



ΣΠΟΥΔΑΙ / SPOUDAI

Journal of Economics and Business
<http://spoudai.org>



Beyond Self-Report: A Neurocognitive Approach to UX Evaluation of Greek Municipal Websites

Olga Tsatsani^a, Yannis Pollalis^b

^a PhD Candidate, Department of Economics, Piraeus University, Greece.

Email: tsn.olga@gmail.com

^b Professor, Department of Economics Piraeus University, Greece.

Email: yannis@unipi.gr

Abstract

Despite growing investment in e-government infrastructure, the adoption of local municipal digital services remains limited, raising questions about how user experience (UX) is currently evaluated. Traditional approaches, based on expert audits and self-reported measures, capture primarily the conscious and observable layers of interaction, overlooking the underlying cognitive and emotional processes that shape actual user behavior. This study introduces a multimodal, neuroscience-informed framework for UX evaluation and presents, on a pilot basis, an exploratory application of the Neurocognitive Engagement (NCE) Index to three Greek municipal websites representing bad, medium, and good UX quality. A pilot experiment (N=10) was conducted combining electroencephalography (EEG), a Post-Session Recognition Memory Task (PSRMT), eye-tracking behavioral data, and self-reported measures. Findings indicate that while self-reports differentiated UX Conditions, the NCE Index revealed near-identical engagement between medium and good UX (5.20 vs. 5.07), despite substantial differences in audit scores. These preliminary findings suggest that conventional UX evaluation methods may not fully capture the underlying neurocognitive processes associated with user interaction, highlighting the potential value of integrating neuroscience-informed measures into existing UX evaluation frameworks. The study presents exploratory results as a proof-of-concept for a novel methodological framework. Future research with larger, gender-balanced samples is needed to validate and extend these preliminary findings. The study carries practical implications for the evidence-based allocation of public digitalization budgets and the adoption of neurocognitive methods in applied public-sector UX evaluation.

JEL Classification: C87, C89, C91, M31

Keywords: user experience, e-government, multimodal evaluation

1. Introduction

In the broader context of public sector digitization, the development of effective e-government services has become a strategic priority for both the European Commission and national governments, including Greece (EC, 2010). A central goal of this transition has been the creation of transparent, efficient, and citizen-oriented digital interfaces that strengthen

democratic participation and reconnect citizens with public administration. This need became even more urgent during and after the COVID-19 period, when digital access to public services was no longer optional but essential.

At the same time, the expansion of e-government infrastructure has not automatically translated into sustained citizen use. Despite substantial public investment in digital governance, the adoption of e-government services remains limited (Perez-Morote et al., 2020). Prior studies indicate that satisfaction, trust, and perceived service quality are strongly associated with citizens' willingness to use digital government platforms (Perez-Morote et al., 2020; Zheng & Schachter, 2017; Nam, 2014; Reddick, 2011). This positions user experience (UX) as a central factor in the success of digital public services.

In the context of e-democracy, municipal websites play a critical role as the primary digital interface between citizens and local governments, supporting access to services, information, and participation (Rodríguez Bolívar, 2016; Ali, 2023). However, in Greece, despite progress in centralized e-government platforms such as gov.gr (Dimitreloy and Fouskas, 2023; Spinellis et al., 2022), local municipal websites often remain outdated, fragmented, and poorly maintained (Patergiannaki and Pollalis, 2022; Georgios and Nikolaos, 2021). These limitations are frequently linked to structural constraints, including limited budgets, decentralized governance, and a lack of specialized digital expertise (Anastasopoulou et al., 2025; Chairistanidis and Rusu, 2024; Chourmouziadis, 2024; Choudrie et al., 2018; Katsikas and Gritzalis, 2017).

Beyond structural issues, practical constraints in UX research further exacerbate the problem. In real-world professional environments, UX practitioners often operate under severe time and resource limitations, leading to design decisions that are not systematically validated. Business practice has shown that traditional UX methods—such as usability testing, interviews, and surveys—are frequently perceived by stakeholders as time-consuming, costly, or overly subjective. As a result, they are often underutilized or entirely omitted, increasing the likelihood of arbitrary design decisions and digital products that require costly revisions after implementation.

At the same time, user expectations toward digital services have significantly evolved. Digitally experienced users, particularly younger generations such as Millennials and Generation Z, expect fast, intuitive, and seamless interactions, while older users increasingly depend on digital platforms for accessing essential services and require clarity and accessibility (Fernandes et al., 2013). This growing diversity of expectations further amplifies the importance of designing municipal platforms that are not only functional but also cognitively and emotionally aligned with user needs.

Building on this context, our previous large-scale study examined 138 Greek municipal websites through an adapted UX Audit framework, combined with citizen-oriented evidence on expectations and frustrations (Tsatsani and Pollalis, 2025; 2024). The findings revealed widespread weaknesses in navigation, task orientation, multilingual accessibility, content relevance, search functionality, and interactive support, alongside significant regional disparities and a broader mismatch between citizen expectations and platform performance. Overall, municipal websites were found to operate primarily as static information repositories rather than as intuitive and participatory digital environments.

While such audit-based and survey-based approaches are valuable for identifying structural usability issues, they also present inherent limitations. Traditional UX methods rely heavily on self-reported data and expert judgment, which may fail to capture users' real-time cognitive and emotional responses during interaction. Consequently, critical experiential dimensions—

such as cognitive load, attentional processes, emotional engagement, and implicit memory—remain insufficiently explored.

Prior research has emphasized that trust, transparency, and perceived system quality are essential for the adoption of digital public services (i.e. Nulhusna et al., 2017; Alzahrani et al., 2017; Zuiderwijk et al., 2015). At the same time, studies have shown that even digitally skilled users may struggle to locate information or complete tasks when usability is poor (Fernandes et al., 2013), reinforcing the importance of experience-driven evaluation approaches.

Despite these insights, the application of neurophysiological and behavioral methods in e-government UX research remains limited. One of the few relevant studies is that of Albayrak and Çağıltay (2013), who combined eye-tracking with qualitative methods to examine user interaction with governmental websites, identifying fixation patterns associated with system visual complexity and reduced task efficiency. However, such approaches remain only recent and rare, particularly in municipal contexts (i.e. Zhang et al., 2021; Yildiz et al., 2020; Bataineh and Mourad, 2017; Demir et al., 2017), and the integration of neurocognitive data into UX evaluation frameworks is still underexplored.

Responding to this gap, the present study extends previous UX Audit findings by moving from structural evaluation to neurophysiological and behavioral investigation. It adopts a multimodal approach combining electroencephalography (EEG), eye-tracking, an implicit memory test, and self-reported measures to examine whether municipal websites previously classified as good, medium, or bad in UX quality also produce different cognitive and experiential responses. In doing so, the study explores the potential of neuroscience-informed methods as a more objective and efficient complement to traditional UX evaluation, with reduced reliance on large samples and time-intensive procedures.

The theoretical background that follows is organized around three core issues. First, it examines the importance of UX in public sector platforms, focusing on local e-government effectiveness and age-related expectations in digital behavior. Second, it discusses the limitations of traditional UX research methods, particularly self-reporting bias and the absence of unconscious data. Third, it explores how cognitive psychology and neuroscience contribute to advancing UX evaluation in governmental contexts.

2. Literature Review

2.1 User Experience and the Effectiveness of Local e-Government

User experience (UX) in public sector websites is not merely a matter of convenience; it is closely tied to digital accessibility, equity, trust, and democratic participation. Municipal websites, in particular, function as the primary digital interface between citizens and local governments, supporting access to information, administrative procedures, and civic engagement opportunities (OECD, 2020; Rodríguez Bolívar, 2016; Ali, 2023). In this sense, local e-democracy depends not only on the existence of digital platforms, but also on the quality of the interaction they enable.

Previous literature (i.e. Camilleri, 2020; Carter and Bélanger, 2005; Welch et al., 2005) has consistently shown that the effectiveness of e-government services cannot be reduced to technical availability alone. The extent to which citizens use and trust digital public services is strongly influenced by how usable, transparent, and relevant these services are perceived to be (Perez-Morote et al., 2020; Nulhusna et al., 2017; Alzahrani et al., 2018; Zuiderwijk et al., 2015). When public websites are difficult to navigate, poorly structured, or lacking in interactivity, citizens may experience confusion, frustration, and disengagement, which in turn

weakens both service adoption and institutional trust (Galvez and Youngblood, 2016; Jun et al., 2014; Voutinioti, 2013; Scott, 2006). Conversely, well-designed municipal websites can streamline bureaucracy, improve transparency, and enhance the inclusiveness of local governance.

This issue is particularly relevant in the Greek context. Although Greece has made visible progress in centralized e-government systems such as gov.gr, local municipal websites often remain fragmented in quality, uneven in digital maturity, and weak in service depth (Dimitreloy and Fouskas, 2023; Spinellis et al., 2022; Patergiannaki and Pollalis, 2024; Georgios and Nikolaos, 2021). Earlier studies have highlighted problems related to outdated infrastructures, limited interactivity, poor content relevance, and inconsistent alignment with citizen needs (Anastasopoulou et al., 2025; Chairistanidis and Rusu, 2024; Chourmouziadis, 2024; Choudrie et al., 2018; Katsikas and Gritzalis, 2017). Technical and heuristic evaluations of Greek municipal websites have also pointed to recurring weaknesses in navigation, accessibility, and trust-related design elements (Patergiannaki et al., 2020; Panayiotou and Stavrou, 2019; Chatzopoulos, 2009;).

These concerns were further reinforced by our previous large-scale UX Audit of 138 Greek municipal websites, which revealed widespread deficits across core usability dimensions (Tsatsani and Pollalis, 2024). Taken together, this evidence suggests that many municipal websites still operate as static repositories of information rather than as user-centered and participatory public service platforms.

These considerations motivate Research Question 1 (RQ1), addressed in Section 1.5, which asks whether municipal websites with higher UX quality are associated with more positive user experience outcomes.

2.2 Age Expectations and Digital Behavior in UX Contexts

The importance of UX in public platforms is further amplified by changing user expectations across age groups. Contemporary digital users, especially Millennials and Generation Z, have developed high standards for speed, clarity, responsiveness, and ease of use through repeated exposure to mature private-sector digital ecosystems (Süleymanov, 2025; Zaelani et al., 2025). As a result, they are quick to detect friction in online environments and are less tolerant of confusing structures, delays, or inefficient navigation. When a system fails to meet these expectations, abandonment may occur almost immediately, often accompanied by negative judgments about the service provider itself.

At the same time, older adults are increasingly engaged with digital platforms in their everyday lives, especially when accessing essential services such as banking, health, or public administration (Xie et al., 2021; Hodge et al., 2017). Although they may display somewhat greater patience than younger users, they also require clear information structures, predictable workflows, and accessible interfaces. In governmental contexts, where the stakes of use are often higher than in commercial environments, poor UX may affect not only convenience but also the ability to complete necessary tasks or access public support.

This means that municipal websites must address a broad spectrum of digital behaviors and expectations. Designing only for digitally fluent users risks excluding older or less confident citizens, while designing only for simplicity without considering contemporary standards risks alienating younger users who expect efficient and intuitive experiences. Research has shown that even highly educated and digitally experienced users may struggle with poorly designed governmental systems, suggesting that usability barriers are not merely a matter of user skill, but of interface quality and system logic (Fernandes et al., 2013). Therefore, effective public-

sector UX requires balancing inclusiveness, clarity, efficiency, and relevance across diverse user groups.

These observations motivate Research Question 2 (RQ2), addressed in Section 1.5, which asks whether differences in UX quality are reflected in users' interaction patterns and engagement levels across diverse user groups, regardless of digital familiarity.

2.3 Limitations of Traditional UX Research Methods

Traditional UX research has made an important contribution to the evaluation of digital systems through methods such as surveys, interviews, usability tests, and expert reviews (i.e. Ardito et al., 2014; Duman and Salzman, 2006). However, these methods are also characterized by significant limitations, especially when the goal is to understand how users actually experience a digital interface in real time (i.e. Albert and Tullis, 2022; Allam and Dahlan, 2013; Schwarz, 2012). Most traditional approaches rely heavily on self-reported perceptions or retrospective accounts, which are shaped by memory bias, social desirability, interpretation errors, and users' limited ability to articulate unconscious emotional or cognitive states (Tsatsani and Pollalis, 2025).

This limitation is especially important in public-sector environments, where usability failures often occur under conditions of time pressure, confusion, or cognitive overload. A citizen may report that a website was "acceptable" overall while still experiencing frustration, attentional disruption, or mental fatigue during specific tasks. As a result, self-reported satisfaction does not always reflect the actual moment-to-moment experience of interacting with a platform. This creates a methodological gap between what users say and what they cognitively and behaviorally undergo (Hekler et al., 2013; Schwarz, 2012).

The problem is further intensified in professional practice. In applied UX environments, especially in public or resource-constrained settings, research teams often face limited budgets, short timelines, and low stakeholder tolerance for complex or repeated testing (Ardito et al., 2014; Kuniavsky, 2003). Traditional UX methods are therefore frequently perceived as too slow, too costly, or too subjective to support rapid decision-making. In many cases, this leads to minimal validation, small convenience samples, or design decisions based primarily on assumptions. Ironically, the avoidance of research for reasons of efficiency may ultimately result in ineffective implementations, costly redesigns, and lower user adoption after launch.

In the case of municipal websites, these constraints are particularly problematic. Such platforms serve heterogeneous populations, support essential tasks, and carry strong implications for trust in public administration (Feeney and Brown, 2017). Yet their evaluation is often reduced to technical compliance or superficial usability review, leaving deeper experiential dimensions underexplored. Thus, while traditional UX methods remain useful for identifying structural problems, they are often insufficient for capturing the full cognitive and emotional complexity of digital public service interaction (Yang, 2025).

These insights motivate Research Question 3 (RQ3), addressed in Section 1.5, which asks to what extent traditional self-reported UX measures capture users' real-time cognitive and behavioral responses during interaction with municipal websites.

2.4 The Role of Neurocognitive Methods

Addressing this limitation requires moving beyond conscious evaluation and examining the underlying cognitive processes that shape user experience. From a cognitive psychology perspective, user experience is not solely a conscious process. Digital interactions generate implicit cognitive and emotional responses that shape perception, trust, and future behavior. These responses are often encoded in implicit memory, defined as the unconscious retention of

experiences that influence judgments and decisions without deliberate awareness (Tucker et al., 2022; Tobias et al., 2014). Importantly, users may not be able to explicitly recall or verbalize these experiences, even though they significantly impact their interaction with a system.

To address these limitations, UX research has increasingly incorporated neuroscientific and behavioral methods that capture real-time user responses. Electroencephalography (EEG), for example, enables the monitoring of brain activity with high temporal resolution, allowing researchers to assess cognitive load, attentional engagement, and emotional activation as they occur (Dan and Reiner, 2017; Antonenko et al., 2010). Specific frequency bands, such as theta, alpha, and beta, provide measurable indicators of mental effort, attentional states, and task engagement. Crucially, EEG captures rapid, unconscious processes that remain inaccessible through self-report methods, offering a more direct reflection of the user's internal state during interaction (Bell et al., 2018).

In parallel, eye-tracking technology provides detailed insights into users' visual behavior. By recording gaze patterns, fixation durations, and scan paths, eye-tracking reveals how users allocate attention, process information, and navigate digital environments. It allows researchers to identify whether key elements are easily discoverable, whether users experience confusion or hesitation, and how efficiently they complete tasks. Unlike verbal feedback, which is retrospective and subjective, eye-tracking captures observable behavioral evidence of interaction in real time (Macdonald and Tatler, 2018).

Beyond neural and visual attention measures, recognition memory testing offers an additional layer of insight into how digital content is encoded and retained. By assessing whether users unconsciously recognize or recall elements encountered during interaction, implicit memory measures reveal the extent to which information has been meaningfully processed (Greenwald and Banaji, 2017; Ellis, 2005). This is particularly relevant in public-sector contexts, where the perceived relevance and memorability of content influence trust, usability, and continued engagement.

While some studies have raised concerns regarding the reliability of implicit measures and their comparative advantage over self-reported methods (Corneille and Gawronski, 2024), in the present study implicit memory is employed as a complementary metric alongside neurophysiological, behavioral, and subjective measures, contributing to a more comprehensive assessment of user experience.

When combined, these methods provide a multidimensional understanding of user experience that goes beyond the limitations of traditional approaches. EEG captures internal cognitive and emotional processes, eye-tracking reflects external behavioral interaction, and implicit memory reveals downstream cognitive encoding. Together, they enable the triangulation of user experience across complementary levels of analysis, reducing reliance on subjective reporting and increasing the robustness of findings.

This multimodal approach is especially valuable in the evaluation of public-sector digital platforms, such as municipal websites, where usability directly affects access to services, citizen participation, and institutional trust. Previous research in e-government contexts remains limited in its use of such methods. As such, the integration of neurophysiological, behavioral, and cognitive measures in a **unified e-governmental framework** remains largely underexplored.

