HEC MONTRÉAL

A Multimodal Approach to Identifying and Predicting Usability Pain Points: An Experimental Study in User Experience Research

par
Kelvin Jacinto

Sylvain Sénécal
HEC Montréal
Codirecteur de recherche

Constantinos Coursaris
HEC Montréal
Codirecteur de recherche

Sciences de la gestion
(Spécialisation Expérience Utilisateur)

Mémoire présenté en vue de l'obtention
du grade de maîtrise ès sciences en gestion
(M. Sc.)

December 2024

Résumé

La complexité croissante des systèmes numériques a mis en évidence les limites des évaluations traditionnelles de la facilité d'utilisation, en particulier celles qui reposent exclusivement sur des mesures autodéclarées sujettes à des biais tels que la mémorisation et la désirabilité sociale. Cette étude répond au besoin de méthodes plus objectives en examinant les biosignaux qui offrent des mesures précises des réactions corporelles afin d'identifier et de prédire les problèmes d'utilisabilité dans le contexte des interfaces numériques. Elle a exploré des signatures psychophysiologiques distinctes d'utilisateurs interagissant avec une application numérique et a évalué leur fiabilité pour prédire les problèmes d'utilisabilité dans des environnements d'entreprises commerciales.

L'étude a introduit une approche multimodale utilisant des mesures psychophysiologiques combinées de l'excitation émotionnelle, de la valence, de la charge cognitive et de l'attention visuelle de 86 participants qui ont effectué des tâches dans trois systèmes d'entreprise. Des points douloureux ont été introduits artificiellement dans des tâches contrôlées afin de susciter des réponses psychophysiologiques. L'analyse en grappes a révélé quatre profils d'utilisateurs distincts pour ces points douloureux artificiellement induits. L'étude a utilisé la régression logistique pour former des modèles prédictifs permettant d'identifier le moment où les utilisateurs rencontrent des points de douleur lors d'une tâche naturelle.

Cette étude expérimentale a impliqué 86 participants, chacun chargé d'effectuer des interactions sur trois plateformes SaaS sélectionnées. Les participants ont été exposés à des perturbations manipulées de l'utilisabilité conçues pour évoquer des réponses naturelles à des points de douleur de l'utilisabilité, et des données ont été collectées sur leur éveil émotionnel, leur valence, leur charge cognitive et leur attention visuelle. En utilisant une combinaison d'outils non invasifs, tels qu'un oculomètre, des capteurs EDA et la reconnaissance des expressions faciales (FER), l'étude a suivi les réponses psychophysiologiques des utilisateurs en temps réel. La recherche a suivi une approche multimodale, intégrant plusieurs mesures psychophysiologiques pour développer des modèles prédictifs capables de détecter et de prévoir avec précision les points de douleur liés à l'utilisabilité. Une analyse en grappes a été réalisée pour identifier les groupes de participants sur la base de leurs réponses psychologiques. Ensuite, les modèles prédictifs ont été entraînés à l'aide de la régression logistique et évalués à l'aide de mesures de rappel et de précision. Enfin, les performances des modèles prédictifs ont été validées par une évaluation d'experts.

Les principaux résultats comprennent l'identification de signatures psychophysiologiques uniques et le succès prédictif modéré des modèles utilisant la dilatation de la pupille et le coefficient k comme indicateurs significatifs. Malgré la variabilité individuelle et les défis de précision modérés, ces résultats ont démontré la faisabilité de l'utilisation de mesures psychophysiologiques pour l'évaluation de l'utilisabilité en temps réel.

Cette recherche a permis de mieux comprendre les réactions des utilisateurs aux problèmes de convivialité dans les environnements des entreprises commerciales. Elle a mis en évidence le potentiel des données psychophysiologiques dans l'évaluation de la convivialité en temps réel. Elle a abordé les défis posés par l'utilisation d'évaluations auto-déclarées.

Mots-clés : signatures psychophysiologiques, utilisabilité, points de douleur, modèle prédictif, analyse en grappes.

Abstract

The increasing complexity of digital systems has highlighted the limitations of traditional usability assessments, particularly those relying exclusively on self-reported measures prone to biases such as recall and social desirability. This study addressed the need for more objective methods by examining biosignals which offers precise measurements of bodily reactions to identify and predict usability challenges in the context of digital interfaces. It explored distinct psychophysiological signatures of users interacting with a digital application and evaluated their reliability in predicting usability issues in business enterprise environments.

