addressing user confusion over its absence. A standard back arrow in the top-left corner on secondary screens, such as challenge or device details, is recommended. Furthermore, redesign the "Devices" menu icon to avoid confusion with "Home," potentially using distinct symbols (e.g., plug or bulb) or adding a "Devices" label underneath. These changes will significantly improve navigation clarity for diverse user groups.

Voice Interaction: Streamline the initiation of voice commands to enhance multimodal integration. Proposals include introducing a persistent microphone button accessible on all screens, possibly in the header or as a floating action button, thus removing the necessity to locate specific voice icons on individual screens. Alternatively, implementing a wake-word (e.g., "Eco Watch, do X") could further simplify voice input. Additionally, clearer on-screen instructions—such as overlay hints clearly stating, "Tap the microphone icon on the top-right, then speak your command"—will reduce confusion and encourage more frequent voice interactions.

Seamless Mode Switching: Enhance fluid transitions between interaction modes by integrating multimodal inputs, eliminating manual mode toggles. Users should effortlessly move between touch, gesture, and voice inputs without explicitly changing modes. For instance, gestures already accepted without toggles could be complemented by alwaysavailable voice commands through a global microphone button or wake-word activation. Settings could include simple mode toggles for continuous listening or automatic gesture detection. Consistent feedback, such as brief confirmations ("Voice command executed") with optional undo functionality, will further reduce cognitive load and encourage diverse interaction methods.

Feedback & Accessibility: Strengthen feedback mechanisms and accommodate accessibility needs to foster user confidence. Immediate, clear confirmations after gestures or voice commands—such as animations, visible checkmarks, or auditory cues—will clarify successful actions, mitigating uncertainty experienced by some users. Accessibility enhancements include implementing a high-contrast colour scheme option tailored for colour-blind users, supplemented with labels, patterns, or shapes to avoid reliance solely on colour. Adding explicit annotations or labels to coloured graphs will facilitate data interpretation. Adjustable text sizes within the app will cater to readability preferences, particularly benefiting older or visually impaired users. These inclusive design modifications align with WCAG guidelines, promoting broader usability without compromising the standard user experience.

Gamification Customization: Adapt gamified elements to user age and preferences to optimize motivational impact. Recognizing varied user responses to gamification, the proposal includes configurable elements. Adult users desiring a more pragmatic experience could deactivate cartoon visuals or prioritize practical metrics (energy or monetary savings) through a settings toggle. Conversely, younger or game-inclined users could activate prominently displayed points, leaderboards, mascots, or additional challenges. Introducing distinct user modes—such as a highly gamified "Family Fun" mode versus a practical, less gamified "Efficiency" mode—would enable personalized user experiences while preserving core functionalities.

Implementing these proposed refinements directly addresses identified usability issues, significantly enhancing the user experience. Navigation clarity, seamless multimodal interactions, immediate inclusive feedback, and adaptable gamification features collectively ensure Eco-Watch meets diverse user needs. The app's intuitive interface, engaging concept, and effectiveness will be maintained and enhanced, positioning Eco-Watch effectively to fulfil its goal of educating and motivating families toward energy-saving behaviours through enjoyable interactions.

3 Ethical Considerations:

Protocols for informed consent, confidentiality, and participant well-being (our study was approved for ethical compliance, and all participants gave informed consent).

4 Figma Prototype Link

https://www.figma.com/proto/01jw3qDFwOrb9R7CNAfI20/Coursework---PG-Dubai---Group-3?page-id=0%3A1&node-id=3-14&p=f&viewport=36%2C38%2C0.16&t=DWsUO6v1FQH0nzAB-1&scaling=scale-down&content-scaling=fixed&starting-point-node-id=3%3A14

5 Participants (Evaluators)

- Asma Damankesh
- Hariharakumar Rathinar
- Husna Ahmed
- Nawshad, Nourin
- Danish, Mariam

REFERENCES

Bangor, A., Kortum, P. T., & Miller, J. T. (2008). **An empirical evaluation of the System Usability Scale (SUS).** *International Journal of Human–Computer Interaction, 24*(6), 574–594.

Blandford, A., Furniss, D., & Makri, S. (2016). **Qualitative HCI Research: Going Behind the Scenes.** Morgan & Claypool Publishers.

Blandford, A. (2022). **Semi-Structured Qualitative Studies.** In Soegaard, M. & Dam, R. F. (eds.), *The Encyclopedia of Human-Computer Interaction* (2nd ed.). Interaction Design Foundation. *(Online)*.