Addressing this gap, the present study adopts a multimodal, neuroscience-informed approach to UX evaluation, combining EEG, eye-tracking, recognition memory testing, and self-reported measures. By doing so, it seeks to capture both consciously reported and less consciously

accessible dimensions of user experience, contributing preliminary evidence on how users interact with municipal digital platforms. Research Questions 4 through 7, addressing the contribution of each modality individually and in combination, are presented in Section 1.5.

2.5 Research Questions and Study Objectives

Building on the theoretical background outlined above, this pilot study is organized around seven research questions rather than confirmatory hypotheses. This reframing reflects the exploratory, proof-of-concept nature of the study and the small pilot sample ($N = 10$), which does not support hypothesis testing in the strict inferential sense. The research questions are:

RQ1. Is municipal website UX quality, as classified through the prior UX Audit, associated with more positive user experience outcomes across behavioral and perceptual responses?

RQ2. Are differences in UX quality reflected in users' interaction patterns and engagement levels, regardless of digital familiarity or generational cohort?

RQ3. To what extent do traditional self-reported UX measures capture users' real-time cognitive and behavioral responses during interaction with municipal websites?

RQ4. Do neurophysiological measures obtained through EEG capture variation in users' cognitive and emotional responses associated with differences in UX quality?

RQ5. Do behavioral efficiency measures obtained through eye-tracking reflect differences in attention allocation and navigation efficiency across websites of varying UX quality?

RQ6. Does recognition memory performance, measured through the Post-Session Recognition Memory Task (PSRMT), and cognitive encoding of website content vary according to UX quality?

RQ7. Does a multimodal approach combining neurophysiological, behavioral, cognitive, and self-reported measures provide a more comprehensive assessment of user experience than any single method alone?

Given the exploratory nature of the pilot design, these research questions are addressed descriptively, using non-parametric statistics where appropriate, rather than through formal hypothesis testing with the current sample size.

Building on the limitations of traditional UX research and the emerging role of neurocognitive methods, the present study aims to investigate how differences in the UX quality of municipal websites are reflected in users' cognitive, behavioral, and experiential responses. Specifically, the study seeks to examine whether websites previously classified as having good, medium, or bad UX produce distinct patterns of interaction, attention, and information processing. To achieve this, a multimodal research framework is employed, integrating electroencephalography (EEG), eye-tracking, recognition memory testing, and self-reported measures. This approach allows for the simultaneous examination of consciously reported and less consciously accessible dimensions of user experience, moving beyond the limitations of single-method evaluations.

Within this context, this study also introduces, on a pilot and exploratory basis, the **Neurocognitive Engagement (NCE) Index**, a composite metric that integrates neurophysiological, behavioral, and cognitive indicators (EEG, eye-tracking, recognition memory performance, and self-reports) into a single evaluative framework. The NCE Index is presented as a proof-of-concept step toward a more holistic assessment of user interaction, intended to inform, rather than replace, design and decision-making in digital public services. Its formula, component operationalization, and limitations are fully disclosed in Section 2.6.

3. Methods

3.1 Study Design

This study adopts a mixed experimental design to evaluate the user experience (UX) of Greek municipal websites through a combination of neuroscientific, behavioral, cognitive, and self-reported measures. The research consists of two distinct but complementary data collection phases, each targeting different dimensions of user experience.

In the first phase (EEG session), a within-subjects design was employed. All participants were exposed to static images from three municipal websites representing distinct levels of UX quality, namely good, medium, and bad, as determined by a prior large-scale UX Audit (Tsatsani & Pollalis, 2024; 2025). EEG data were recorded continuously while participants observed the stimuli across three experimental blocks, each corresponding to a different cognitive context (free exploration, goal-oriented search, and emergency scenario). This design enables within-subject comparisons of neurophysiological responses across UX quality conditions and contextual framings.

In the second phase (eye-tracking session), a between-subjects design was applied. Each participant was randomly assigned to interact with one of the three municipal websites and completed a standardized set of tasks simulating realistic citizen interactions. During task execution, visual behavior was recorded using eye-tracking technology. Following task completion, participants completed an Implicit Memory Test (IMT) and a self-reported questionnaire, capturing both cognitive and subjective evaluations of the experience. This between-subjects structure allows for comparison of behavioral and perceptual responses across participants interacting with websites of differing UX quality, while minimizing learning effects.

Together, the two phases enable the triangulation of conscious and unconscious dimensions of user experience. The multimodal and mixed-design approach supports both within-subject comparisons (in the EEG phase) and between-condition comparisons (in the eye-tracking phase), providing a comprehensive and ecologically valid evaluation of UX in local e-government platforms.

3.2 Participants

The sample consisted of **N = 10 adult male participants**, recruited through convenience sampling. The inclusion of only male participants was a deliberate methodological decision to control for documented gender-related variability in baseline neurophysiological activity and EEG band power distributions (Klimesch, 1999; Harmon-Jones & Gable, 2018; Cahill, 2006). The authors acknowledge that this decision, while reducing a potential confound within a small pilot sample, limits the ecological validity of the findings and precludes conclusions regarding gender differences in neurocognitive UX responses; this is addressed explicitly as a limitation in Section 4.10 (L1). Future studies should recruit gender-balanced samples.

Participants were distributed across three generational cohorts: Generation Z (18–27 years), Millennials (28–43 years), and older adults (44–55+ years), ensuring coverage of a range of digital experience levels and interaction patterns. All participants provided written informed consent prior to participation. Participant characteristics are summarized in Table 1.

Table 1. Participant Characteristics (N=10)

Part.	Age	Cohort	Education	Digital Literacy	Municipal Site Use	Assigned Website (eye-tracking)
P1	28–35	Millennial	University	High	Occasional	Mytilene (Medium UX)
P2	36–43	Millennial	University	High	Occasional	Mytilene (Medium UX)
P3	28–35	Millennial	University	High	Rare	Alexandroupolis (Good UX)
P4	44–55	Older Adult	University	Moderate	Occasional	Alexandroupolis (Good UX)
P5	18–27	Gen Z	University	High	Rare	Mytilene (Medium UX)
P6	44–55	Older Adult	Secondary	Moderate	Occasional	Didymoteicho (Bad UX)
P7	55+	Older Adult	Secondary	Low	Frequent	Didymoteicho (Bad UX)
P8	18–27	Gen Z	University	High	Rare	Alexandroupolis (Good UX)
P9	28–35	Millennial	Postgraduate	High	Occasional	Mytilene (Medium UX)
P10	36–43	Millennial	University	Moderate	Occasional	Alexandroupolis (Good UX)

Note. All participants were adult males. Digital literacy is self-reported (Low/Moderate/High). Website assignment refers to the eye-tracking session (between-subjects).

3.3 Stimuli and Materials

The experimental stimuli consisted of three Greek municipal websites, each representing a distinct level of UX quality, as classified through a prior UX Audit of over 138 municipal websites using a nine-dimension evaluation framework (Tsatsani & Pollalis, 2024; 2025). UX categories were defined based on composite scores (0–100%), with thresholds set at: **good UX (>71%)**, **medium UX (50–70%)**, and **bad UX (<50%)**. The three selected municipalities were: the **Municipality of Alexandroupolis** (good UX, score: 84%), the **Municipality of Mytilene** (medium UX, score: 59%), and the **Municipality of Didymoteicho** (bad UX, score: 28%).

For the EEG session, standardized full-page images of each website's homepage were prepared as static visual stimuli. For the eye-tracking session, participants interacted with the live website assigned to their condition via a standardized browser interface.

The EEG session used static homepage screenshots to ensure stimulus control and inter-participant comparability. Static images capture early visual appraisal responses rather than full interactive navigation. The neurophysiological findings should therefore be interpreted as indicators of first-impression processing; the eye-tracking phase separately captures interactive behavioral engagement. These two phases measure complementary but distinct aspects of user experience and are not treated as interchangeable in the interpretation of results (Frey et al., 2016; Bellos et al., 2025).

3.4 Procedure

The study was conducted in a quiet, controlled environment to minimize external interference and ensure data quality. Each session lasted approximately 45 minutes per participant and was organized into two sequential phases.

3.4.1 Phase 1: EEG Session

EEG data were recorded using the **Emotiv Insight headset** (5 channels: AF3, T7, Pz, T8, AF4; sampling rate: 128 Hz), a portable, non-invasive device suitable for field-based research. Prior to stimulus presentation, a calibration phase was conducted, comprising a standard eyes-open fixation period (15 seconds) and an eyes-closed resting period (15 seconds) to establish individual baseline neurophysiological activity.

The experimental stimuli were presented using the Emotiv Pro Builder platform. Each participant was exposed to all nine images (three per UX condition) across three experimental blocks, each preceded by a brief instruction designed to prime a specific cognitive orientation:

- **Block 1 (Free Exploration):** Participants observed each image freely, focusing on general visual impression.
- **Block 2 (Goal-Oriented Task):** Participants imagined completing a specific administrative task (e.g., submitting an application or locating a municipal service).
- **Block 3 (Emergency Scenario):** Participants imagined a time-sensitive situation (e.g., water outage) and attempted to locate relevant information rapidly.

Each trial followed a standardized sequence: a **fixation cross (+)** presented at the centre of the screen for baseline stabilization, followed by **stimulus presentation for 8 seconds**, and a **2-second inter-stimulus interval** before the next trial. The order of stimuli within each block was fixed and consistent across participants to ensure comparability.

Following the EEG recording, participants completed a **Self-Assessment Manikin (SAM)-based questionnaire**, rating each of the nine stimuli on two dimensions: *Emotional Valence* (pleasant–unpleasant) and *Arousal* (low–high activation), using a 7-point Likert scale. These self-reported ratings were subsequently used as input for the Emotional Intensity (EI) component of the NCE Index.

3.4.2 Phase 2: Eye-Tracking Session

In the second phase, participants were randomly assigned to interact with one of the three municipal websites. Eye-tracking data were recorded using an ASL Mobile Eye-XG eye-tracker (Applied Science Laboratories), consisting of two digital high-resolution cameras attached to lightweight eyeglasses: one camera tracked and recorded the movement of the right eye (monocular, dark-pupil tracking technique), while the second recorded the scene image. Data were recorded at 30 Hz with a manufacturer-reported accuracy of 0.5–1.0° of visual angle. The device features embedded firmware, an integrated microphone, automatic calibration with calibration validation and post-calibration checks, outdoor enhancement, automatic data mapping, a parallax compensation tool, and synchronization capability with external sources. A calibration procedure was carried out for each participant prior to task execution to ensure recording accuracy. No participant was excluded on data-quality grounds.

Participants completed **four standardized tasks** designed to simulate realistic citizen interactions and elicit different cognitive and behavioral responses. Task instructions were identical across all three website conditions to ensure comparability of findings:

- **Task 1 — Exploratory Browsing:** Participants were asked to freely navigate the homepage and form an initial overall impression of the website, without a specific goal.

- **Task 2 — Information Retrieval:** Participants were instructed to locate content related to municipal cleaning and recycling services, assessing the discoverability and clarity of service-related information.
- **Task 3 — Online Request Submission:** Participants were asked to identify how to submit an online request or application, reflecting one of the most common citizen interaction needs in the context of local e-government services (Tsatsani & Pollalis, 2025).
- **Task 4 — Language Accessibility:** Participants were required to locate and activate an English-language version of the website, evaluating the accessibility and discoverability of multilingual support features.

Eye-tracking recordings were captured as video files and subsequently subjected to manual coding for the identification of Areas of Interest (AOIs), fixation counts, fixation durations, and time-to-first-fixation metrics. AOIs were defined a priori based on standard municipal website structural zones (header/navigation, hero content, service links, footer), with task-specific AOIs added for each task. Fixations were identified using a standard dispersion threshold ($< 1^\circ$ visual angle, minimum 100 ms duration). Coding was conducted by a single trained coder; inter-coder reliability was not assessed. This is acknowledged as a limitation of the pilot implementation (Section 4.10); future studies should employ independent double-coding with kappa reliability reporting, or automated fixation extraction, to increase precision.

NB. These behavioral indicators were used to derive the **Relevance (R)** component of the NCE Index.

3.4.3 Post-Session Measures

Following the eye-tracking session, participants completed two additional measures administered via the Emotiv Pro Builder platform:

- **Post-Session Recognition Memory Task (PSRMT):** Following the eye-tracking session, participants completed a Post-Session Recognition Memory Task (PSRMT) administered via the Emotiv Pro Builder platform — a recognition task consisting of 29 terms (15 targets, 14 distractors). While the recognition format requires a conscious old/new judgment and therefore does not constitute an implicit memory measure in the classical priming sense (Schacter, 1987), the task was designed as an incidental-encoding measure: participants were not instructed to memorize content during the preceding session, and target terms were embedded within task-relevant content rather than presented for deliberate study. Recognition performance therefore reflects the strength of memory traces formed through incidental rather than deliberate encoding, consistent with the literature on incidental memory and its relationship to less consciously mediated processing (Tucker et al., 2022; Greenwald & Banaji, 2017; Ellis, 2005). Accuracy and response patterns were recorded and used to calculate a Memory Retention (MR) score based on Signal Detection Theory (Hit Rate – False Alarm Rate), scaled to a 0–10 range (Wixted, 2020; Macmillan, 2002; Macmillan and Kaplan, 1985). The authors acknowledge that process-dissociation procedures would be needed in future work to more precisely isolate implicit from explicit contributions to recognition performance.
- **Self-Reported Questionnaire (SRQ):** A 7-item questionnaire assessing subjective experience of the assigned website on a 7-point Likert scale (1 = strongly disagree / very low, 7 = strongly agree / very high). Items covered: ease of use, information findability, content clarity, emotional engagement, content relevance, reuse intention, and perceived cognitive fatigue. Item 7 (cognitive fatigue) was reverse-scored, and all items were scaled to a 0–10 range for inclusion in the NCE Index.

3.5 EEG Data Processing and Analysis

Raw EEG data were exported from the Emotiv Pro platform in CSV format, retaining continuous time-series recordings alongside event markers corresponding to stimulus onset and offset timestamps. Data processing and analysis were conducted using custom Python scripts (pandas, NumPy, SciPy).

3.5.1 Stimulus Window Extraction

Individual stimulus windows were extracted using the event markers recorded in each participant's *intervalMarker.csv* file, which contained precise latency and duration timestamps for each stimulus element. Absolute timestamps were computed by adding stimulus latency values to each participant's session start timestamp. EEG data rows falling within each stimulus window (onset to offset) were selected for analysis, yielding **27 stimulus windows per participant** (9 images $\times$ 3 blocks), and 270 windows in total across the sample.

3.5.2 Band Power Extraction

For each stimulus window, mean spectral power values were extracted across five frequency bands, namely *Theta* (4–8 Hz), *Alpha* (8–13 Hz), *Beta-Low* (13–20 Hz), *Beta-High* (20–30 Hz), and *Gamma* (30–45 Hz), across all five EEG channels (AF3, T7, Pz, T8, AF4), using Emotiv's pre-computed band power output (POW columns). These values represent the mean oscillatory power within each frequency band during each stimulus exposure. Oscillatory power in these frequency bands is widely used in cognitive neuroscience as **an indirect indicator** of underlying mental processes, such as **attention allocation**, **cognitive effort**, and **emotional engagement**, enabling the interpretation of neurophysiological responses in relation to user experience.

3.5.3 Derived EEG Metrics

Three composite metrics were derived from the band power data to serve as neurophysiological indicators within the NCE Index:

- **Frontal Theta Power (Cognitive Load indicator):** Computed as the mean of AF3.Theta and AF4.Theta across stimulus windows. Frontal Theta activity is associated with increased cognitive effort and working memory engagement (i.e. Yu et al., 2022).
- **Parietal Alpha Power (Relaxation/Disengagement indicator):** Derived from Pz.Alpha. Alpha oscillations in parietal regions are inversely related to attentional engagement and reflect cognitive relaxation states (i.e. Bonnefond and Jensen, 2025; Chen et al., 2020).
- **Frontal Beta-Low Power (Engagement proxy):** Computed as the mean of AF3.BetaL and AF4.BetaL. Beta-Low activity in frontal regions is associated with active cognitive processing and task engagement (i.e. Halderman et al., 2021; Nuamah and Seong, 2018).
- **Theta/Alpha Ratio (Cognitive Engagement Index):** Computed as Frontal Theta / (Parietal Alpha + 0.001) to avoid division by zero. This ratio provides a composite indicator of cognitive load relative to attentional state, with higher values indicating greater mental effort (i.e. Kamzanova et al., 2015).

All derived metrics were normalized to a 0–10 scale across the full sample prior to inclusion in the NCE Index. The Cognitive Load metric was additionally inverted ($CL_inverted = 10 - CL_scaled$), so that higher values consistently indicate a more positive user experience (i.e., lower cognitive effort).

3.5.4 Statistical Analysis of EEG Data

Given the small sample size ($N = 10$) and the non-normal distribution of EEG band power values, non-parametric statistical tests were employed. Differences in EEG metrics across UX conditions (Bad, Medium, Good) were assessed using the Friedman test for repeated measures, and effect sizes were computed as Kendall's coefficient of concordance (W), the standard companion effect-size statistic for the Friedman test, calculated as $W = \chi^2/[N(k-1)]$. Post-hoc pairwise comparisons were conducted using Wilcoxon signed-rank tests with appropriate correction for multiple comparisons. All analyses were performed in Python (SciPy). Results are reported with test statistics, p-values, and effect sizes. Given the exploratory and pilot nature of the study, a significance threshold of $\alpha = .05$ was applied, while non-significant trends are reported descriptively and interpreted with appropriate caution.