The study introduced a multimodal approach using combined psychophysiological measures of emotional arousal, valence, cognitive load and visual attention from 86 participants who performed tasks in three enterprise systems. Pain points were introduced artificially in controlled tasks to elicit psychophysiological responses. Cluster analysis revealed four distinct user profiles to these artificially induced pain points. The study used logistic regression to train predictive models to identify when users encounter usability pain points on a natural task.

This experimental study involved 86 participants, each tasked with completing interactions on three selected SaaS platforms. Participants were exposed to manipulated usability disruptions designed to evoke natural responses to usability pain points, with data collected on their emotional arousal, valence, cognitive load, and visual attention. Using a combination of non-invasive tools—such as an eye tracker, EDA sensors, and Facial Expression Recognition (FER)—the study tracked users' psychophysiological responses in real-time. The research followed a multimodal approach, integrating several psychophysiological measures to develop predictive models that could accurately detect and forecast usability pain points. A cluster analysis was performed to identify group of participants based on their psychological responses. Then, the predictive models were trained using logistic regression and evaluated using recall and precision metrics. Lastly, the predictive models' performance was validated through expert evaluation.

The key results included the identification of unique psychophysiological signatures and the moderate predictive success of models using pupil dilation and k-coefficient as significant indicators. Despite individual variability and moderate precision challenges, these results demonstrated the feasibility of using psychophysiological measures for real-time usability assessment.

This research advanced the understanding of user responses to usability pain points in business enterprise environments. It underscored the potential for psychophysiological data in real-time usability evaluation. It addressed the challenges when using self-reported assessment.

Keywords : psychophysiological signatures, usability, pain points, predictive model, cluster analysis

Table of contents

List of tables and figures

List of abbreviations and acronyms

CRM	Customer Relations Management
ECG	Electrocardiogram
EDA	Electrodermal Activity
EEG	Electroencephalography
ERP	Enterprise Resource Planning
FER	Facial Expression Recognition
HCI	Human Computer Interaction
HRMS	Human Resource Management System
LC-NE	Locus coeruleus-norepinephrine
PM	Predictive Model
SaaS	Software as a Service
UPP	Usability Pain Point
UX	User Experience
VR	Virtual Reality

Preface

My academic and professional background has always centered on the intersection of technology and human behavior. With a strong foundation in Human Resources and User Experience, I have been fascinated by how users, specifically employees, interact with digital systems, particularly in business enterprise environments. This fascination became the driving force behind my pursuit of understanding the challenges and opportunities within enterprise systems—a realm where usability and efficiency often collide.

The inspiration for this research stemmed from observing how seemingly minor usability issues can have profound implications in business settings, from reduced productivity to employee frustration and even attrition. Witnessing these challenges firsthand in conversations with professionals motivated me to explore how advanced methodologies, like psychophysiological measures, could provide unique insights into user experiences and help alleviate these challenges.

This work is significant because it addresses a critical gap in usability research: the need for objective, real-time data that reflects users' emotional, cognitive, and attentional states. While traditional usability methods rely heavily on subjective feedback, this research explores the potential of psychophysiological data to offer deeper, actionable insights. I believe that these findings have implications beyond academia, benefiting businesses striving for more user-centered systems and fostering more productive and satisfying work environments.

The primary audience for this work includes researchers and practitioners in the fields of human-computer interaction, user experience, psychophysiology, and enterprise system design. It may also resonate with professionals tasked with technology adoption and implementation in organizational settings.

In the chapters that follow, readers can expect a deep dive into the methodology, findings, and implications of leveraging psychophysiological data to identify usability pain points. This thesis explored how psychophysiological signatures uncover various user behavior and predict usability challenges, ultimately contributing to more effective system design.

One particularly interesting insight from this research is the discovery of distinct psychophysiological signatures linked to user experiences. These signatures not only confirm the emotional and cognitive impact of usability challenges but also highlight the potential for predictive modeling in enhancing user experience. It is my hope that this work will inspire further exploration into the integration of psychophysiological data in usability research and system development.

Acknowledgements

This thesis represents the culmination of a challenging yet rewarding journey that would not have been possible without the support, guidance, and encouragement of many individuals to whom I am deeply grateful.