Braun, V. & Clarke, V. (2006). **Using thematic analysis in psychology.** *Qualitative Research in Psychology, 3*(2), 77–101.

Brooke, J. (1996). **SUS: a "quick and dirty" usability scale.** In P. W. Jordan, B. Thomas, B. Weerdmeester, & I. L. McClelland (Eds.), *Usability Evaluation in Industry* (pp. 189–194). London: Taylor & Francis.

Field, A. (2013). **Discovering Statistics Using IBM SPSS Statistics** (4th ed.). Sage Publications.

Fessenden, T. (2022). **Focus Groups 101.** Nielsen Norman Group. *(Online)* Available at: https://www.nngroup.com/articles/focus-groups-definition/ (Accessed 31 July 2022).

ISO (2018). **Ergonomics of human-system interaction – Part 11: Usability: Definitions and concepts (ISO 9241-11:2018).** International Organization for Standardization.

Jacob, R. J. K., & Karn, K. S. (2003). **Eye tracking in human-computer interaction and usability research: Ready to deliver the promises.** In J. Hyönä, R. Radach, & H. Deubel (Eds.), *The Mind's Eye: Cognitive and Applied Aspects of Eye Movement Research* (pp. 573–605). Amsterdam: Elsevier.

Lazar, J., Feng, J. H., & Hochheiser, H. (2017). **Research Methods in Human-Computer Interaction** (2nd ed.). Morgan Kaufmann.

Lewis, C. (1982). **Using the "thinking-aloud" method in cognitive interface design.** IBM Corporation, Yorktown Heights, NY.

Lewis, J. R. (2006). **Usability testing.** In G. Salvendy (Ed.), *Handbook of Human Factors and Ergonomics* (3rd ed., pp. 1275–1316). John Wiley & Sons.

MacKenzie, I. S. (2013). **Human-Computer Interaction: An Empirical Research Perspective.** Morgan Kaufmann (Elsevier).

McDonald, S., Edwards, H. M., & Zhao, T. (2012). **Exploring think-alouds in usability testing: An international survey.** *IEEE Transactions on Professional Communication, 55*(1), 2–19.

Nielsen, J. (1993). **Usability Engineering.** Academic Press.

Preece, J., Rogers, Y., & Sharp, H. (2015). **Interaction Design: Beyond Human-Computer Interaction** (4th ed.). John Wiley & Sons.

Rubin, J. & Chisnell, D. (2008). **Handbook of Usability Testing: How to Plan, Design, and Conduct Effective Tests** (2nd ed.). Indianapolis, IN: Wiley.

Sauro, J., & Lewis, J. R. (2016). **Quantifying the User Experience: Practical Statistics for User Research** (2nd ed.). Morgan Kaufmann.

Soegaard, M. (2023). **What is Eye Tracking in UX?** Interaction Design Foundation. *(Online)* Available at: https://www.interaction-design.org/literature/article/eye-tracking-ux (Accessed 2024).

APPENDIX 1

Participant Form

Evaluation Of the Eco-Watch Gamified Energy-Saving Prototype
Course: Advanced Human-Computer Interaction (F21AD/F20AD)
Study Conductor: Syed Arif Ali
Institution: Heriot Watt University

You are Invited to Participate in A User Evaluation Study for The **Eco-Watch Gamified Energy-Saving Prototype**. This Study Aims to Assess the Usability, Engagement, And Interaction Effectiveness of The Prototype. Your Participation Is Voluntary, And You May Withdraw at Any Time Without Penalty.

This Study Is Conducted as Part of An Academic Project to Understand How Users Interact with The **Eco-Watch** App, Focusing on **Multimodal Interactions (Gesture, Voice, Touch)** And **Gamification Features**. Your Feedback Will Help Improve the Design.

What Will I Be Asked to Do?

If You Agree to Participate, You Will Be Asked To:

1. Use The Eco-Watch Prototype and Complete a Series of Tasks (E.G., Navigating the Dashboard, Completing an Energy-Saving Challenge).
2. Provide Feedback Through a Think-Aloud Protocol, Interviews, And Surveys.
3. Allow Recording of Observations (E.G., Time Taken to Complete Tasks, Usability Challenges).
4. The Session Will Take Approximately 30-40 Minutes.

Voluntary Participation & Withdrawal

- Your Participation Is Entirely Voluntary.
- You May Withdraw at Any Time Without Giving a Reason.
- If You Choose to Withdraw, Your Data Will Be Deleted Immediately.