3.6 The Neurocognitive Engagement (NCE) Index

This study introduces the **Neurocognitive Engagement (NCE) Index**, a novel composite metric developed by the authors to integrate neurophysiological, behavioral, cognitive, and self-reported indicators of user experience into a single, unified evaluative score. The NCE Index is presented as a pilot-stage, proof-of-concept framework, intended to demonstrate the feasibility of multimodal UX integration rather than to provide a validated or finalized evaluation tool. The framework is proposed as a more objective and comprehensive complement to traditional single-method UX evaluation approaches, specifically designed to capture both conscious and unconscious dimensions of user experience. The NCE Index is structured around six core dimensions, each targeting a distinct layer of user experience:

Table 1. NCE Index Components, Definitions and Data Sources

Component	Full Name	Description	Data Source
EI	Emotional Intensity	Captures the emotional quality and intensity of the user's response to the interface	SAM questionnaire (post-EEG)
ENG	Engagement	Reflects the degree of active cognitive and neural engagement during interface exposure	EEG neurophysiological data
CL	Cognitive Load	Assesses the mental effort required to process and navigate the interface	EEG neurophysiological data
R	Relevance	Evaluates the perceived meaningfulness and utility of the interface content, based on implicit behavioral indicators	EEG + IMT + Eye-tracking
MR	Memory Retention	Measures the implicit encoding and retention of content encountered during interaction	Implicit Memory Test (IMT)
SRQ	Self-Reported Questionnaire	Captures the participant's subjective evaluation of the overall interaction experience	Post-session questionnaire

NB. The NCE Index integrates all five components into a single composite score. Full technical specifications regarding component derivation and aggregation are **subject to intellectual property protection** (see note below).

The resulting NCE score ranges from 0 to 10, with higher values indicating more positive and cognitively engaged user experiences. For interpretive purposes, the following qualitative

thresholds are applied: **Low Engagement (0–4)**, namely the digital environment fails to elicit meaningful interest or cognitive involvement; **Moderate Engagement (4–6)**, namely the interface partially meets user needs but presents barriers to full engagement; **Good Engagement (6–8)**, namely the system responds effectively to user needs with room for improvement; **High Engagement (8–10)**, the interface is highly comprehensible, relevant, and cognitively engaging.

3.6.1 Formula and Weighting

In this pilot study, all six components are weighted equally (1/6 each), yielding the following formula: $NCE = (EI \times 1/6) + (ENG \times 1/6) + (CL^{-1} \times 1/6) + (R \times 1/6) + (MR \times 1/6) + (SRQ \times 1/6)$. Equal weighting was adopted as a conservative, theoretically neutral starting point for this pilot, pending empirical validation of differential component contributions. A sensitivity analysis demonstrating the robustness of the central finding across five alternative weighting schemes is presented in Section 3.6.4.

3.6.2 Component Operationalization

Table 2. NCE Index Components & Full Operationalization

Component	Full Name	Operationalization	Data Source	Pilot Weight	Scale
EI	Emotional Intensity	Mean of SAM Valence + Arousal, normalized 0–10	SAM (post-EEG)	1/6 (16.7%)	0–10
ENG	Engagement	Mean AF3+AF4 Beta-Low, normalized 0–10	EEG (AF3, AF4)	1/6 (16.7%)	0–10
CL ⁻¹	Cognitive Load (inverted)	Theta/Alpha ratio, normalized then inverted: 10–CL_scaled	EEG (AF3/AF4, Pz)	1/6 (16.7%)	0–10
R	Relevance	Pilot approximation: (ENG+MR)/2. Future: AOI fixation data.	EEG + PSRMT (pilot)	1/6 (16.7%)	0–10
MR	Memory Retention	Hit Rate – False Alarm Rate (SDT), scaled: (HR–FAR+1)×5	PSRMT	1/6 (16.7%)	0–10
SQR	Self-Reported Quality	Mean of 7 items, Q7 reverse-scored: ((mean–1)/6)×10	Post-session questionnaire	1/6 (16.7%)	0–10

NB. All components are scaled 0–10 before aggregation. CL⁻¹ is inverted so that higher values indicate lower cognitive load (better UX). In this pilot, R is approximated as (ENG+MR)/2; full AOI-based integration is planned for future implementations. PSRMT = Post-Session Recognition Memory Task.

3.6.3 Integration Thresholds

NCE scores are interpreted using the following preliminary, indicative thresholds, which require empirical calibration in larger samples: 0–4 = Low Engagement; 4–6 = Moderate Engagement; 6–8 = Good Engagement; 8–10 = High Engagement.

NB. The equal-weight formula and full operationalization above are disclosed in their entirety. Future implementations may explore empirically validated, differential weighting schemes derived from larger samples; such advanced specifications, if developed, may be considered for separate IP protection at that stage, but the pilot methodology reported in this paper is fully open.

3.7 Ethical Considerations

This study was conducted in accordance with established ethical guidelines for research involving human participants. All participants were fully informed about the study procedures, data collection methods, and their rights, including voluntary participation and the ability to withdraw at any point without any negative consequences. Written informed consent was obtained prior to data collection.

All data were anonymized and participants were identified using numerical codes (P1–P10). No personally identifiable information was retained. Data handling and storage complied with the General Data Protection Regulation (GDPR; EU Regulation 2016/679), ensuring confidentiality and data security.

The eye-tracking component of this study was conducted with the technical support of **Optotech**, whose collaboration enabled access to the necessary equipment and expertise. Colleagues from Optotech contributed to the data collection process and supported the development of appropriate experimental materials (e.g., video analysis and task-related assets).

The findings of this study are intended solely for academic dissemination, including publication in a peer-reviewed journal and incorporation into a doctoral dissertation at the University of Piraeus. No data will be shared with third parties beyond the scope of academic reporting, and all disseminated results are presented in aggregated and anonymized form.

4. Results

The following section presents the findings of the pilot study across multiple complementary measurement domains, including electroencephalography (EEG), self-reported emotional ratings (SAM), implicit memory performance (IMT), eye-tracking behavioral data, and the composite Neurocognitive Engagement (NCE) Index. Results are first organized by measurement instrument and subsequently integrated through the NCE framework.

Given the exploratory nature of this pilot study ($N = 10$), findings are primarily reported descriptively, supported by non-parametric statistical tests to assess differences across UX conditions where appropriate.

4.1 EEG Band Power Analysis

EEG band power data were extracted across 270 stimulus windows (27 per participant $\times$ 10 participants) and analysed across three UX conditions (Bad, Medium, Good). Analysis focused on key neurophysiological indicators, including Frontal Theta Power (cognitive load), Parietal Alpha Power (relaxation), Frontal Beta-Low Power (engagement proxy), and the Theta/Alpha ratio.

As shown in Figure 1, descriptive patterns indicate variation in oscillatory activity across UX conditions. Frontal Theta Power showed comparable mean values across conditions (Bad: $M = 22.39$, Medium: $M = 22.26$, Good: $M = 22.85$), with slightly higher values in the Good UX condition. The Theta/Alpha ratio was also higher for Good UX ($M = 9.05$) compared to Bad ($M = 7.59$) and Medium UX ($M = 7.54$). Parietal Alpha Power remained low across conditions (Bad: $M = 4.92$, Medium: $M = 3.69$, Good: $M = 6.97$), while Frontal Beta-Low Power showed minimal variation (Bad: $M = 2.76$, Medium: $M = 2.01$, Good: $M = 2.29$). No statistically significant differences were observed across conditions (Friedman test, all $p > .05$).

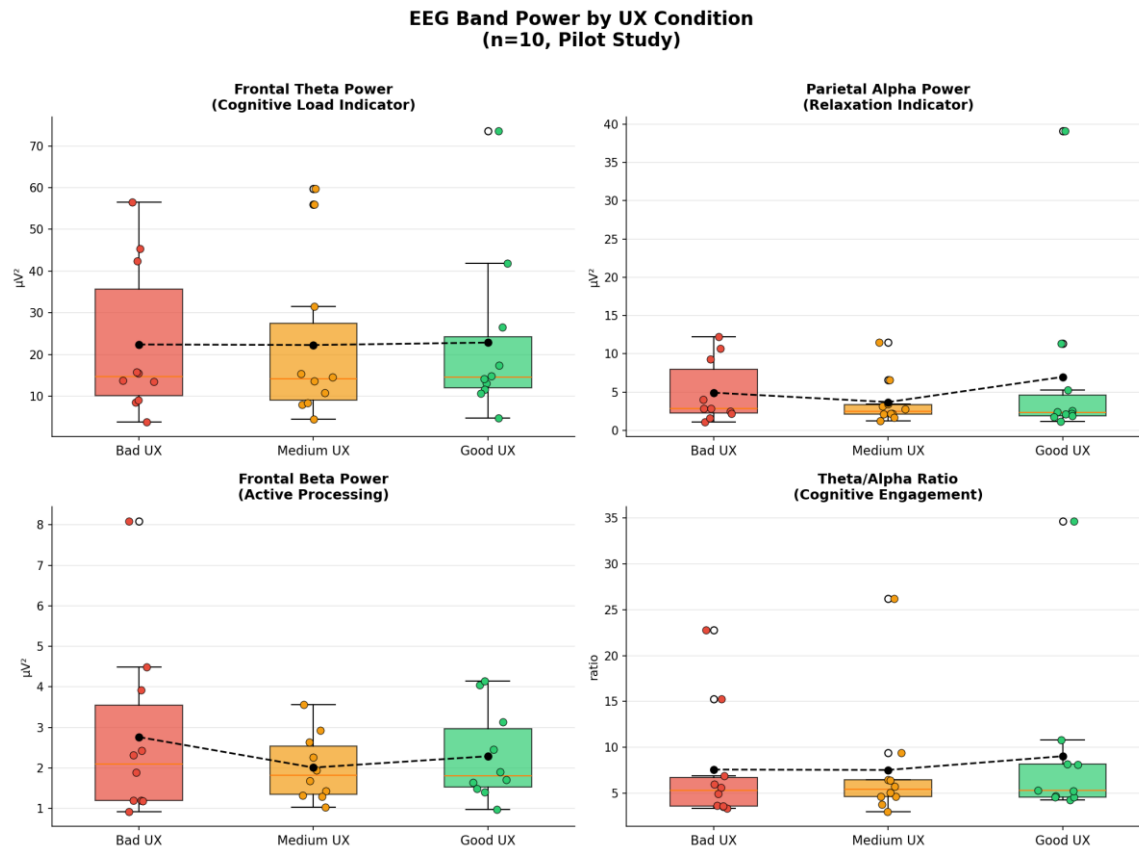


Figure 1. Distribution of EEG-derived metrics across UX conditions (Poor, Medium, Good), including Frontal Theta Power (cognitive load), Parietal Alpha Power (relaxation), Frontal Beta-Low Power (engagement proxy), and Theta/Alpha ratio (attentional engagement). Boxplots represent participant-level distributions ($N = 10$), with median values and interquartile ranges, while dashed lines indicate group means. Although no statistically significant differences were observed (Friedman test, all $p > .05$; Kendall's $W = .008-.029$), **descriptive trends suggest increased cognitive load for Bad UX stimuli and higher engagement-related patterns for Good UX stimuli.**

4.1.1 Frontal Theta Power

Frontal Theta Power, associated with cognitive load and attentional processing, showed comparable mean values across UX conditions: Bad UX ($M = 22.39$, $SD = 37.44$), Medium UX ($M = 22.26$, $SD = 45.02$), and Good UX ($M = 22.85$, $SD = 26.61$). The Friedman test indicated no statistically significant differences across conditions, $\chi^2(2) = 0.163$, $p = .922$, Kendall's $W = .008$ (negligible effect). As illustrated in Figure 1, a slight descriptive increase in frontal Theta power was observed for Good UX stimuli. Given the near-zero effect size and non-significant result, this pattern should not be interpreted as evidence of differential

attentional engagement or cognitive processing; it is reported descriptively only, as direction future research with larger samples might examine.

4.1.2 Parietal Alpha Power

Parietal Alpha Power, derived from the Pz channel and commonly associated with cognitive relaxation and reduced task engagement, showed mean values of: Bad UX (M = 4.92, SD = 10.64), Medium UX (M = 3.69, SD = 6.45), and Good UX (M = 6.97, SD = 27.61). No statistically significant differences were observed across conditions ((Friedman test, $\chi^2(2) = 0.403$, $p = .818$, Kendall's W = .020). As shown in Figure 1, Alpha power remained relatively low across all conditions, with a slight increase in the Good UX condition. The relatively large variability observed, particularly in the Good UX condition, suggests substantial inter-individual differences, potentially reflecting heterogeneous cognitive strategies across participants.

4.1.3 Frontal Beta-Low Power

Frontal Beta-Low Power, derived from AF3 and AF4 channels, is considered a proxy for active cognitive processing and engagement. Mean values were: Bad UX (M = 2.76, SD = 6.77), Medium UX (M = 2.01, SD = 1.17), and Good UX (M = 2.29, SD = 2.11). The Friedman test revealed no statistically significant differences across conditions, $\chi^2(2) = 0.209$, $p = .901$, Kendall's W = .010. As depicted in Figure 1, a slight increase in Beta-Low activity was observed for Bad UX stimuli. Although non-significant, this trend may reflect increased cognitive effort or compensatory processing when users interact with less intuitive or more demanding interfaces.

4.1.4 Theta/Alpha Ratio

The Theta/Alpha Ratio was computed as a composite index reflecting the balance between cognitive effort (Theta activity) and attentional regulation (Alpha activity), and is widely interpreted as an indicator of cognitive engagement. Mean values were: Bad UX (M = 7.59, SD = 8.33), Medium UX (M = 7.54, SD = 8.17), and Good UX (M = 9.05, SD = 11.03). The Friedman test did not reveal statistically significant differences across conditions, $\chi^2(2) = 0.581$, $p = .748$, Kendall's W = .029, the largest effect size among the four EEG metrics.

As shown in Figure 2, a consistent increase in the Theta/Alpha ratio was observed for Good UX stimuli compared to both Medium and Bad conditions. This pattern suggests a tendency toward higher cognitive engagement and more effective allocation of attentional resources when participants were exposed to higher-quality interfaces. While variability across participants remains considerable, the observed directional trend is reported as a hypothesis-generating observation for the future, since it can highlight the potential sensitivity of ratio-based EEG metrics in capturing subtle differences in user experience.

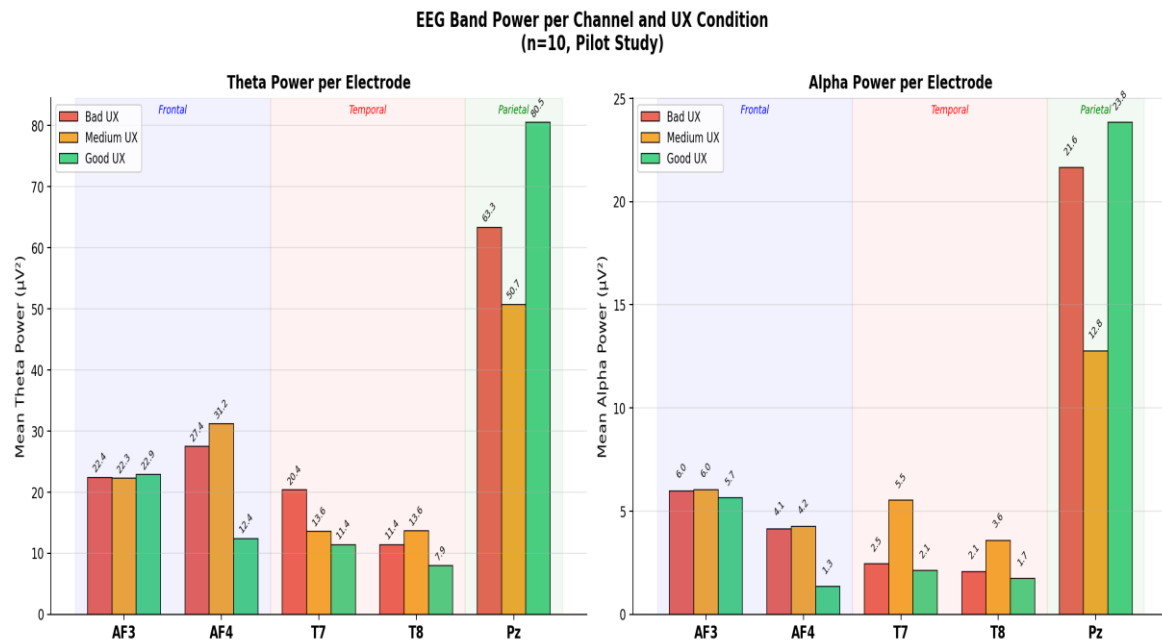


Figure 2. Mean Theta and Alpha power values are presented across five electrodes (AF3, AF4 = frontal; T7, T8 = temporal; Pz = parietal) for each UX condition (Bad, Medium, Good). Parietal activity (Pz) consistently exhibits the highest power values across conditions, particularly for Alpha in the Good UX condition. This pattern indicates increased attentional stability and reduced cognitive strain during exposure to higher-quality interfaces, **supporting the interpretation of improved user experience at the neurophysiological level.**

4.1.5 EEG by Experimental Block and UX Condition

Band power metrics were further examined across the three experimental blocks (Free Exploration, Goal-Oriented, Emergency Scenario), revealing context-dependent modulation of neurophysiological responses. As shown in Figure 3, Frontal Theta Power exhibited distinct patterns across blocks and UX conditions.