First and foremost, I extend my heartfelt thanks to my supervisors, Pierre-Majorique Léger, Sylvain Sénécal, Constantinos Coursaris, Alexander Karran and Thaddé Rolon-Mérette for their invaluable expertise, thoughtful feedback and unwavering support throughout this process. Their insights and encouragement have influenced the direction and quality of this research.

I would also like to express my gratitude to the Tech3Lab Research Operations Team for their support during the data collection phase of my study. A special thanks to Shang Lin for assisting me in the data analysis phase, Emilie Dubois for collaborating with me on preparing the project's stimuli, and Andrada Toma for being an excellent project partner. I am also thankful to the study participants, whose willingness to contribute their time and effort made this research possible. Their participation provided the critical data that underpin the findings of this work.

Thank you to my colleagues and peers for the stimulating discussions, feedback, and moral support that enriched my research journey.

A warm thank you to my family and friends, who have stood by me through this journey, offering encouragement and patience during moments of stress and uncertainty. Your belief in me has been my most significant source of determination.

Finally, I acknowledge Tech3Lab, HEC Montreal, NSERC, PROMPT, and Deloitte for their financial support, which enabled me to conduct this research and complete this thesis.

To everyone who contributed in ways both big and small—thank you. This work is as much a reflection of your support as it is of my efforts

Chapter 1
Introduction

1.1 Background

The ongoing evolution of digital systems has led to their increased complexity, requiring sophisticated methods for assessing and optimizing user experience (UX). While the increasing popularity of enterprise software solutions such as Enterprise Resource Planning (ERP) and Software-as-a-Service (SaaS) platforms have revolutionized business processes (Klaus et al., 2000; Haselmann & Vossen, 2011; Yusuf et al., 2018), their usability remains a critical challenge. In Canada, the SaaS market is experiencing significant growth due to trends in collaboration-focused markets, CRM and HRMS software adoption, customer preferences for cloud-based solutions, government support for digital transformation and a diverse business landscape such as remote work (Statista). With the rise of enterprise software solutions, research on user experience (UX) in these digital systems need to be adequately addressed in UX research.

The International Organization for Standardization (2019) defined UX as "the user's perceptions and responses resulting from using or anticipating a system, product or service." The concepts of usability assessment are widely used to measure user experience to ensure that design solutions are practical and centered around user needs. This user-centered design philosophy emphasizes the importance of understanding user experiences and expectations, which can lead to more effective and engaging digital products (Klotins et al., 2018; Berni et al., 2023). Research showed that poor usability in digital interfaces and applications, such as health tracking platforms, can result in user abandonment (Saleh et al., 2021). Moreover, Pittet and Barthélemy (2015) highlighted the importance of optimizing user interface design and tailoring SaaS systems to users' preferences, essential for enhancing user experience. Factors such as user expectations, environmental conditions, and the system's capacity to meet the current user needs impact overall user sentiment (Roto & Kaasinen, 2008).

1.2 Research Problem and Gap

User Experience (UX) research has historically relied heavily on self-reported assessments, such as surveys, interviews, and questionnaires (Bargas-Avila & Hornbæk, 2011; Inan Nur et al.,2021; Perrig et al., 2024), to understand users' perceptions, emotions, and interactions with products and services (Brunn et al., 2016). Studies like those by Bargas-Avila and Hornbæk (2011) and Inan Nur et al. (2021) have demonstrated the dominance of these methods, with over 50% of UX studies employing questionnaires. However, while self-reported assessments are perceived as practical and easy to administer (Law et al., 2013), they are prone to various biases that undermine the insights' accuracy, such as social desirability and recall bias (Kwak et al., 2021; Vrijheid et al., 2008; Horwitz et al., 2024), which can distort the accuracy of the data. Therefore, one of the primary challenges is finding reliable and objective ways to assess usability challenges experienced by users beyond self-reporting.