Confidentiality & Data Protection

- Your Responses Will Be Kept Confidential and Used Only for Academic Purposes.
- No Personally Identifiable Information Will Be Published.
- Data Will Be Anonymized and Securely Stored.

Potential Risks & Benefits

- Risks: There Are No Known Risks Associated with Participation.
- Benefits: Your Feedback Will Contribute to Improving the Usability of Energy-Saving Applications.

Consent Declaration
Please Read the Following Statements Carefully and Check Each Box If You Agree:

☐ I have read and understood the information provided in this consent form.

☐ I voluntarily agree to participate in this study.

☐ I understand that my participation is confidential and my data will be anonymized.

☐ I understand that I can withdraw at any time without any consequences.

☐ I consent to my responses being recorded and used for research purposes.

Participant ID

User Feedback Interview (Demographic Data)
This Interview/Survey Section Collects Demographic Data from Users After Interacting with The Eco-Watch Gamified Energy-Saving Prototype to Identify User Age Sentiment and Experience.

Age Group:
☐ 10-18 ☐ 19-30 ☐ 31-50 ☐ 51+
Familiarity With Energy-Saving Apps:
☐ Very Familiar ☐ Somewhat Familiar ☐ Not Familiar

Feedback Interview on Usability, Multimodal Interaction, Engagement & Gamification
Open-Ended Questions (Qualitative)
(Users Rate Each Statement on A Likert Scale: 1 = Strongly Disagree, 5 = Strongly Agree)

The Interface Was Easy to Navigate. ☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5
The Gesture-Based Interaction (e.g., Waving to Turn Off the Light) Was Intuitive. ☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5
The Voice Commands Worked Accurately and Consistently. ☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5
The Touch Interface Was Easy to Use and Responsive. ☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5
Switching Between Different Interaction Modes (Voice, Gesture, Touch) Was Smooth and Seamless. ☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5
The "Zaps" Reward System Made Saving Energy Feel More Engaging. ☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5
Competing In Energy-Saving Challenges Motivated Me to Interact More with The App. ☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5
The App Provided Useful Feedback About My Energy-Saving Behaviour. ☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5
I Would Recommend This App to My Family and Friends. ☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5
What Did You Like the Most About the App?
What Features or Interactions Did You Find Confusing or Difficult to Use?
How Could the App's Multimodal Interactions (Voice, Gesture, Touch) Be Improved?
Did The Game-Based Rewards (Zaps, Leaderboards) Encourage You to Save Energy? Why Or Why Not?
Any Additional Suggestions for Improving the App?

System Usability Scale (SUS) Questionnaire
This Survey Section Collects a Standardized 10-Question Usability Survey That Is Used to Measure the Perceived Usability of The Eco-Watch Prototype.
Instructions: Please Indicate Your Level of Agreement with The Following Statements. (1 = Strongly Disagree, 5 = Strongly Agree)
SUS Question Rating Scale (1 = Strongly Disagree, 5 = Strongly Agree)

I Think That I Would Like to Use This System Frequently. **☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5**
I Found the System Unnecessarily Complex. (Reverse Scored) **☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5**
I Thought the System Was Easy to Use. **☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5**
I Think That I Would Need the Support of a Technical Person to Use This System. (Reverse Scored) **☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5**
I Found the Various Functions in This System Were Well Integrated. **☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5**
I Thought There Was Too Much Inconsistency in This System. (Reverse Scored) **☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5**
I Would Imagine That Most People Would Learn to Use This System Very Quickly. **☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5**
I Found the System Very Cumbersome to Use. (Reverse Scored) **☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5**
I Felt Very Confident Using the System. **☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5**
I Needed to Learn a Lot of Things Before I Could Get Going with This System. (Reverse Scored) **☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5**

Thank You for Participating!

Dear Participant,

Thank You for Taking the Time to Participate in Our Evaluation Study for The Eco-Watch Gamified Energy-Saving Prototype. Your Valuable Insights and Feedback Will Help Us Improve the Usability, Engagement, And Effectiveness of Our System.

Your Participation Contributes to Advancements in Human-Computer Interaction (HCI) Research, Particularly in Designing Gamified and Multimodal Interaction Experiences for Energy-Saving Applications.

If You Have Any Further Thoughts or Suggestions, Feel Free to Reach Out. Your Input Is Greatly Appreciated!