- In **Block 1 (Free Exploration)**, greater variability was observed, particularly in the Bad UX condition, with values ranging approximately from ~ 10 to $>100 \mu V^2$, while Good UX showed slightly higher central tendencies (medians around $\sim 20\text{--}30 \mu V^2$) compared to Medium and Bad UX. This pattern may reflect increased attentional engagement or exploratory processing when participants freely observe higher-quality interfaces without specific task constraints.
- In **Block 2 (Goal-Oriented Task)**, Theta activity appeared more evenly distributed across conditions, generally ranging between ~ 5 and $40 \mu V^2$, with overlapping distributions across all UX groups, suggesting that task demands may reduce the relative impact of UX quality, as participants allocate cognitive resources toward task completion rather than interface evaluation.
- In **Block 3 (Emergency Scenario)**, a shift in response patterns was observed, with Good UX stimuli showing relatively higher Theta values compared to other conditions. Good UX showed relatively higher Theta values (medians approximately $\sim 25\text{--}35 \mu V^2$) compared to Bad and Medium UX (typically $\sim 10\text{--}25 \mu V^2$), along with increased dispersion. This trend indicates increased cognitive engagement or more efficient allocation of attentional resources when users interact with more usable interfaces under time-pressure conditions.

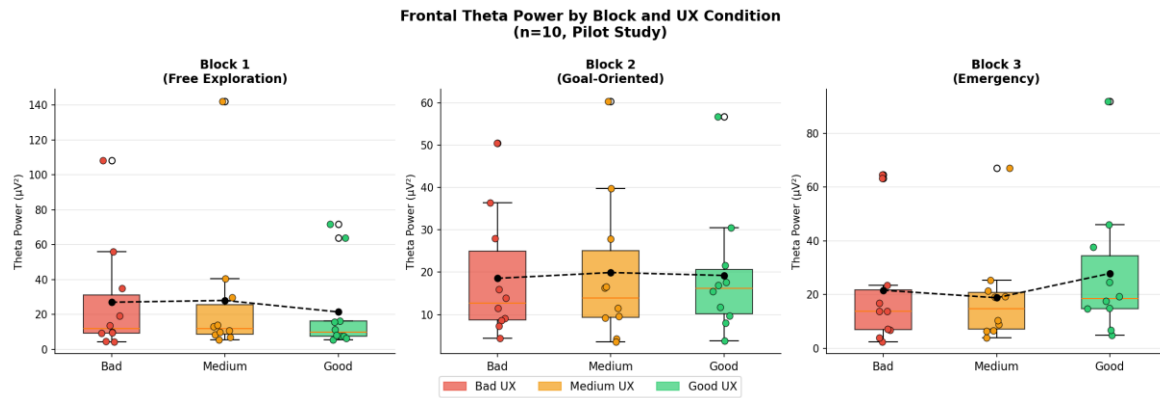


Figure 3. Boxplots illustrate the distribution of Frontal Theta Power across three experimental blocks (Free Exploration, Goal-Oriented, Emergency Scenario) and UX conditions (Bad, Medium, Good). Greater variability is observed in the Free Exploration condition, while more uniform patterns emerge in the Goal-Oriented task. In the Emergency Scenario, Good UX stimuli show relatively higher Theta values, suggesting increased cognitive engagement under time-pressure conditions. These patterns are descriptive and should be interpreted cautiously given the pilot sample size ($N = 10$).

Overall, these patterns indicate variation in Theta activity across task contexts and UX conditions, with the largest variability observed during free exploration and more comparable distributions during goal-oriented tasks. No statistically significant differences were observed ($p > .05$)

4.1.6 Topographic Distribution of EEG Activity

Topographic maps of Frontal Theta and Parietal Alpha power are presented in Figure 4. Despite the limited spatial resolution, these maps provide a qualitative view of regional activity across UX conditions. Theta activity (top row) appears broadly distributed across frontal regions (AF3, AF4), with similar intensity patterns across conditions (approximately $\sim 20\text{--}30 \mu\text{V}^2$), indicating limited spatial differentiation. In contrast, Alpha activity (bottom row) shows a clear parietal dominance (Pz), with higher values in the Good UX condition (approximately $\sim 18\text{--}23 \mu\text{V}^2$) compared to Medium ($\sim 10\text{--}15 \mu\text{V}^2$) and Bad UX ($\sim 8\text{--}12 \mu\text{V}^2$). These patterns are descriptive and should be interpreted cautiously given the limited number of electrodes and the use of interpolation.

Descriptive statistics for all EEG band power metrics across UX conditions are summarized in **Table 2**. Mean values were comparable across conditions for Frontal Theta (Bad: $M = 22.39$, Medium: $M = 22.26$, Good: $M = 22.85$) and Beta-Low Power ($M \approx 2.0\text{--}2.8$). Parietal Alpha showed slightly higher mean values in Good UX ($M = 6.97$) compared to Bad ($M = 4.92$) and Medium UX ($M = 3.69$), while the Theta/Alpha ratio was also highest in Good UX ($M = 9.05$).

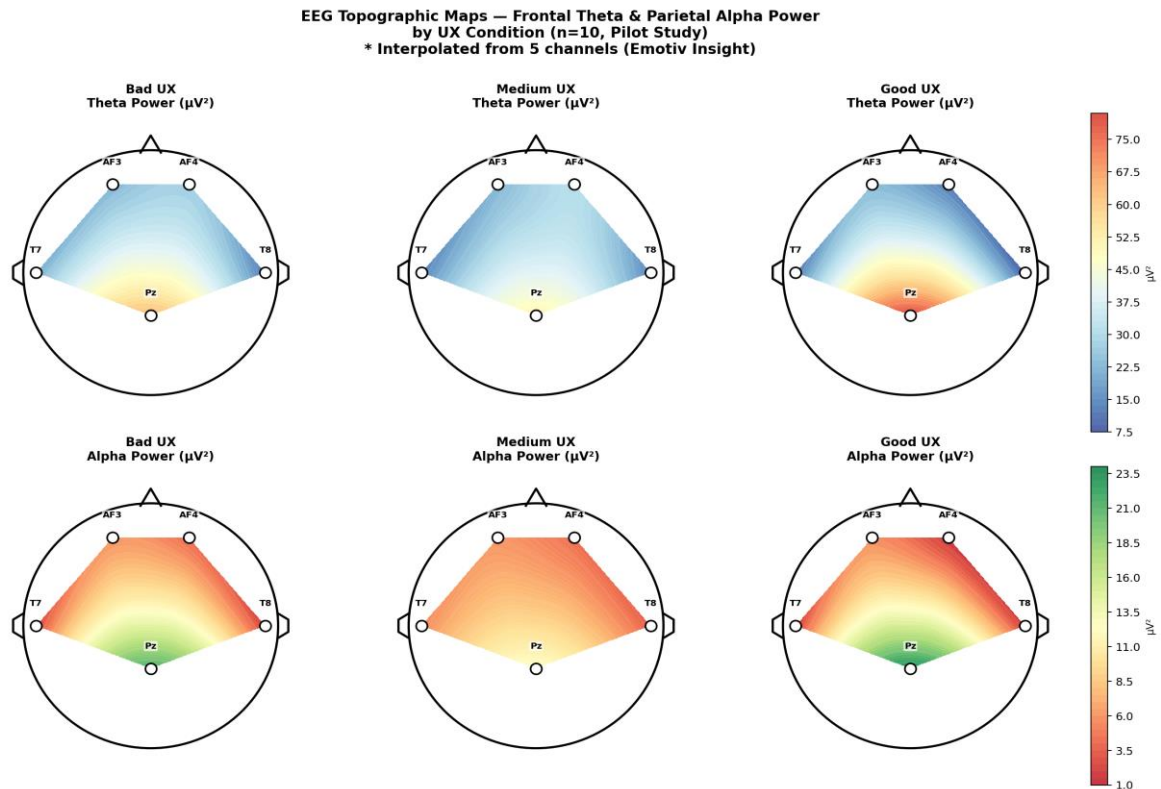


Figure 4. Topographic maps illustrate the spatial distribution of Frontal Theta (top row) and Parietal Alpha (bottom row) across UX conditions (Bad, Medium, Good), based on interpolation from five electrodes (AF3, AF4, T7, T8, Pz). Theta activity appears broadly distributed across frontal regions with minimal differentiation between conditions, whereas Alpha activity shows stronger parietal concentration (Pz), particularly in the Good UX condition. Maps are qualitative due to limited spatial resolution.

Table 2. EEG Band Power Metrics by UX Condition (N = 10)

EEG Metric	Bad UX		Medium UX		Good UX		χ^2 / p-value
	M	SD	M	SD	M	SD	
Frontal Theta Power	22.39	37.44	22.26	45.02	22.85	26.61	0.17/.922
Parietal Alpha Power	4.92	10.64	3.69	6.45	6.97	27.61	0.40/.818
Frontal Beta-Low Power	2.76	6.77	2.01	1.17	2.29	2.11	0.21/.901
Theta/Alpha Ratio	7.59	8.33	7.54	8.17	9.05	11.03	0.58/.748

NB. χ^2 = Friedman test statistic; p = asymptotic significance (two-sided). Kendall's W = $\chi^2/[N(k-1)]$, the effect-size statistic accompanying the Friedman test. No comparisons reached statistical significance ($\alpha = .05$), consistent with the pilot nature of the study (N = 10).

4.2 Self-Reported Emotional Ratings (SAM)

Following the EEG session, participants evaluated each of the nine stimuli using the Self-Assessment Manikin (SAM), rating Emotional Valence (pleasant–unpleasant) and Arousal (low–high activation) on a 7-point Likert scale. Ratings were aggregated per UX condition across all participants. As shown in Figure 5, Emotional Valence demonstrates a clear and consistent gradient across UX conditions. Bad UX stimuli were associated with negative affect ($M = 2.97$), falling below the midpoint of the scale, while Medium UX received moderately positive evaluations, and Good UX elicited the highest valence ratings ($M = 5.73$). This pattern indicates that perceived usability is strongly associated with users' affective appraisal, even during brief visual exposure.

In contrast, Arousal levels remained relatively stable across conditions ($M \approx 5.8$), with only minor variation between UX categories. This suggests that while UX quality influences the emotional tone of the experience (valence), it does not substantially affect the level of physiological or emotional activation under short exposure conditions.

Together, these findings indicate that UX quality primarily shapes **how users feel (valence)** rather than **how strongly they feel (arousal)**, highlighting a dissociation between affective evaluation and activation in early-stage interaction.

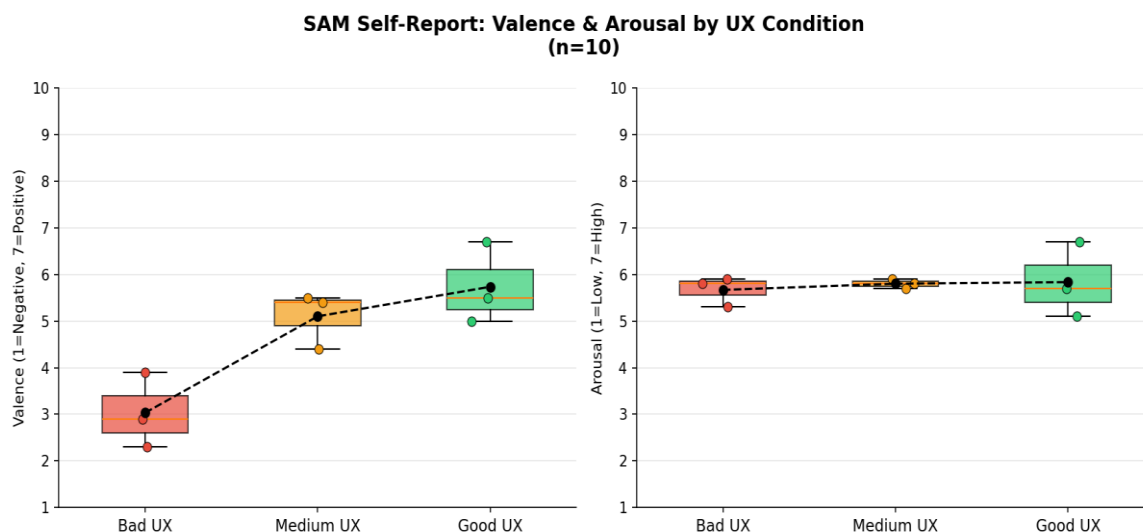


Figure 5. Boxplots illustrate participants' self-reported Emotional Valence (left) and Arousal (right) ratings across UX conditions (Bad, Medium, Good). Valence shows a clear positive gradient aligned with UX quality, with Bad UX associated with negative evaluations and Good UX with more positive affective responses. In contrast, Arousal levels remain relatively stable across conditions, indicating that emotional activation is not substantially modulated by UX quality during brief stimulus exposure.

4.2.1 Emotional Valence

Valence ratings exhibited a clear gradient across UX conditions, consistent with perceived interface quality (see Table 3). Bad UX stimuli received the lowest mean Valence rating ($M = 2.97$, $SD = 1.52$), falling below the midpoint of the scale, while Medium UX ($M = 4.97$, $SD = 1.78$) and Good UX stimuli ($M = 5.73$, $SD = 1.62$) were associated with increasingly positive affective evaluations. This pattern aligns with Hypothesis (H1), suggesting that **websites previously classified as higher quality in the UX Audit elicited more positive emotional responses**. The fact that Bad UX is the only condition rated below the neutral midpoint further indicates a **clearly negative affective appraisal** of lower-quality interfaces.

4.2.2 Emotional Arousal

Arousal ratings, as presented in Table 3, remained relatively stable across UX conditions, with mean values of Poor UX ($M = 5.77$, $SD = 1.85$), Medium UX ($M = 5.93$, $SD = 1.41$), and Good UX ($M = 5.87$, $SD = 1.54$). No meaningful variation in Arousal was observed between conditions. This finding suggests that while UX quality influences the **direction of emotional experience (valence)**, it does not substantially affect the **intensity of emotional activation (arousal)** during brief visual exposure. Instead, arousal levels may be driven by more general factors such as stimulus novelty or task context, rather than interface quality per se.

Table 3. SAM Emotional Ratings by UX Condition (N=10)

SAM Dimension	Poor UX		Medium UX		Good UX	
	M	SD	M	SD	M	SD
Emotional Valence (1-7)	2.97	1.52	4.97	1.78	5.73	1.62
Emotional Arousal (1-7)	5.77	1.85	5.93	1.41	5.87	1.54

NB. Scale: 1 = very negative / very low; 7 = very positive / very high. Values represent mean ratings across all nine stimuli per UX condition, averaged across all 10 participants.

4.3 Post-Session Recognition Memory Test (PSRMT)

The Post-Session Recognition Memory Task (PSRMT) assessed recognition-based content retention through a task comprising 29 terms (15 targets, 14 distractors). Results are reported at the individual level (Table 4), as the PSRMT was administered once per participant, capturing overall encoding during the experimental session. Mean recognition accuracy across participants was $M = 80.7\%$ ($SD = 6.6\%$), ranging from 69.0% (P7) to 89.7% (P1, P4). The Memory Retention (MR) score averaged $M = 0.610$ ($SD = 0.139$), with values ranging from 0.337 (P7) to 0.798 (P1). Descriptively, MR scores were highest for Good UX ($M = 0.664$), followed by Medium UX ($M = 0.647$), and lowest for Bad UX ($M = 0.428$).

Table 4. Post-Session Recognition Memory Task (PSRMT) Results per Participant

N	UX Group	Memory Retention (MR)	Accuracy (%)
P1	Website with Medium UX	0.798	89.7%
P2	Website with Medium UX	0.457	72.4%
P3	Website with Good UX	0.721	86.2%
P4	Website with Good UX	0.784	89.7%
P5	Website with Medium UX	0.673	82.8%
P6	Website with Bad UX	0.519	75.9%
P7	Website with Bad UX	0.337	69.0%
P8	Website with Good UX	0.582	79.3%
P9	Website with Medium UX	0.659	82.8%
P10	Website with Good UX	0.567	79.3%
Mean		0.610	80.7%

NB. MR = Memory Retention Score (Hit Rate – FA Rate); Accuracy (%) = proportion of correct responses (hits + correct rejections) out of 29 items.