1.3 Significance of the Study

Psychophysiological signatures are objective biological signals that reflect an individual's emotional, cognitive, and psychological states, derived from physiological responses such as heart rate variability, skin conductance, and brain activity (Chang et al., 2015; Ajenaghughrure et al., 2020). These signatures hold significant potential in complementing self-reported assessments, which are often susceptible to biases and inaccuracies stemming from subjective interpretation and social desirability effects (Ajenaghughrure et al., 2020; Wager et al., 2013). For instance, while self-reports can offer insights into an individual's conscious experience and emotional states, psychophysiological measures provide a more objective perspective, allowing researchers to observe the underlying biological processes that correspond to these experiences (Chang et al., 2015; Visser et al., 2017). While psychophysiological measures may not always precisely capture bodily reactions, they provide a valuable complementary perspective. For instance, Barreto et al. (2007) demonstrated the effectiveness of these measures in detecting stress levels through heart rate variability. Similarly, Maia and Furtado (2019) used galvanic skin response to monitor emotional responses, while

Ferreira et al. (2014) assessed cognitive load during interactions using electroencephalography (EEG).

Despite the growing interest in psychophysiological approaches to usability assessment (Apraiz et al., 2021), limited research integrated multiple psychophysiological measures to capture a user's real-time response to usability challenges comprehensively. Furthermore, to our knowledge, prior research has yet to develop predictive models that could anticipate usability issues in future tasks. Therefore, integrating psychophysiological methods and predictive models to detect usability challenges would be a promising avenue for enhancing real-time usability monitoring and system optimization.

The study's primary purpose is to explore and leverage psychophysiological data to identify and predict usability challenges users encounter when interacting with a business enterprise platform. Specifically, this research identified distinct users' psychophysiological response patterns when encountering usability issues. Moreover, the study examined the reliability of the uncovered psychophysiological signature to identify usability obstacles that users experienced in another similar task. This study provided a novel methodology for improving real-time usability assessment by addressing these objectives.

1.4 Implications

This study has significant implications for both academic and practical domains. The study enriched our understanding of user responses to usability challenges by introducing a multimodal approach to usability assessment. In the real world, this research offered a valuable framework for system designers, managers, and UX professionals to anticipate usability obstacles, enabling them to prioritize usability improvements based on objective, real-time data proactively. Such advancements can enhance employee satisfaction, reduce training costs, and increase the efficiency of enterprise systems, aligning with organizational goals.

1.5 Theoretical Framework

The study is grounded in established psychophysiological theories of emotional and cognitive responses, including the arousal-valence model and adaptive gain theory. These frameworks helped interpret the psychophysiological signatures captured in the study, linking changes in emotional arousal, valence, cognitive load, and visual attention to specific user experiences during task performance. The study built a robust conceptual foundation for understanding usability challenges by situating the research within these theoretical models.

1.6 Methodological Approach

The research employed a mixed-subjects experimental design involving 86 participants who performed various tasks on business enterprise platforms. Psychophysiological data were collected around the artificially manipulated usability disruption, which was placed on Task 1 and Task 2, including electrodermal activity, facial expressions, pupil dilation, and k-coefficient. Cluster analysis was used to identify distinct psychophysiological response patterns to the usability disruptions experienced by users during their interactions with a digital interface, and logistic regression models were trained to predict the occurrence of these disruptions on a natural task. The models were tested on a task free of manipulated usability obstacles, Task 3, to evaluate their performance and reliability.

1.7 Scope and Limitations

The study focused on business SaaS platforms, specifically Microsoft Dynamics 365, Salesforce, and ServiceNow. While the research provided novel insights into real-time usability assessment, its findings were constrained by the controlled laboratory setting, specific stimuli and tasks used and the moderate predictive accuracy of the models. Future research should validate these findings in diverse environments, other business enterprise platforms and task sets and explore advanced machine-learning techniques to improve model reliability.

This research study used the term "*usability pain point (UPP)*" to refer to the psychophysiological response to usability challenges experienced by users during their interaction with a digital interface.

1.8 Thesis Structure

This thesis started with an introduction to the context of the study, including the problem statement, importance, and objectives. This is followed by an in-depth literature review, examining prior work on conventional usability assessments and psychophysiological methods in UX research. Chapter 3 presented the scientific article prepared to be submitted to the journal Computers in Human Behavior Reports. This article introduced a novel multimodal approach to identifying and predicting usability pain points, a detailed experimental method was presented, and key findings were discussed. In Chapter 4 of this thesis, a short managerial article was written, which included a summary of the study, the key findings, and best practices and recommendations. The last chapter provided a thesis conclusion that provides a summary of the entire study.

This thesis was completed in the Tech3Lab, involving multiple collaborators with differing input levels throughout various stages. The student's intellectual contributions to each part of the thesis are detailed in the **Table 1** below.