Next Steps:

- **Your Feedback Will Be Analysed and Used to Enhance the Prototype.**
- **If You Wish to Receive a Summary of The Study's Findings, Please Let Us Know.**
- **Your Contributions Make a Difference—Thank You Once Again!**

Best Regards,

Syed Arif Ali
Heriot Watt University
Sa4001@Hw.Ac.Uk

APPENDIX 2

CONSENT DECLARATION , PARTICIPANT ID AND FEEDBACK INTERVIEW ON USABILITY, MULTIMODAL INTERACTION, ENGAGEMENT & GAMIFICATION

	PARTICIPANT 1	PARTICIPANT 2	PARTICIPANT 3	PARTICIPANT 4	PARTICIPANT 5
Timestamp	25/03/2025 12:37:02 AM	25/03/2025 05:49:58 PM	25/03/2025 10:44:06 PM	26/03/2025 10:54:26 AM	26/03/2025 03:28:16 PM
Consent Declaration	YES	YES	YES	YES	YES
Participant ID	H00463082	H00473375	H00502683	H00504162	H00457771
Age Group of Participant	31-50	31-50	19-30	19-30	31-50
Familiarity with energy-saving apps	VERY FAMILIAR	NOT FAMILIAR	SOMEWHAT FAMILIAR	VERY FAMILIAR	NOT FAMILIAR
The interface was easy to navigate.	5	5	4	5	4
The gesture-based interaction (e.g., waving to turn off the light) was intuitive.	4	4	3	5	5
The touch interface was easy to use and responsive.	5	4	5	4	4
Switching between different interaction modes (voice, gesture, touch) was smooth and seamless.	3	4	3	4	4
The "zaps" reward system made saving energy feel more engaging.	4	4	4	5	4
The app provided useful feedback about my energy-saving behaviour.	5	4	4	4	5
I would recommend this app to my family and friends.	5	4	4	5	4

FEEDBACK INTERVIEW ON USABILITY, MULTIMODAL INTERACTION, ENGAGEMENT & GAMIFICATION

Open ended qualitative questions	Participant 1	Participant 2	Participant 3	Participant 4	Participant 5
What did you like the most about the app?	Difference in engagement based on age group. Really well thought process	The overall idea of the app	I get to know the daily usage and how much power each device use	The interactive features like zaps	Dashboard has a comprehensive data and is very well structured
What features or interactions did you find confusing or difficult to use?	I don't think there was any part where the feature or interaction which was difficult to use	Nothing as such	There was no back button you must navigate through below banner	I did not find anything confusing in the app.	Tab menu, the icon for devices could be easily mis-interpreted as homepage
How could the app's multimodal interactions (voice, gesture, touch) be improved?	The features to be more explicitly marked to identify the same but to the validation of the app there doesn't seem to be any major gap.	It could have more features	Rather than clicking on each voice button there should be a voice button on the top or bottom so it can be used for throughout the app	It could be improved by performing more tasks related to the app when you are away from home, like putting all devices in sleep mode.	The voice interaction, says "to begin click button" but did not mention what is the button label
Did the game-based rewards (zaps, leader boards) encourage you to save energy? Why or why not?	Yes, it will help users to be encouraged to use the app. It's helpful to take this up further.	Yes I liked the zaps idea a lot	For the kids it's fun, engaging and interactive	Did not really like it as an adult. It would be engaging for kids for sure.	I find leader board rather encouraging to save energy
Any additional suggestions for improving the app?	Colour scheme is tagged to the eco mode. The forward backward navigation can be further improved.	No	Text little bit bigger	I suggest adding a back button to navigate back to the home page easily.	To recommend the app to a friend in manage account section was a bit odd!

SYSTEM USABILITY SCALE (SUS) QUESTIONNAIRE

QUESTIONS	PARTICIPANT 1	PARTICIPANT 2	PARTICIPANT 3	PARTICIPANT 4	PARTICIPANT 5
I think that I would like to use this system frequently	4	4	4	5	4
I found the system unnecessarily complex.	1	2	2	1	1
I thought the system was easy to use.	5	4	4	4	5
I think that I would need the support of a technical person to use this system.	1	2	2	1	1
I found the various functions in this system were well integrated.	5	4	4	4	4
I thought there was too much inconsistency in this system.	1	1	2	2	1
I would imagine that most people would learn to use this system very quickly.	5	4	4	5	4
I found the system very cumbersome to use.	3	2	3	2	1
I felt very confident using the system.	5	4	4	5	5
I needed to learn a lot of things before I could get going with this system.	2	1	2	2	1