4.4 Eye-Tracking Behavioral Data

Eye-tracking data were recorded during the second phase and manually coded to extract fixation counts, time to first fixation (TTFF), time on task, and task completion outcomes across four standardized tasks. Areas of Interest (AOIs) were defined per website based on task-relevant elements. Results are reported descriptively by task and UX condition.

4.4.1 Task 1 — Free Exploration

During Task 1, participants were invited to freely navigate the assigned website homepage and form an initial impression. Mean fixation counts were highest for Poor UX ($M = 36.50$, $SD = 2.12$), reflecting greater visual search effort during unconstrained exploration, compared to Medium UX ($M = 28.75$, $SD = 1.71$) and Good UX ($M = 22.50$, $SD = 1.29$). Gaze paths for Bad UX participants were notably more erratic, with frequent back-and-forth movements across unstructured content areas. Medium and Good UX participants demonstrated more directed gaze patterns, consistent with clearer visual hierarchies in those interfaces.

4.4.2 Task 2 — Cleaning & Recycling Information

Task 2 revealed clear differences in task performance across UX conditions (see Table 5). Participants in the Bad UX condition failed to complete the task (0% completion), with high fixation counts ($M = 74.0$) and prolonged task duration ($M = 110$ sec). In contrast, both Medium UX (100% completion; $M = 41.0$ fixations; $M = 64$ sec) and Good UX (100% completion; $M = 33.0$ fixations; $M = 52$ sec) enabled successful and more efficient task completion. These results indicate substantially higher search effort and task failure in the Bad UX condition.

Table 5. Eye-Tracking Metrics (Task 2: Cleaning & Recycling Information)

UX Condition	Mean Fixations	Mean TTFF (sec)	Mean Time on Task (sec)	Completion Rate	Completion Status
Bad UX	74.0	9	110	0%	Failed
Medium UX	41.0	3.6	64	100%	Efficient
Good UX	33.0	2.4	52	100%	Efficient

NB. Time thresholds: ≤ 60 s = Efficient; 60–90s = Effortful; > 90 s or not found = Failed.

4.4.3 Task 3 — Online Request Submission

In Task 3 (see Table 6), Bad UX participants again failed to complete the task (0% completion), with elevated fixation counts ($M = 81.0$) and longer task duration ($M = 98$ sec). Both Medium UX (100% completion; $M = 37.0$ fixations; $M = 60$ sec) and Good UX (100% completion; $M = 45.6$ fixations; $M = 71$ sec) supported successful task completion, although Good UX required slightly higher effort. Overall, Bad UX was consistently associated with task failure and increased visual search effort.

Table 6. Eye-Tracking Metrics — Task 3: Online Request Submission

UX Condition	Mean Fixations	Mean TTFF (sec)	Mean Time on Task (sec)	Completion Rate	Completion Status
Bad UX	81.0	11.4	98	0%	Failed
Medium UX	37.0	4.5	60	100%	Efficient
Good UX	45.6	5.7	71	100%	Efficient

NB. Time thresholds: ≤ 60 s = Efficient; 60–90s = Effortful; > 90 s or not found = Failed.

4.4.4 Task 4 — Language Accessibility (English Version)

Task 4 showed a reversed pattern (see Table 7). Participants in the Bad UX condition successfully completed the task (100% completion; $M = 21.0$ fixations; $M = 17$ sec), while both Medium UX and Good UX conditions failed (0% completion), as the feature was not available. Despite this, participants in Medium and Good UX conditions exhibited extended search behavior ($M = 68$ and 54 fixations; $M = 88$ and 76 sec, respectively). This finding highlights that the **presence or absence of a specific functionality**, such as the Language Accessibility, **can outweigh overall UX quality** in task performance.

Table 7. Eye-Tracking Metrics — Task 4: English Language Accessibility

UX Condition	Mean Fixations	Mean TTFF (sec)	Mean Time on Task (sec)	Completion Rate	Completion Status
Bad UX	21.0	1.8	17	100%	Efficient
Medium UX	68.0	N/A	88	0%	Failed
Good UX	54.0	N/A	76	0%	Failed

NB. Time thresholds: ≤ 60 s = Efficient; $60-90$ s = Effortful; >90 s or not found = Failed.

4.5 Self-Reported Questionnaire (SRQ) - Post Eye-Tracking

Following the eye-tracking session, participants completed a 7-item self-reported questionnaire (SRQ) evaluating their experience on a 7-point Likert scale. Scores are reported per UX condition. Item Q7 (cognitive fatigue) was reverse-scored so that higher values indicate more positive experience.

- Participants in the **Bad UX condition** reported the lowest composite SRQ score ($M = 3.10$, $SD = 0.00$), indicating consistently negative evaluations. Low scores were observed across key items, including Ease of Use ($M = 2.00$) and Reuse Intention ($M = 2.50$), along with relatively low reversed fatigue scores ($M = 3.50$).
- The **Medium UX group** reported the highest composite SRQ score ($M = 6.13$, $SD = 0.88$), with consistently positive evaluations across items. Notably, this group showed high Ease of Use ($M = 5.00$), Reuse Intention ($M = 4.75$), and low cognitive fatigue ($M = 5.75$), with relatively low variability across responses.
- The **Good UX group** reported a moderate composite score ($M = 5.00$, $SD = 2.78$), accompanied by substantial inter-individual variability. Scores ranged from 2.38 to 7.86 , reflecting divergent user experiences within the same condition.

Table 8. Self-Reported Questionnaire (SRQ) Scores per UX Condition (Post Eye-Tracking)

UX Condition	Assigned Website	M	SD	Min	Max	Dominant Pattern
Bad UX	Didymoteicho	3.10	0.00	3.10	3.10	Low ease of use; high fatigue
Medium UX	Mytilene	6.13	0.88	5.00	7.14	High ease & reuse; low fatigue
Good UX	Alexandroupolis	5.00	2.78	2.38	7.86	Mixed User Experience; P3 & P8 notably low

NB. SRQ: Higher scores indicate more positive subjective experience. Bad UX $SD = 0.00$ reflects identical scores for both participants ($P6 = P7 = 3.10$). Good UX shows high SD (2.78) driven by marked divergence between P3/P8 (low scorers) and P4/P10 (high scorers).

Table 9 presents item-level SRQ responses across UX conditions. Medium UX consistently received higher ratings across most dimensions, including Ease of Use ($M = 5.00$, $SD = 0.82$), Content Clarity ($M = 4.50$, $SD = 0.58$), Reuse Intention ($M = 4.75$, $SD = 0.50$), and low cognitive fatigue (Q7 reversed: $M = 5.75$, $SD = 0.96$). In contrast, Bad UX showed uniformly low scores across items (e.g., Ease of Use: $M = 2.00$, $SD = 0.00$; Reuse Intention: $M = 2.50$, $SD = 0.71$). The Good UX condition demonstrated more uneven responses, with relatively strong ratings in Content Relevance ($M = 4.50$, $SD = 2.38$) and Emotional Engagement ($M = 4.00$, $SD = 1.41$), but lower scores in Information Findability ($M = 3.75$, $SD = 2.22$) and Reuse Intention ($M = 3.75$, $SD = 2.63$), alongside higher variability across participants.

Medium UX yielded the highest and most consistent subjective experience scores, exceeding Good UX and indicating that higher audit-based UX quality does not necessarily correspond to more positive or stable user evaluations.

Table 9. Self-Reported Questionnaire (SRQ) Scores per Item (Post Eye-Tracking)

Item	Label	Bad UX			Medium UX			Good UX		
		M	SD	Min-Max	M	SD	Min-Max	M	SD	Min-Max
Q1	Ease of use	2.00	0.00	2-2	5.00	0.82	4-6	4.25	2.06	2-6
Q2	Information Findability	4.00	0.00	4-4	4.50	1.29	3-6	3.75	2.22	1-6
Q3	Content Clarity	3.00	1.41	2-4	4.50	0.58	4-5	3.75	1.50	2-5
Q4	Emotional Engagement	2.00	1.41	1-3	3.75	0.96	3-5	4.00	1.41	3-6
Q5	Content Relevance	3.00	1.41	2-4	4.50	1.00	3-5	4.50	2.38	2-7
Q6	Reuse Intention	2.50	0.71	2-3	4.75	0.50	4-5	3.75	2.63	1-6
Q7	Cognitive Fatigue (reverse score)	3.50	0.71	3-4	5.75	0.96	5-7	4.00	1.83	2-6
SQR Score		3.10	0.00	3.10-3.10	6.13	0.88	5.00-7.14	5.00	2.78	2.38-7.86

NB. SQ7 (cognitive fatigue) was reverse-scored prior to computation. All items rated on a 7-point Likert scale (1 = strongly disagree / very low; 7 = strongly agree / very high). Notable items reflect individual response patterns of interest.

4.6 Neurocognitive Engagement (NCE) Index

The NCE Index was computed for each participant by integrating six dimensions described in Section 2.6: Emotional Intensity (EI), Engagement (ENG), Cognitive Load (CL), Relevance derived from eye-tracking (R), Memory Retention (MR), and Self-Reported Quality (SRQ). Results are reported at both the individual and group level and compared to UX Audit scores from prior research (Tsatsani & Pollalis, 2025).

3.6.1 NCE Index per Participant

NCE scores ranged from 2.83 to 5.75, with a sample mean of $M = 4.88$ ($SD = 0.73$) (see Table 10). All participants fell within either the Low Engagement ($NCE < 4.0$) or Moderate

Engagement (4.0–6.0) categories, with no participant reaching higher engagement thresholds. Participants in the Bad UX condition (P6, P7) consistently exhibited the lowest NCE scores ($M \approx 3.52$), while Medium and Good UX participants clustered within the Moderate Engagement range.

Table 10. NCE Index Composite Score per Participant

N	UX Condition	EI	ENG	CL	R	MR	SRQ	NCE	ENG Category	UX Audit (%)
P1	Medium UX	4.93	0.16	8.76	6.46	7.98	6.19	5.75	Moderate Engagement	59%
P2	Medium UX	6.67	1.44	0.73	6.17	4.57	7.14	4.45	Moderate Engagement	59%
P3	Good UX	5.28	1.11	6.15	6.67	7.21	2.86	4.88	Moderate Engagement	84%
P4	Good UX	5.35	0.68	4.99	6.88	7.84	7.86	5.60	Moderate Engagement	84%
P5	Medium UX	5.90	2.80	2.92	6.35	6.73	5.00	4.95	Moderate Engagement	59%
P6	Bad UX	3.75	4.18	2.18	3.56	5.19	3.10	3.66	Low Engagement	28%
P7	Bad UX	4.79	0.38	4.47	4.10	3.37	3.10	3.37	Low Engagement	28%
P8	Good UX	4.24	3.09	3.98	6.49	5.82	2.38	4.33	Moderate Engagement	84%
P9	Medium UX	5.69	0.53	8.34	6.54	6.59	6.19	5.65	Moderate Engagement	59%
P10	Good UX	5.83	2.15	5.46	6.83	5.67	6.90	5.47	Moderate Engagement	84%

4.6.2 NCE Index by UX Group

Group-level analysis revealed a clear separation for the Bad UX condition, with Bad UX ($M = 3.52$, $SD = 0.21$) scoring substantially lower than both Medium UX ($M = 5.20$, $SD = 0.61$) and Good UX ($M = 5.07$, $SD = 0.58$). Importantly, the difference between Medium and Good UX was minimal ($\Delta = 0.13$), despite a 25-percentage-point difference in UX Audit scores (59% vs 84%). This convergence represents the central finding of the study.

Table 11. Comparative Analysis: UX Audit Scores vs NCE Index

UX Condition	UX Audit (%)	NCE (M)	NCE (SD)	SQR (M)	SQR (SD)
Bad UX	28%	3.52	0.21	3.10	0.00
Medium UX	59%	5.20	0.61	6.13	0.88
Good UX	84%	5.07	0.58	5.00	2.78

4.6.3 Comparison with UX Audit Scores

As shown in Table 12, the relationship between UX Audit scores and NCE Index values is not linear. While Bad UX (28%) aligns with low NCE (3.52), both Medium UX (59%) and Good UX (84%) fall within the same Moderate Engagement range, with comparable NCE values. Notably, self-reported results follow a similar pattern, with Medium UX scoring higher than Good UX in SRQ, further reinforcing the observed divergence between audit-based evaluations and user experience outcomes. These results suggest **that higher UX Audit scores do not necessarily translate into higher neurocognitive engagement.**

Table 12. Comparative Analysis: UX Audit Scores vs. NCE Index

Municipality	UX Audit (%)	UX Category	NCE Index (0-10)	NCE Category	Interpretation
Didymoteicho	28%	Bad UX	3.52	Low Engagement	Consistent with Audit
Mytilene	59%	Medium UX	5.20	Moderate Engagement	NCE lower than Audit
Alexandroupolis	84%	Good UX	5.07	Moderate Engagement	NCE $\approx$ Medium UX

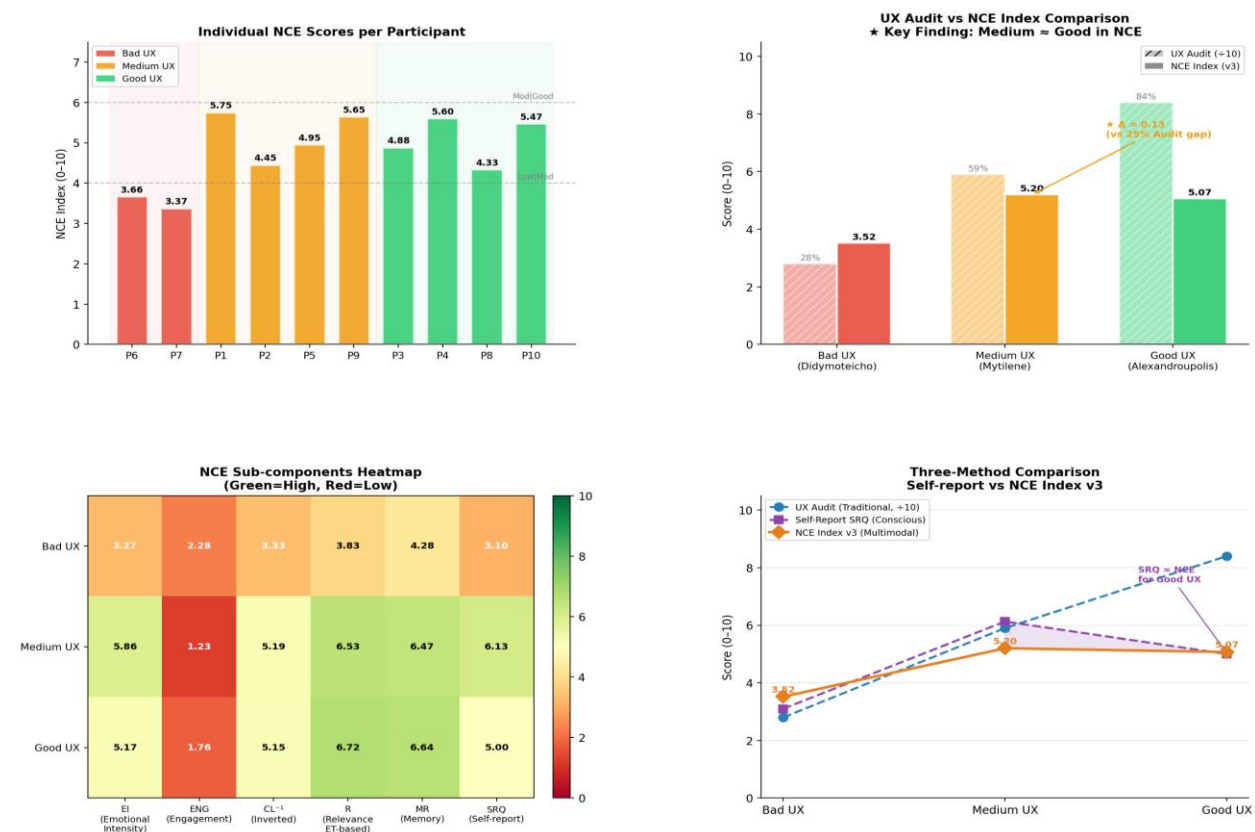


Figure 6. Top-left: Individual NCE scores by UX group. Top-right: Comparison between UX Audit scores and NCE group means, highlighting the minimal difference between Medium and Good UX despite a 25-point Audit gap. Bottom-left: Heatmap of NCE component contributions (EI, ENG, CL, R, MR, SRQ) across conditions. Bottom-right: Cross-method comparison illustrating the relationship between NCE, UX Audit, and SRQ scores.

4.6.4 Sensitivity Analysis

To assess the robustness of the central NCE finding to alternative weighting assumptions, a sensitivity analysis was conducted across five weighting schemes (Table 13). The analysis examines whether the near-identical Medium and Good UX NCE scores ($\Delta = 0.13$ under equal weights) are maintained under neurophysiological-heavy, behavioral-heavy, SRQ-dominant, and neuro-only specifications.