Table 1.
Student's contribution and responsibilities in the realization of this thesis

Stage in the process	Contribution
Research Question	Identified gaps in current literature and defined the research problem [80%] • Defined research questions • Identified the constructs to be tested
Literature Review	Conducted relevant literature search, read scientific articles relevant to the research. [100%]
Experimental Design	Applied to the Research Ethics Board (REB) [60%] • Prepared documentation related to the submission of the application to the CER Developed experimental protocol and stimuli [80%] • Created experimental protocol, questionnaires, task instructions, short onboarding videos for the experiment • Determined the tasks to be performed by the participants on the stimuli

	• In collaboration with a research assistant, configured stimuli to apply the artificial pain point
Recruitment, Pre-testing and Data Collection	Recruited participants for data collection [20%] • Provided inclusion and exclusion criteria for participant recruitment (The Tech3Lab operations team oversaw the guidelines, collected data using the institution's recruitment panel, and distributed the compensation for this study) • Coordinated participant's schedules; this includes cancellation, rescheduling, and other requests. Managed Pre-testing and data collection [100%] • Oversaw the data collection and managed participants' experience during the study • Monitored and managed stimuli assignment (randomly assigned, but switching from one stimulus to the other is not done automatically and required manual intervention; this includes applying the manipulated pain point to the stimuli and removing it after)
Data Analysis	Prepared data for analysis and analyzed the results [60%] (The data file for the analysis statistics was formatted by the lab statistician)
Writing the thesis	Wrote the thesis and the articles [100%] (The student was guided by their supervisor with their constructive feedback through the process)

*These percentages did not consider the support and input of the directors during this project.

Chapter 2

Literature review

2.1 Challenges in Usability Assessment

User experience (UX) relies heavily on self-reported assessment to gather subjective data about users' feelings, perceptions, and experiences with products or services. This is heavily supported by research paper reviews conducted in the past and recent years. A study conducted by Bargas-Avila & Hornbæk (2011), which critically analyzed 66 empirical studies on UX conducted between 2005 and 2009, found that the dominant method used was questionnaires appearing in 53% of the studies reviewed. Inan Nur et al. (2021) found that 95% used self-reported measures for UX evaluation on the 61 research papers reviewed from 2000 to 2019. Also, a recent systematic review by Perrig et al. (2024), which screened 153 research papers from the ACM Conference on Human Factors in Computing Systems proceedings from 2019 to 2022, identified 85 survey scales used in the reviewed research papers. Indeed, self-reported assessments are widely used in UX research.

However, this reliance on self-reported measures has its challenges. Law et al. (2013) explored the attitudes of UX researchers and practitioners toward UX measurements and found that while most respondents' views are generally positive, UX professionals showed mixed feelings and were often skeptical of self-reported measures. Self-reported assessments were seen as practical for capturing subjective experiences but were criticized for potential bias as they rely on users' interpretations and memory (Law et al., 2013).

One of the primary biases in self-reported assessments is the social desirability bias, where users may provide answers, they believe are more socially acceptable rather than their true feelings or experiences (Nederhof, 1985). For instance, a study conducted by Kwak et al. (2021), which examined social desirability bias focusing on survey-based studies that deal with mobile loafing – non-related mobile internet use during work hours, found that social desirability bias significantly affected mobile internet addiction and mobile-loafing intentions. According to the study's findings, respondents would underreport their perceptions of mobile internet addiction and their intentions of committing mobile loafing, as this will make them look bad (Kwak et al., 2021). In the context of UX, this bias can lead to inflated reports as users may feel compelled to present themselves in a more favourable light, thereby skewing the data collected from surveys or interviews.

Additionally, recall bias is another significant challenge affecting the reliability of self-reported assessments. This bias occurs when users have difficulty remembering past experiences and behaviour, leading to inaccurate responses. A study conducted by Vrijheid et al. (2008) examined the recall bias in self-reported mobile phone use and found that participants tend to underestimate the number of calls they made by 19% and overestimate call duration by roughly 40%. The study concluded that recall bias posed challenges for accurate risk assessment in epidemiological studies relying on self-reported mobile phone usage data (Vrijheid et al., 2008). Tapping to the recall bias is the peak-end rule. The peak-end rule refers to the phenomenon where user's retrospective evaluations of past affective experiences are heavily influenced by the most intense moment, the "peak," and the final moment, the "end" of the experience (Kahneman et al., 1993). A recent study conducted by Horwitz et al. (2024) examined the peak-end rule and found that it significantly affects retrospective mental health assessments. The study suggested that retrospective self-reports of symptoms are often aligned with the peak experiences of distress rather than the average daily experience, which can lead to recall biases in clinical assessments (Horwitz et al., 2024). The recall bias and peak-end rule are problematic in UX research, where understanding interactions over time is critical for practical UX evaluation.