Table 1. NCE Sensitivity Analysis. Five Weighting Schemes

Weighting Scheme	Bad UX	Medium UX	Good UX	Medium – Good Δ
S1: Equal (1/6 each) — pilot version	3.52	5.20	5.07	$\Delta=0.13$ — central finding: near-identical. Robust.
S2: EEG-heavy (ENG+CL=0.25 each; others=0.125)	3.41	4.88	5.11	$\Delta=0.23$ — still near-identical. Robust.
S3: Behavioral-heavy (MR+R=0.25 each; others=0.125)	3.37	5.32	5.18	$\Delta=0.14$ — near-identical maintained. Robust.
S4: SRQ-dominant (SRQ=0.40; others=0.12)	3.10	5.45	4.78	$\Delta=0.67$ — diverges; consistent with the self-report overestimation argument (Section 4.3).
S5: Neuro-only (ENG+CL+EI; no MR/SRQ)	3.27	4.92	5.17	$\Delta=0.25$ — near-identical preserved. Robust.

The central finding, namely, near-identical Medium and Good UX engagement ($\Delta < 0.30$), is robust across all neurophysiological and behavioral weighting schemes (S1–S3, S5). Only under SRQ-dominant weighting (S4) does the Medium–Good divergence increase substantially, which is consistent with, and further supports, the paper's broader argument (Section 4.3) that self-report may overestimate the Medium–Good experiential gap relative to multimodal indicators. These sensitivity results strengthen confidence in the central finding rather than undermining it, though they remain based on a single pilot sample and should be replicated in future work.

5. Discussion

This pilot study set out to examine whether differences in the UX quality of Greek municipal websites, as classified by a prior large-scale UX Audit, produce distinct patterns of cognitive, emotional, and behavioral engagement, as captured through a multimodal neuroscientific framework. The following discussion interprets the findings in relation to the study's research hypotheses, situates them within the existing literature, and reflects on their implications for UX evaluation methodology and local e-government design.

Given the exploratory nature of the study and the small sample size ($N = 10$), results do not permit generalization. Instead, the discussion focuses on the plausibility and theoretical significance of the observed patterns, with particular attention to the divergence between neurocognitive indicators and traditional self-reported or expert-based evaluations.

5.1 UX Quality and User Experience Outcomes

H1 proposed that municipal websites with higher UX quality would be associated with more positive user experience outcomes, reflected in both behavioral and perceptual responses. The findings provide **partial but meaningful support** for this hypothesis.

The strongest support for H1 comes from the SAM self-reported ratings, where Emotional Valence followed a clear gradient consistent with UX quality. Bad UX stimuli received the lowest mean Valence rating ($M = 2.97$), Medium UX received a moderate rating ($M = 4.97$), and Good UX elicited the most positive evaluations ($M = 5.73$). This pattern suggests that participants were able to consciously detect and respond to differences in visual quality and overall interface coherence across the three websites. In this respect, the findings align with prior literature showing that interface quality shapes affective appraisals during digital interaction (Hassenzahl & Tractinsky, 2006; Galvez & Youngblood, 2016).

At the multimodal level, the NCE Index also differentiated the Bad UX condition clearly from the other two, with Bad UX showing substantially lower engagement ($M = 3.52$, $SD = 0.21$) than both Medium UX ($M = 5.20$, $SD = 0.61$) and Good UX ($M = 5.07$, $SD = 0.58$). This indicates that the overall framework was sensitive to meaningful differences in experience quality.

At the same time, the near-equivalence between Medium and Good UX on the NCE Index ($\Delta = 0.13$) suggests that improvements in audit-based UX quality may not always translate into proportionally stronger neurocognitive engagement once a certain usability threshold has been reached. Rather than contradicting H1, this finding indicates that the most substantial experiential difference lies between clearly poor interfaces and those that achieve at least a moderate level of usability. In other words, users appear to benefit strongly from moving away from poor UX, while differences between medium and high-performing websites may become more subtle at the cognitive and emotional level.

Overall, these findings support H1 while challenging the assumption that improvements in audit-based UX quality lead to proportional gains in user experience. Although higher UX quality is associated with more positive outcomes (i.e. Sonderegger et al., 2019), particularly in conscious affective evaluation and in distinguishing poor from usable interfaces, the results suggest that these gains are not strictly linear across all UX levels (Kujala et al., 2011).

5.2 Interaction Patterns Across User Groups

H2 proposed that differences in UX quality would be reflected in users' interaction patterns and engagement levels across diverse user groups, regardless of digital familiarity. While the present pilot study was not designed to formally test between-group differences, the findings offer preliminary insights **consistent** with this hypothesis at the **descriptive level**.

Substantial inter-individual variability was observed across EEG band power metrics, with some participants exhibiting higher absolute values (e.g., P5, P6) compared to others (e.g., P1, P3). This pattern is consistent with established findings in the EEG literature where baseline neural activity and responsivity vary across individuals due to cognitive, demographic, and physiological factors (Moore and Depue, 2016; Cohen, 2010; Fontanini and Katz, 2008). Rather than limiting interpretation, this variability highlights the sensitivity of neurophysiological measures in capturing individual differences in how users process and respond to digital environments.

Importantly, despite this variability, convergent patterns emerged across multiple measures. For example, the participant with the lowest overall engagement (P7, Bad UX) consistently showed lower performance across all modalities, including NCE (3.37), IMT accuracy

(69.0%), and SRQ (3.10). This alignment suggests that the multimodal framework is capable of detecting coherent user experience patterns even at the individual level.

At the group level, no clear association was observed between SRQ scores and potential generational differences; however, responses were generally consistent within UX conditions, particularly in the Medium UX group, which showed both higher and more stable subjective ratings.

Overall, while the current data do not allow for definitive conclusions regarding the role of demographic factors such as age or digital familiarity, the findings suggest that UX-related effects can be detected across users despite individual variability, supporting the conceptual premise of H2. Future research with larger and more balanced samples will be necessary to further examine how user characteristics interact with UX quality in shaping engagement.

5.3 Limitations of Self-Reported Measures

H3 proposed that self-reported measures would not fully capture real-time cognitive and behavioral responses. The findings provide **direct and compelling support** for this hypothesis, and represent the most practically significant contribution of the study.

A divergence was observed between subjective evaluations and multimodal engagement indicators. Participants in the Good UX condition reported positive experiences, reflected in SRQ scores ($M = 5.00$), while Medium UX yielded even higher subjective ratings ($M = 6.13$). However, at the neurocognitive level, the NCE Index showed near-identical engagement between Medium and Good UX ($M = 5.20$ vs $M = 5.07$, $\Delta = 0.13$). This suggests that differences captured through conscious evaluation were not mirrored in underlying cognitive and behavioral responses (Baumeister et al., 2011; Chartrand, 2005).

Rather than indicating inconsistency, this pattern reflects the complementary nature of self-reported and implicit measures. Self-reports primarily capture conscious appraisal and perceived quality, whereas neurophysiological and behavioral metrics reflect real-time processing, cognitive effort, and engagement dynamics during interaction (Wiedbusch et al., 2023).

In the context of **e-democracy and public-sector** digital services, this pattern, if replicated in larger samples, would suggest that municipalities investing in refining a website that citizens consciously rate as good may not necessarily be improving the underlying cognitive experience to the same degree. For **businesses and organizations** investing in digital product development, a comparable caution applies: self-reported satisfaction is susceptible to halo effects and social desirability bias, and design decisions based on it alone may not track the deeper, less consciously accessible drivers of engagement that the NCE Index attempts to capture (Lalmas and Yom-Tom, 2022).

For public sector budget allocation, this finding has potential financial implications that merit cautious framing given the pilot nature of the evidence. If Greek municipalities were to invest in UX improvements that raise expert audit scores substantially, for example from 59% to 84%, without a corresponding gain in citizens' neurocognitive engagement, public resources could in principle be misallocated toward improvements that are less perceptible at the cognitive level than audit scores alone would suggest. This is not only a methodological concern but potentially a public value and governance efficiency issue (Moore, 1995; Milakovich, 2012), and one worth testing directly in future, adequately powered studies. Resources directed at structural refinements beyond a functional usability threshold might instead be evaluated against the specific experience dimensions the NCE pilot data tentatively suggest could matter most: cognitive load reduction (CL^{-1}), content relevance (R), and emotional resonance (EI). For private sector organizations and digital product teams, the implications are similarly worth

noting, with the same caution. User satisfaction surveys, susceptible to halo effects, social desirability bias, and post-hoc rationalization, may lead to overinvestment in interface refinements that improve conscious ratings without a proportional improvement in underlying adoption drivers. The NCE Index, as piloted here, is offered as one way to probe dimensions of engagement, cognitive load, and emotional investment that self-report may under-capture, pending validation in larger samples (Lalmas and Yom-Tom, 2022).

Overall, the results support H3 by demonstrating that **self-reported experience and real-time engagement represent related but distinct dimensions of UX**, both of which are necessary for a complete assessment.

5.4 EEG as a Measure of Cognitive and Emotional Variation

H4 proposed that EEG-based neurophysiological measures would capture variations in users' cognitive and emotional responses across different levels of UX quality. While the present pilot study did not provide statistical confirmation of this hypothesis, the observed descriptive patterns were directionally consistent with H4 and are discussed as exploratory findings rather than confirmatory evidence.

Across conditions, numerically consistent patterns emerged in key EEG indicators. The Theta/Alpha Ratio, frequently associated with cognitive engagement in previous research (i.e. Kamzanova et al., 2015), was highest in the Good UX condition ($M = 9.05$) compared with the Bad ($M = 7.59$) and Medium UX ($M = 7.54$) conditions. Similarly, Parietal Alpha Power was numerically higher for the Good UX condition ($M = 6.97$) than for the Bad ($M = 4.92$) and Medium UX ($M = 3.69$) conditions. While these differences cannot be interpreted as evidence of enhanced engagement or reduced cognitive strain, they follow a direction broadly consistent with previous neuroergonomics findings (i.e. Bonnefond & Jensen, 2025; Chen et al., 2020) and may warrant further investigation in larger samples.

Interestingly, the descriptive EEG patterns were not entirely mirrored by the conventional UX measures. Whereas self-report and audit-based evaluations primarily reflected conscious and perceived aspects of user experience, the EEG measures showed gradual numerical variation across conditions. Although these observations should not be interpreted as demonstrating greater sensitivity of EEG, they suggest that neurophysiological measures may capture complementary aspects of user processing that are not directly observable through conventional evaluation methods alone.

Block-level analyses also indicated that EEG responses varied descriptively across task contexts. During the Emergency Scenario, the Good UX condition exhibited numerically higher Theta activity than the Medium and Bad UX conditions, whereas differences appeared less pronounced during the Goal-Oriented Task. Given the exploratory nature of the pilot study and the absence of inferential testing at the block level, these observations should be regarded as preliminary indications of possible context-dependent neurophysiological responses rather than established effects.

The topographic visualizations likewise suggested differences in the spatial distribution of neural activity across UX conditions. Although these maps were not subjected to statistical comparison, they provide an initial qualitative indication that interface quality may influence patterns of neural activation in addition to overall signal magnitude. Future studies using larger samples and spatial statistical analyses will be required to determine whether these apparent differences reflect reliable neurophysiological effects.

Taken together, these findings do not provide confirmatory evidence that EEG differentiates UX quality in this pilot sample. Nevertheless, the observed descriptive trends suggest that EEG may offer a complementary perspective on users' cognitive and emotional responses during

interaction, supporting its continued investigation alongside conventional UX evaluation methods.

5.5 Eye-Tracking and Attention Allocation (H5)

H5 proposed that eye-tracking measures would reflect differences in attention allocation and navigation efficiency. The findings provide strong support for this hypothesis, demonstrating that behavioral metrics capture clear and actionable differences in user interaction across UX conditions.

Distinct patterns emerged in task performance, fixation behavior, and time on task. The Bad UX condition was consistently associated with unsuccessful task completion, elevated fixation counts (e.g., $M \approx 74\text{--}81$), and prolonged interaction times (up to ~ 110 seconds), indicating increased cognitive effort and inefficient navigation. These patterns reflect disorientation and difficulty in locating task-relevant information.

In contrast, both Medium and Good UX conditions demonstrated high completion rates (100%), reduced fixation counts (e.g., $\sim 33\text{--}41$), and shorter task durations (approximately 50–65 seconds), suggesting more efficient information processing and clearer navigation structures. Gaze patterns in these conditions were also more directed, indicating improved information architecture and reduced search effort.

Importantly, differences between Medium and Good UX were more nuanced. While Good UX facilitated faster access in some cases (e.g., visible entry points), it did not consistently outperform Medium UX across all tasks. In more complex navigation scenarios, additional steps or structural depth introduced measurable friction, reflected in increased fixations and longer completion times (e.g., ~ 71 seconds in Task 3).

These findings suggest that eye-tracking measures are particularly effective in capturing behavioral efficiency and interaction friction, offering direct insight into how users navigate and process interface structures (Wegner et al., 2020; Macdonald and Tatler, 2018). While poor UX clearly disrupts usability, improvements beyond a functional level appear to yield more incremental gains, highlighting the importance of optimizing not only structure but also task-specific pathways and content accessibility.

5.6 Recognition Memory and Cognitive Encoding

H6 proposed that recognition memory performance (PSRMT) and cognitive encoding of website content would vary according to UX quality. While the present pilot study was not designed to provide formal inferential comparisons across UX conditions, the findings revealed several descriptive patterns that were directionally consistent with this hypothesis. Accordingly, the interpretation below focuses on exploratory trends rather than confirmed statistical effects.

Overall, participants demonstrated relatively high recognition accuracy ($M = 80.7\%$, $SD = 6.6\%$) and moderate Memory Retention scores ($M = 0.610$, $SD = 0.139$), suggesting that the PSRMT was effective in capturing measurable recognition memory during the experimental session (Gehrke et al., 2018).

At the group level, participants in the Good UX condition showed the highest mean MR scores ($M = 0.664$), followed closely by the Medium UX condition ($M = 0.647$), whereas the Bad UX condition showed lower retention ($M = 0.428$). Although these descriptive differences should not be interpreted as evidence that UX quality caused the observed variation, the numerical pattern is consistent with previous research suggesting that more usable and well-structured interfaces may facilitate cognitive encoding and subsequent recognition memory (Grunwald & Corsbie-Massay, 2006).

The observed pattern was also broadly consistent with the behavioral and neurophysiological measures included in the multimodal framework. Differences between the Medium and Good UX conditions were relatively small, whereas both showed higher memory retention than the Bad UX condition, mirroring the descriptive pattern observed in the NCE Index. While these converging observations do not constitute confirmatory evidence, they provide preliminary support for the internal coherence of the proposed multimodal evaluation framework.

At the individual level, the lowest memory performance was observed in a participant assigned to the Bad UX condition (P7: MR = 0.337; accuracy = 69.0%). Although this represents a single observation and should not be overinterpreted, it follows the overall descriptive trend observed across the pilot sample.

Taken together, these preliminary findings do not establish that UX quality influences recognition memory performance. Rather, they suggest that recognition memory measures such as the PSRMT may provide a useful complementary perspective within multimodal UX evaluation. Future studies with larger samples and condition-specific inferential analyses are needed to determine whether the descriptive patterns observed here represent reliable effects, particularly in domains such as e-government, where successful interaction depends on users' ability to locate, understand, and retain digital information (Jia et al., 2026; Zhang et al., 2025; Aldrees & Gračanin, 2023).

5.7 The Value of a Multimodal Approach

H7 proposed that a multimodal approach combining neurophysiological, behavioral, and self-reported measures would provide a more comprehensive assessment of user experience than any single method alone. The present findings offer strong empirical support for this hypothesis and represent one of the central contributions of the study.

The value of multimodality becomes particularly evident when comparing how individual methods capture UX quality. When considered in isolation, each measurement approach provides a partial and method-specific perspective:

The UX Audit clearly differentiated all three websites (28%, 59%, 84%), but as an expert-based evaluation, it reflects structural quality rather than actual user experience during interaction.

Self-reported measures (SRQ) successfully identified Poor UX as the lowest-performing condition but suggested a different ranking between Medium and Good UX (Medium: M = 6.13, Good: M = 5.00), highlighting that perceived experience does not always align with expert evaluations.

EEG measures captured continuous variation in cognitive engagement and effort, revealing subtle differences across conditions, but without a single standalone metric fully distinguishing between UX levels.

The Implicit Memory Test (IMT) provided evidence of differential content encoding (Good: M = 0.664, Medium: M = 0.647, Bad: M = 0.428), yet as a global measure, it cannot independently resolve UX quality differences across conditions.

When these measures are integrated through the NCE Index, a more coherent and informative pattern emerges. The composite results clearly differentiate the Bad UX condition (M = 3.52) from the other two, while revealing near-identical engagement levels between Medium and Good UX (M = 5.20 vs M = 5.07; $\Delta = 0.13$). This pattern is not visible in any single method alone.

This convergence is critical: it demonstrates that user experience is not a single-layer construct, but the outcome of interacting perceptual, behavioral, and neurocognitive processes. Methods such as self-reports capture conscious evaluation, eye-tracking reflects observable behavior, EEG reveals underlying cognitive dynamics, and memory measures capture encoding outcomes. Only through their integration can these layers be meaningfully aligned.