Going back to the study conducted by Bargas-Avila & Hornbæk (2011), their review showed that the majority of UX assessments were conducted after interaction (70%), 58% of the reviewed papers included during interaction assessment, and before interaction assessments were rare at only 20%. Whereas Inan Nur et al.'s (2021) findings showed that most UX evaluations took place after interaction, making up 44 studies out of 61 reviewed papers, with fewer assessing UX during interaction (12 studies) or over long-term interactions (6 studies). The context in which self-reported assessments are collected can also introduce bias. For example, in virtual reality (VR) studies, the transition between immersive experiences and the physical world when collecting self-reported assessments can disrupt users' presence and can cause participants to provide feedback that does not accurately reflect their true feelings during the immersive experience (Putze et al., 2020; Alexandrovsky et al., 2020), thus compromising the validity of the findings.

2.2 Shift to psychophysiological methods in usability assessment

These UX challenges highlight the need for more reliable assessment methods. Thus, in recent years, the exploration of psychophysiological measures has gained significant traction in UX research, supported by prior research reviews on UX papers. Based on Bargas-Avila & Hornbæk's (2011) findings, physiological measures were less common, appearing only in 5% of the 66 empirical studies reviewed from 2005 to 2009. Meanwhile, Inan Nur et al. (2021) found that 14% of the 61 research papers reviewed from 2000 to 2019 included physiological measures. Also, a recent review paper by Apraiz Iriarte et al. (2021) systematically examined a total of 33 research studies that applied physiological measures in UX evaluations spanning from 2006 to 2020, with a notable increase in publications from 2016 onwards reflects the growing interest in incorporating psychophysiological measures in UX research.

Psychophysiological measures provide a non-invasive and implicit approach to understanding a user's emotional or cognitive processes (Dirican & Göktürk, 2011). The psychophysiology theories examine the complex relationships between psychological processes and physiological responses (Lovallo, 2013) by understanding emotional states, cognitive processes, and mental well-being through measuring and interpreting physiological indicators like heart rate, skin conductance, and brain activity (Dair et al., 2023).

2.3 Prior research using psychophysiological measures in UX research

Recent advancements in HCI and UX research have highlighted the potential of psychophysiological measures to provide a more objective and nuanced understanding of the user's experience. One example is the work of Barreto et al. (2007), where the study demonstrated the use of non-intrusive psychophysiological measures like galvanic skin response, blood volume pulse (BVP), skin temperature, and pupil diameter combined with machine learning to effectively detect stress in real-time.

2.3.1 Emotional Response

Emotion is one of the most researched aspects of a user's response measured through psychophysiological methods. Arousal and valence are the most studied dimensions of emotion and commonly utilized psychophysiological measures of a user's emotion (Partala & Kangaskorte, 2009). A widely used model to represent emotion is the arousal-valence model (Russell, 1980), which describes emotion as a two-dimensional space: arousal (vertical axis) and valence (horizontal axis). The valence dimension represents a range of emotions from negative to neutral to positive, while the arousal dimension ranges from calm to neutral to aroused (Partala & Kangaskorte, 2009). With arousal and valence all having values in the same range, between -1 and 1, the arousal-valence model is widely used and effectively describes a person's emotional change (Yang & Sun, 2017). In 2013, Alexandros and Michalis (2013) proposed using heart rate, EDA, respiration rate, and muscle tension to analyze the duration, intensity, and transitions between emotional states during interactions. Recent research by Maia and Furtado (2019) highlights the use of electroencephalography (EEG), electrodermal activity (EDA), and heart rate to capture emotional states. The study showed significant correlations between psychophysiological signals and emotional dimensions during pleasure-driven tasks (Maia & Furtado, 2019). The study by Vignaux et al. (2021) examined the impact of collective immersion in a learning environment on emotional engagement. The study used psychophysiological methods, EDA and electrocardiogram (ECG), to measure emotional engagement, and the results showed greater emotional engagement in immersive and collective environments (Vignaux et al. 2021). Furthermore, a study conducted by

Swoboda et al. (2022) highlighted the effectiveness of speech and physiological measures in detecting emotional responses during interactions with voice user interfaces.