Importantly, the multimodal approach does not simply aggregate data; it resolves contradictions between methods. For example, while audit scores and subjective ratings suggest a clear superiority of Good UX, the integrated neurocognitive profile indicates that Medium and Good UX produce comparable levels of engagement. This finding challenges the assumption that improvements in structural UX quality necessarily translate into proportionally improved user experience.

This divergence between self-reported evaluations and neurocognitive measures can be interpreted through the lens of dual-process theories of cognition. Self-reported ratings primarily reflect **System 2 processing**, involving conscious, reflective judgment, whereas EEG, eye-tracking, and implicit memory measures capture aspects of **System 1 processing**, which operates automatically and often outside conscious awareness (i.e. Da Silva, 2023; Wickens and Carswell, 2021; Simon, 1981). The present findings suggest that while users are able to consciously evaluate interface quality, their underlying cognitive and emotional engagement may remain unaffected. This reinforces the idea that UX is not solely a reflective judgment, but a multi-layered process shaped by both conscious and unconscious mechanisms.

Taken together, the findings support the growing view in UX and Neuroimaging research that methodological triangulation is essential for capturing the full complexity of human–computer interaction. In applied contexts such as e-government, where user experience directly impacts accessibility, trust, and service effectiveness, multimodal frameworks offer a more accurate and actionable basis for evaluation than any single method alone.

5.8 The NCE Index as a Novel Evaluative Contribution

Beyond the specific hypotheses discussed above, the primary methodological contribution of this study is the introduction of the Neurocognitive Engagement (NCE) Index as a multimodal framework for UX evaluation. Rather than relying on a single measurement domain, the NCE Index represents an initial attempt to operationalize user experience as a multidimensional construct by integrating both self-reported and less consciously accessible indicators within a transparent, fully disclosed evaluative framework (Section 2.6).

The index combines six complementary dimensions—Emotional Intensity (EI), Engagement (ENG), Cognitive Load (CL^{-1}), Relevance (R), Memory Retention (MR), and Self-Reported Quality (SRQ)—each reflecting a different aspect of human–computer interaction. By integrating neurophysiological (EEG), behavioral (eye-tracking), cognitive (PSRMT), and subjective (SAM and SRQ) measures, the NCE Index is intended to capture complementary dimensions of user experience that may not be fully represented when these methods are considered independently.

The pilot application provides initial evidence for the conceptual coherence and practical feasibility of the proposed framework rather than validation of the index itself. Within the present sample, the NCE Index differentiated the Bad UX condition ($M = 3.52$) from the Medium and Good UX conditions, which clustered closely together ($M = 5.20$ vs. 5.07 ; $\Delta = 0.13$). This descriptive pattern remained stable across the sensitivity analyses (Section 3.6.4) and differed from the ranking obtained through the UX Audit (28%, 59%, 84%). Although these observations require replication in larger samples, they suggest that improvements in

expert-rated UX quality may not necessarily correspond to proportional differences across the multimodal indicators incorporated within the proposed framework.

One possible interpretation of this descriptive pattern is the existence of a functional engagement threshold, whereby improvements beyond a certain level of usability are associated with comparatively smaller changes in the multimodal measures contributing to the NCE Index. In the present pilot sample, the Medium UX condition appeared to reach levels of multimodal engagement broadly comparable to those observed for the Good UX condition, despite their substantially different UX Audit scores. This interpretation should be regarded as exploratory and hypothesis-generating rather than as an established empirical finding.

If similar patterns are observed in future studies, they may have implications for the evaluation of public-sector digital services. Specifically, they could suggest that incremental improvements in expert-rated usability do not always correspond to proportional changes across multimodal indicators of user experience. Under this perspective, optimization efforts may benefit from focusing not only on overall usability scores but also on specific experience dimensions, such as reducing cognitive load, improving content relevance, and supporting goal-directed navigation.

More broadly, the present work contributes to UX methodology by proposing a transparent multimodal framework that integrates neurophysiological, behavioral, cognitive, and subjective measures within a single composite index. Rather than positioning the NCE Index as a replacement for established UX evaluation methods, the framework is intended to complement existing approaches by combining multiple sources of evidence into a unified representation of user experience. As such, the primary contribution of this study lies not in validating a new measurement instrument, but in demonstrating the feasibility of integrating heterogeneous measures within a coherent methodological framework for UX evaluation.

The modular architecture of the NCE Index was designed to facilitate adaptation across different research and applied settings, allowing individual components to be modified according to available measurement tools and evaluation objectives. As demonstrated in this pilot study, the framework provides one possible approach for integrating multimodal data into a single interpretable metric. The complete computational framework—including the index formula, weighting strategy, and operationalization of each component—is fully disclosed (Section 2.6) to support transparency, replication, and further methodological refinement by future researchers.

Future research should focus on validating the NCE Index across larger and more diverse samples, examining its psychometric properties, and evaluating its applicability across different digital environments and interaction contexts. Viewed from this perspective, the present study should be considered a methodological proof-of-concept that establishes a foundation for the continued development and empirical validation of multimodal UX evaluation frameworks.

5.9 Implications for UX Evaluation in Local e-Government

The findings of this pilot study offer preliminary implications for the evaluation and design of municipal digital platforms, both in Greece and in broader e-government contexts. Given the exploratory nature of the study, these implications should be viewed as directions for future research and validation rather than established recommendations.

First, the descriptive divergence observed between UX Audit scores (28%, 59%, 84%) and the NCE Index (3.52, 5.20, 5.07) suggests that structural UX audits, although valuable, may not fully reflect the cognitive and emotional dimensions of users' experiences. These preliminary observations support further investigation into whether public-sector evaluation frameworks

could benefit from combining expert-based assessments with measures that capture users' real-time cognitive and behavioral responses.

Second, the findings highlight the potential value of broadening UX evaluation beyond task completion and self-reported satisfaction. The inclusion of measures such as cognitive load, emotional intensity, and recognition memory offers additional perspectives on how users process, interpret, and retain digital information. While these measures require further validation, they may be particularly relevant in e-government environments, where successful interaction depends on efficient information processing and decision making.

Third, the study demonstrates the practical feasibility of implementing portable consumer-grade neurophysiological equipment (Emotiv Insight) within an applied UX research protocol. Although the pilot study was not designed to establish the effectiveness or scalability of these methods, it shows that multimodal neurophysiological data can be collected and integrated alongside conventional UX measures outside highly controlled laboratory environments. This provides an encouraging foundation for future applied research using similar approaches.

Fourth, this study introduces the Neurocognitive Engagement (NCE) Index as a transparent multimodal framework for integrating neurophysiological, behavioral, cognitive, and subjective measures within a single evaluative approach. Rather than replacing established UX methods, the proposed framework is intended to complement existing evaluation practices by integrating multiple sources of evidence into a unified representation of user experience. Its modular architecture and fully disclosed computational framework are designed to facilitate replication, critical evaluation, and methodological refinement by future researchers.

From a public-sector perspective, the descriptive patterns observed in this pilot study also raise an important question for future investigation. If similar findings are replicated in larger studies, investment decisions based primarily on structural UX audits and satisfaction measures may not always correspond to proportional differences across multimodal indicators of user experience. Exploring this relationship further could contribute to more informed prioritization of digital service improvements and public digitalization investments.

Overall, the present findings support continued investigation into integrated, multimodal approaches to UX evaluation as a complement to conventional assessment methods in digital governance. Although the proposed framework requires further validation, this pilot study demonstrates the feasibility of combining multiple measurement domains within a single methodological approach and provides a foundation for future research in evidence-informed UX evaluation.

5.10 Limitations and Future Research Directions

The findings of this study should be interpreted in light of several limitations inherent to its pilot design. While these limitations constrain the statistical interpretation and generalizability of the findings, they do not diminish the primary methodological contribution of the study, namely the development and pilot implementation of the proposed multimodal UX evaluation framework.

First, the study involved a small convenience sample ($N = 10$) consisting exclusively of adult male participants. Consequently, the findings should be regarded as exploratory rather than confirmatory and cannot be generalized beyond the present sample. Future research should include larger, gender-balanced samples to enable adequately powered statistical analyses.

Second, EEG recordings were obtained using a portable five-channel Emotiv Insight device. Although suitable for demonstrating the feasibility of the proposed framework in an applied

setting, future studies would benefit from higher-density EEG systems with access to raw signals and more advanced preprocessing techniques.

Third, EEG data were collected using static website screenshots rather than live interaction. Accordingly, the recorded responses primarily reflect early visual processing and first impressions rather than continuous navigation. Future work should extend the framework to real-time interaction scenarios.

Fourth, eye-tracking data were manually coded by a single researcher, and the Recognition Memory Task (PSRMT) was administered once per participant rather than separately for each UX condition. These design choices limited condition-specific analyses and should be refined in future implementations through automated eye-tracking analysis and condition-specific memory assessments.

Fifth, the current version of the NCE Index applies equal weighting across its six components. Although the sensitivity analysis suggested that the overall descriptive pattern remained stable across most alternative weighting schemes, future studies should empirically validate component weights using larger datasets and appropriate psychometric methods.

Finally, each UX category was represented by a single municipal website. Consequently, some observed differences may reflect characteristics of the selected websites rather than UX quality categories themselves. Future studies should include multiple websites per category using counterbalanced experimental designs.

Overall, the present study should be viewed as a methodological proof-of-concept rather than a validation study. Future research should focus on replicating and refining the proposed multimodal framework across larger samples, more diverse digital environments, and higher-resolution measurement techniques to further evaluate the applicability of the Neurocognitive Engagement (NCE) Index.

6. Conclusion

6.1 Summary of Findings

This study contributes to the growing field of neuroscience-informed UX research by proposing and piloting a transparent multimodal framework for evaluating digital user experience. The Neurocognitive Engagement (NCE) Index represents an initial methodological attempt to integrate neurophysiological, behavioral, cognitive, and subjective measures within a single evaluation framework. While the pilot findings are exploratory and require validation in larger samples, they demonstrate the feasibility of this approach and provide a foundation for future research aimed at developing more comprehensive methods for UX evaluation across public and commercial digital systems.

6.2 Tentative Implications for Practice

With these limitations in mind, the present findings suggest three directions for future investigation in municipal digital governance rather than established recommendations. First, expert-based UX audits could be complemented by behavioral or neurocognitive measures to further examine whether improvements in structural UX quality correspond to broader dimensions of user experience. Second, the component profile of the proposed NCE Index may serve as a hypothesis-generating framework for exploring the relative contribution of cognitive load, content relevance, emotional intensity, and memory retention to digital user experience. Third, neuroscience-informed UX evaluation may represent a promising complementary

approach for future assessment of public digital services, provided that its reliability, validity, and practical applicability are established through larger-scale empirical studies.

6.3 Future Research Directions

Future research should focus on validating and refining the proposed multimodal framework through larger, gender-balanced samples, live-interaction paradigms, automated eye-tracking analysis, condition-specific recognition memory assessment, and the inclusion of multiple websites within each UX category. Further work is also needed to empirically evaluate the weighting and psychometric properties of the NCE Index to support its broader applicability across different digital contexts.

Acknowledgments

The authors would like to sincerely thank Optotech Company for their generous support and collaboration throughout this study. Their willingness to provide access to eye-tracking equipment was essential for the successful implementation of the behavioral component of the research. The authors are especially grateful to the Optotech team for their hands-on support during the data collection process, as well as for their contribution to the preparation of experimental materials, including video analysis and task-related assets. Their expertise and practical assistance significantly enhanced the quality and feasibility of this work. This study would not have been possible without their support.

References

- Albayrak, D. and Çagiltay, K., 2013, October. Analyzing Turkish e-government websites by eye tracking. In *2013 Joint Conference of the 23rd International Workshop on Software Measurement and the 8th International Conference on Software Process and Product Measurement* (pp. 225-230). IEEE. Doi: 10.1109/IWSM-Mensura.2013.39.
- Albert, B. and Tullis, T., 2022. *Measuring the user experience: Collecting, analyzing, and presenting UX metrics*. Morgan Kaufmann.
- Aldrees, A. and Gračanin, D., 2023. UX in E-government Services for Citizens: A Systematic Literature Review. *Journal of User Experience*, 18(3).
- Ali, M., 2023. E-governance and E-democracy: a Digital Revolution. Available at SSRN 4623414. Available at SSRN: <http://dx.doi.org/10.2139/ssrn.4623414>.
- Allam, A.H. and Dahlan, H.M., 2013. User experience: challenges and opportunities. *Journal of information systems research and innovation*, 3(1), pp.28-36.
- Alzahrani, L., Al-Karaghoul, W. and Weerakkody, V., 2017. Analysing the critical factors influencing trust in e-government adoption from citizens' perspective: A systematic review and a conceptual framework. *International business review*, 26(1), pp.164-175. Available at: <https://doi.org/10.1016/j.ibusrev.2016.06.004>
- Anastasopoulou, E., Gkika, K., Tsagri, A., Travlou, C., Spanou, A. and Katsivelou, V., 2025. ICT in Greek Primary Education: A Literature Review. *Asian Journal of Education and Social Studies*, 51(6), pp.959-973. Available at <https://doi.org/10.9734/ajess/2025/v51i62049>
- Antonenko, P., Paas, F., Grabner, R. and Van Gog, T., 2010. Using electroencephalography to measure cognitive load. *Educational psychology review*, 22(4), pp.425-438. Available at: <https://doi.org/10.1007/s10648-010-9130-y>

- Ardito, C., Buono, P., Caivano, D., Costabile, M.F. and Lanzilotti, R., 2014. Investigating and promoting UX practice in industry: An experimental study. *International Journal of Human-Computer Studies*, 72(6), pp.542-551. Available at: <https://doi.org/10.1016/j.ijhcs.2013.10.004>
- Bataineh, E., Al Mourad, B. and Kammoun, F., 2017, July. Usability analysis on Dubai e-government portal using eye tracking methodology. In *2017 Computing conference* (pp. 591-600). IEEE. doi: 10.1109/SAI.2017.8252156
- Baumeister, R.F., Masicampo, E.J. and Vohs, K.D., 2011. Do conscious thoughts cause behavior?. *Annual review of psychology*, 62(1), pp.331-361. Available at: <https://doi.org/10.1146/annurev.psych.093008.131126>
- Bell, L., Vogt, J., Willemse, C., Routledge, T., Butler, L.T. and Sakaki, M., 2018. Beyond self-report: A review of physiological and neuroscientific methods to investigate consumer behavior. *Frontiers in psychology*, 9, p.1655. Available at: <https://doi.org/10.3389/fpsyg.2018.01655>
- Bellos, C., Stefanou, K., Tzallas, A., Stergios, G. and Tsipouras, M., 2025. Methods and approaches for user engagement and user experience analysis based on electroencephalography recordings: A systematic review. *Electronics*, 14(2), p.251. Available at: <https://doi.org/10.3390/electronics14020251>
- Bonnefond, M. and Jensen, O., 2025. The role of alpha oscillations in resisting distraction. *Trends in Cognitive Sciences*, 29(4), pp.368-379. DOI: 10.1016/j.tics.2024.11.004
- Camilleri, M.A., 2020. The online users' perceptions toward electronic government services. *Journal of Information, Communication and Ethics in Society*, 18(2), pp.221-235. Available at: <https://doi.org/10.1108/JICES-09-2019-0102>
- Carter, L. and Bélanger, F., 2005. The utilization of e-government services: citizen trust, innovation and acceptance factors. *Information systems journal*, 15(1), pp.5-25. Available at: <https://doi.org/10.1111/j.1365-2575.2005.00183.x>Digital Object Identifier
- Chairistanidis, E. and Rusu, L., 2024, September. Barriers to Digital Transformation in Public Organizations: A Case Study of a Greek Public Organization. In *European, Mediterranean, and Middle Eastern Conference on Information Systems* (pp. 335-349). Cham: Springer Nature Switzerland.
- Chartrand, T.L., 2005. The role of conscious awareness in consumer behavior. *Journal of Consumer Psychology*, 15(3), pp.203-210. Available at: https://doi.org/10.1207/s15327663jcp1503_4Digital Object Identifier (DOI)
- Chatzopoulos, K.C. and Economides, A.A., 2009. A holistic evaluation of Greek municipalities' websites. *Electronic Government, an International Journal*, 6(2), pp.193-212. Available at <https://doi.org/10.1504/EG.2009.024442>
- Chen, Z., He, Y. and Yu, Y., 2020. Attention restoration during environmental exposure via alpha-theta oscillations and synchronization. *Journal of Environmental Psychology*, 68, p.101406. Available at: <https://doi.org/10.1016/j.jenvp.2020.101406>
- Choudrie, J., Tsatsou, P. and Kurnia, S. eds., 2018. Social inclusion and usability of ICT-enabled services. New York City: Routledge.
- Chourmouziadis, I., 2024, June. Unveiling the Digital Odyssey: Explaining the Challenges Behind eVoting Adoption in Greece.
- Cohen, J.R., 2010. Decoding developmental differences and individual variability in response inhibition through predictive analyses across individuals. *Frontiers in human neuroscience*, 4, p.1287. Available at <https://doi.org/10.3389/fnhum.2010.00047>

- Corneille, O. and Gawronski, B., 2024. Self-reports are better measurement instruments than implicit measures. *Nature Reviews Psychology*, 3(12), pp.835-846. Available at: <https://doi.org/10.1038/s44159-024-00376-z>
- Dan, A. and Reiner, M., 2017. Real time EEG based measurements of cognitive load indicates mental states during learning. *Journal of Educational Data Mining*, 9(2), pp.31-44. Available at: <https://doi.org/10.5281/zenodo.3554719>
- Da Silva, S., 2023. System 1 vs. system 2 thinking. *Psych*, 5(4), pp.1057-1076. Available at: <https://doi.org/10.3390/psych5040071>
- Demir, Ö., Tuluk, A., Ay, K., Özçelik, F. and Tüzün, H., 2017. The evaluation of E-government web portal in Turkey through eye-tracking method. In *ECER (European Conference on Educational Research)*.
- Dimitrelou, S. and Fouskas, K., 2023. Digital Government Platforms in Greece. Current Trends: The Case of Gov. gr. *KnE Social Sciences*, pp.335-359. Doi: 10.18502/kss.v8i1.12655
- Dumas, J.S. and Salzman, M.C., 2006. Usability assessment methods. *Reviews of human factors and ergonomics*, 2(1), pp.109-140. Available at: <https://doi.org/10.1177/1557234X0600200105>
- Ellis, N.C., 2005. At the interface: Dynamic interactions of explicit and implicit language knowledge. *Studies in second language acquisition*, 27(2), pp.305-352. doi:10.1017/S027226310505014X
- European Commission (2010). A digital agenda for Europe. Publications Office of the European Union.
- Fernandes, F.R., Paschoarelli, L.C. and Da Silva, J.C.P., 2013, July. Evaluating interaction with websites: Case study of a government website of the Brazilian Ministry of labor and employment. In *International Conference of Design, User Experience, and Usability* (pp. 265- 273). Berlin, Heidelberg: Springer Berlin Heidelberg. Available at https://doi.org/10.1007/978-3-642-39253-5_29.
- Fontanini, A. and Katz, D.B., 2008. Behavioral states, network states, and sensory response variability. *Journal of neurophysiology*, 100(3), pp.1160-1168. Available at: <https://doi.org/10.1152/jn.90592.2008>
- Frey, J., Daniel, M., Castet, J., Hachet, M. and Lotte, F., 2016, May. Framework for electroencephalography-based evaluation of user experience. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (pp. 2283-2294). Available at: <https://doi.org/10.1145/2858036.2858525>
- Galvez, R.A. and Youngblood, N.E., 2016. e-Government in Rhode Island: what effects do templates have on usability, accessibility, and mobile readiness?. *Universal Access in the Information Society*, 15(2), pp.281-296. Available at <https://doi.org/10.1007/s10209-014-0384-x>
- Gehrke, L., Iversen, J.R., Makeig, S. and Gramann, K., 2018, July. The invisible maze task (IMT): Interactive exploration of sparse virtual environments to investigate action-driven formation of spatial representations. In *German Conference on Spatial Cognition* (pp. 293-310). Cham: Springer International Publishing. Available at: https://doi.org/10.1007/978-3-319-96385-3_20
- Georgios, K. and Nikolaos, V., 2021. Reinstating greek e-governance: a framework for e-Government Benchmarking, improvement and government policies. *Вопросы государственного и муниципального управления*, (6), pp.103-127. Doi: 10.17323/1999-5431-2021-0-6-103-127.
- Greenwald, A.G. and Banaji, M.R., 2017. The implicit revolution: Reconceiving the relation between conscious and unconscious. *American Psychologist*, 72(9), p.861.

- Grunwald, T. and Corsbie-Massay, C., 2006. Guidelines for cognitively efficient multimedia learning tools: educational strategies, cognitive load, and interface design. *Academic medicine*, 81(3), pp.213-223. Available at: <https://doi.org/10.1097/00001888-200603000-00003>
- Halderman, L.K., Finn, B., Lockwood, J.R., Long, N.M. and Kahana, M.J., 2021. EEG correlates of engagement during assessment. *ETS Research Report Series*, 2021(1), pp.1-17. Available at: <https://doi.org/10.1002/ets2.12312>Digital Object Identifier (DOI)
- Hassenzahl, M. and Tractinsky, N., 2006. User experience-a research agenda. *Behaviour & information technology*, 25(2), pp.91-97. Available at: <https://doi.org/10.1080/01449290500330331>
- Hekler, E.B., Klasnja, P., Froehlich, J.E. and Buman, M.P., 2013, April. Mind the theoretical gap: interpreting, using, and developing behavioral theory in HCI research. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 3307-3316). Available at: <https://doi.org/10.1145/2470654.2466452>
- Hodge, H., Carson, D., Carson, D., Newman, L. and Garrett, J., 2017. Using Internet technologies in rural communities to access services: The views of older people and service providers. *Journal of Rural Studies*, 54, pp.469-478. Available at: <https://doi.org/10.1016/j.jrurstud.2016.06.016>
- Jia, Z., Li, J., Kang, Y., Wang, Y., Wu, T., Wang, Q., Wang, X., Zhang, S., Shen, J., Li, Q. and Qi, S., 2026. The AI Hippocampus: How Far are We From Human Memory?. *arXiv preprint arXiv:2601.09113*. Available at: <https://doi.org/10.48550/arXiv.2601.09113>
- Jun, K.N., Wang, F. and Wang, D., 2014. E-Government use and perceived government transparency and service capacity: Evidence from a Chinese local government. *Public Performance & Management Review*, 38(1), pp.125-151. Available at <https://doi.org/10.2753/PMR1530-9576380106>.
- Kamzanova, A., Matthews, G. and Kustubayeva, A., 2015. EEG Indices of Task Engagement. *Farabi Journal of Social Sciences*, 1(1).
- Katsikas, S.K. and Gritzalis, S., 2017. Digitalization in Greece: State of play, barriers, challenges, solutions. In *Beyond Bureaucracy: Towards Sustainable Governance Informatisation* (pp. 355- 375). Cham: Springer International Publishing. Available at https://doi.org/10.1007/978-3-319-54142-6_19.
- Kujala, S., Roto, V., Väänänen-Vainio-Mattila, K., Karapanos, E. and Sinnelä, A., 2011. UX Curve: A method for evaluating long-term user experience. *Interacting with computers*, 23(5), pp.473-483. Available at: <https://doi.org/10.1016/j.intcom.2011.06.005>
- Kuniavsky, M., 2003. *Observing the user experience: a practitioner's guide to user research*. Elsevier.
- Lalmas, M., O'Brien, H. and Yom-Tov, E., 2022. *Measuring user engagement*. Springer Nature.
- Macdonald, R.G. and Tatler, B.W., 2018. Gaze in a real-world social interaction: A dual eye-tracking study. *Quarterly Journal of Experimental Psychology*, 71(10), pp.2162-2173. Available at: <https://doi.org/10.1177/1747021817739221>
- Macmillan, N.A., 2002. Signal detection theory. *Stevens' handbook of experimental psychology: Methodology in experimental psychology*, 3, pp.43-90.
- Macmillan, N.A. and Kaplan, H.L., 1985. Detection theory analysis of group data: estimating sensitivity from average hit and false-alarm rates. *Psychological bulletin*, 98(1), p.185.
- Moore, S.R. and Depue, R.A., 2016. Neurobehavioral foundation of environmental reactivity. *Psychological bulletin*, 142(2), p.107.
- Nam, T., 2014. Determining the type of e-government use. *Government information quarterly*, 31(2), pp.211-220. Available at: <https://doi.org/10.1016/j.giq.2013.09.006>

- Nuamah, J.K. and Seong, Y., 2018. Support vector machine (SVM) classification of cognitive tasks based on electroencephalography (EEG) engagement index. *Brain-Computer Interfaces*, 5(1), pp.1-12. Available at: <https://doi.org/10.1080/2326263X.2017.1338012>
- Nulhusna, R., Sandhyaduhita, P.I., Hidayanto, A.N. and Phusavat, K., 2017. The relation of e-government quality on public trust and its impact on public participation. *Transforming Government: People, Process and Policy*, 11(3), pp.393-418. Available at: <https://doi.org/10.1108/TG-01-2017-0004>
- Panayiotou, N. and Stavrou, V., 2019. A proposed maturity assessment framework of the Greek local government Web Electronic Services. *Transforming Government: People, Process and Policy*, 13(3/4), pp.237-256. Available at <https://doi.org/10.1108/TG-03-2019-0018>.
- Patergiannaki, Z. and Pollalis, Y.A., 2024. E-government quality from the citizen's perspective: the role of perceived factors, demographic variables and the digital divide. *International Journal of Public Sector Management*, 37(2), pp.232-254. Available at: <https://doi.org/10.1108/IJPSM-07-2023-0229>
- Patergiannaki, Z., Tsatsani, O. and Pollalis, Y., 2020. Analysing the key factors impacting on the digitalization and penetration of e-Government: A case study analysis of Greek Municipalities. In *XIV. International Conference on e-Government and e-Justice, Amsterdam*.
- Patergiannaki, Z. and Pollalis, Y.A., 2022. A review of the e-Government maturity models: are they still relevant?. *Journal of Politics and Ethics in New Technologies and AI*, 1(1), pp.e30872- e30872. Available at <https://doi.org/10.12681/jpentai.30872>.
- Pérez-Morote, R., Pontones-Rosa, C. and Núñez-Chicharro, M., 2020. The effects of e-government evaluation, trust and the digital divide in the levels of e-government use in European countries. *Technological forecasting and social change*, 154, p.119973. Available at: <https://doi.org/10.1016/j.techfore.2020.119973>
- Punchoojit, L. and Hongwarittorn, N., 2017. Usability studies on mobile user interface design patterns: a systematic literature review. *Advances in Human-Computer Interaction*, 2017(1), p.6787504. Available at: <https://doi.org/10.1155/2017/6787504>Digital Object Identifier (DOI)
- Reddick, C.G., 2011. Citizen interaction and e-government: Evidence for the managerial, consultative, and participatory models. *Transforming Government: People, Process and Policy*, 5(2), pp.167-184. Available at: <https://doi.org/10.1108/17506161111131195>
- Rodríguez Bolívar, M.P., Navarro Galera, A., Alcaide Muñoz, L. and López Subirés, M.D., 2016. Risk factors and drivers of financial sustainability in local government: An empirical study. *Local Government Studies*, 42(1), pp.29-51. Available at <https://doi.org/10.1080/03003930.2015.1061506>.
- Schwarz, N., 2012. Why researchers should think" real-time": A cognitive rationale.
- Scott, J.K., 2006. "E" the people: Do US municipal government web sites support public involvement?. *Public administration review*, 66(3), pp.341-353. Available at <https://doi.org/10.1111/j.1540-6210.2006.00593.x>.
- Simon, H.A., 1981. Information-processing models of cognition. *Journal of the American Society for Information Science*, 32(5), pp.364-377. Available at: <https://doi.org/10.1002/asi.4630320517>Digital Object Identifier (DOI)
- Sonderegger, A., Uebelbacher, A. and Sauer, J., 2019, August. The UX construct—does the usage context influence the outcome of user experience evaluations?. In *IFIP Conference on Human-Computer Interaction* (pp. 140-157). Cham: Springer International Publishing. Available at: https://doi.org/10.1007/978-3-030-29390-1_8

- Spinellis, D., Pouloudi, A., Xirogiannis, G. and Makantasi, E., 2022. Intelligent modeling of eGovernment initiatives in Greece. Available at <http://eprints.lse.ac.uk/id/eprint/113730>.
- Süleymanov, E., 2025. Generation Z in Digitalizing Labor Processes: A Phenomenological Study Based on Private Sector Experiences. *MALUMAT*, (1), pp.45-68. Available at: <https://doi.org/10.65298/malumat.4>
- Tobias, B.A., Kihlstrom, J.F. and Schacter, D.L., 2014. Emotion and implicit memory. In *The handbook of emotion and memory* (pp. 67-92). Psychology Press.
- Tsatsani, O., Patergiannaki, Z. and Pollalis, Y., 2024. Unraveling the Dynamics of e-Government Digitization, Penetration and User Experience: A Case Study of Greek Municipalities. *Journal of Strategic Innovation & Sustainability*, 19(3). DOI: 10.33423/jsis.v19i3.7376
- Tsatsani, O. and Pollalis, Y., 2025. E-Government and User Experience: Evidence from Greek municipalities. *SPOUDAI Journal of Economics and Business*, 75(3-4), pp.35-79. Available at: <https://spoudai.org/index.php/journal/article/view/593>
- Tucker, D.M., Luu, P. and Johnson, M., 2022. Neurophysiological mechanisms of implicit and explicit memory in the process of consciousness. *Journal of neurophysiology*. Available at: <https://doi.org/10.1152/jn.00328.2022>
- Voutinioti, A., 2013. Determinants of user adoption of e-Government services in Greece and the role of citizen service centres. *Procedia Technology*, 8, pp.238-244. Available at <https://doi.org/10.1016/j.protcy.2013.11.033>.
- Wegner, S., Lohmeyer, Q., Wahlen, D., Neumann, S., Groebli, J.C. and Meboldt, M., 2020. Value of eye-tracking data for classification of information processing-intensive handling tasks: Quasi-experimental study on cognition and user interface design. *JMIR human factors*, 7(2), p.e15581. DOI:10.2196/15581
- Welch, E.W., Hinnant, C.C. and Moon, M.J., 2005. Linking citizen satisfaction with e-government and trust in government. *Journal of public administration research and theory*, 15(3), pp.371-391. Available at: <https://doi.org/10.1093/jopart/mui021>
- Wickens, C.D. and Carswell, C.M., 2021. Information processing. *Handbook of human factors and ergonomics*, pp.114-158. Available at: <https://doi.org/10.1002/9781119636113.ch5>Digital Object Identifier (DOI)
- Wiedbusch, M., Dever, D., Li, S., Amon, M.J., Lajoie, S. and Azevedo, R., 2023. Measuring multidimensional facets of SRL engagement with multimodal data. In *Unobtrusive observations of learning in digital environments: Examining behavior, cognition, emotion, metacognition and social processes using learning analytics* (pp. 141-173). Cham: Springer International Publishing. Available at: https://doi.org/10.1007/978-3-031-30992-2_10
- Wixted, J.T., 2020. The forgotten history of signal detection theory. *Journal of experimental psychology: learning, memory, and cognition*, 46(2), p.201.
- Xie, B., Charness, N., Fingerma, K., Kaye, J., Kim, M.T. and Khurshid, A., 2021. When going digital becomes a necessity: Ensuring older adults' needs for information, services, and social inclusion during COVID-19. In *Older adults and COVID-19* (pp. 181-191). Routledge.
- Yang, Y., 2025. Heterogeneous effects of e-government on urban and rural public services: the perspective of the digital divide. *Applied Economics*, pp.1-16. Available at: <https://doi.org/10.1080/00036846.2025.2586293>
- Yildiz, M., Ocak, N., Yildirim, C., Cagiltay, K. and Babaoglu, C., 2020. Usability in local e-government: analysis of Turkish metropolitan municipality Facebook pages. In *Open Government: Concepts, Methodologies, Tools, and Applications* (pp. 966-984). IGI Global. DOI: 10.4018/978-1-5225-9860-2.ch045

- Yu, S., Mückschel, M. and Beste, C., 2022. Superior frontal regions reflect the dynamics of task engagement and theta band-related control processes in time-on task effects. *Scientific Reports*, 12(1), p.846. Available at: <https://doi.org/10.1038/s41598-022-04972-y>
- Zaelani, R., Taufikurohman, M.R., Astuti, S.Y. and Setiawan, E., 2025. Transformation of Digital Consumption Patterns of Millennials and Generation Z: Sustainable Creative Economy Growth in the Age of Technology. *International Journal of Advanced Technology and Social Sciences*, 3(10), pp.1351-1366. Available at: <https://doi.org/10.59890/ijatss.v3i10.117>
- Zhang, Z., Zhang, Y., Tan, H., Li, R. and Chen, X., 2025. Explicit vs implicit memory: Exploring multi-hop complex reasoning over personalized information. *arXiv preprint arXiv:2508.13250*. Available at: <https://doi.org/10.48550/arXiv.2508.13250>
- Zhang, J., Chang, D. and Zhang, Z., 2021, December. Review on the Application of Eye-tracking Technology in Usability Evaluation of E-government Apps. In *2021 IEEE International Conference on Industrial Engineering and Engineering Management (IEEM)* (pp. 1646-1650). IEEE. DOI 10.1109/IEEM50564.2021.9672792.
- Zheng, Y. and Schachter, H.L., 2017. Explaining citizens' E-participation use: The role of perceived advantages. *Public Organization Review*, 17(3), pp.409-428. Available at: <https://doi.org/10.1007/s11115-016-0346-2>
- Zuiderwijk, A., Janssen, M. and Dwivedi, Y.K., 2015. Acceptance and use predictors of open data technologies: Drawing upon the unified theory of acceptance and use of technology. *Government information quarterly*, 32(4), pp.429-440. Available at: <https://doi.org/10.1016/j.giq.2015.09.005>