2.3.2 Cognitive Process

Another construct to assess user's response is cognitive load. Cognitive load theory posits that working memory has a limited capacity that can be easily overwhelmed by excessive information or complex tasks (Sweller et al., 1998). In human factors literature, cognitive load is defined as the quantity of mental activity necessary to execute a task and is commonly termed as mental workload, mental effort, or mental demand in the field (Van Acker et al., 2018). In 2014, Ferreira et al. (2014) concluded that a real-time cognitive load assessment is feasible for both younger and older adults using low-cost, non-invasive physiological sensors. Vanneste et al. (2020) directly examined how multimodal physiological measures can assess cognitive load, mainly through EDA, EEG, and eye tracking, and indirectly measuring emotional arousal via EDA. The study concluded that combining these three physiological measures provides a nuanced picture of cognitive load by effectively capturing both arousal and mental components (Vanneste et al., 2020). A recent study by Hudon et al. (2021) investigated how different visualization methods for explaining AI predictions impact user cognitive load and confidence in AI systems. The study measured cognitive load using pupillary dilation, precisely the task-evoked pupillary response (TEPR), which is a well-established proxy for cognitive effort (Hudon et al., 2021).

Pupil size has been associated with cognitive processes (Kucewicz et al., 2018) and has shown that pupil dilation increases with increasing task demands (Van Der Wel & Van Steenbergen, 2018). Prior research studies provided empirical evidence that the locus coeruleus-norepinephrine (LC-NE) system regulates task engagement, which correlates with pupil size fluctuations (Gilzenrat et al., 2010; Murphy et al., 2014; Hopstaken et al., 2015). The Adaptive Gain Theory posits that the LC-NE system operates in phasic and tonic modes (Aston-Jones & Cohen, 2005). Phasic mode is characterized by moderate NE levels and intense stimulus-triggered bursts of NE release, associated with high task engagement, where attention is concentrated on task-relevant stimuli to optimize performance (Minzenberg et al., 2008). In tonic mode, both baseline and stimulus-induced NE levels are elevated, which is associated with disengagement of the current task, where attention is no longer primarily focused on task-relevant stimuli but also responds to irrelevant stimuli (Cohen et al., 2007). In relation, Gilzentrat et al. (2010) explored how pupil diameter corresponds to the LC-NE modes, where larger baseline pupils indicate tonic mode (task disengagement) and smaller pupils indicate phasic mode (task engagement).

2.3.3 Visual Attention Behaviour

Moreover, visual attention is another aspect of the user's behaviour measured through psychophysiological methods. When engaging in a specific task, users actively seek, gather, share, and consume information in their environment. This is aligned with Pirolli and Card's (1999) information foraging theory, which assumes that individuals maximize their rate of gaining valuable information by modifying their strategies or the structure of the environment. In real-life computer-based tasks, users are required to allocate their attention effectively by focusing on the most critical aspects of the display and ignoring the rest (Wals & Wichary, 2022). Visual attention is a selection process that allows certain stimuli to be processed more thoroughly than others (Lamme, 2003). Krejtz et al. (2016) introduced a novel non-invasive visual search measure to characterize ambient and focal visual attention modes. Building up on Krejtz's findings, a recent study by Lounis et al. (2020) assessed visual attention in pilots during different flight phases (take-off, cruise, and landing) by tracking eye movement using novel eye-tracking device, Tobii Pro Glasses. Moreover, Carmichael et al. (2022) explored how information disclosure nudges affect users' information disclosure behaviours when interacting with chatbots. The study measured visual attention using eye-tracking technology and validated that the information disclosure nudges successfully drew participants' attention, ensuring their potential to influence user behaviour (Carmichael et al., 2022).

2.4 The use of psychophysiological measures in pain point detection

Platzer's (2018) genealogical approach to trace the historical development of the term "pain point" in business and UX context found that "pain point" refers to specific user problems or frustrations that design changes can alleviate. UX professionals use this term to elicit stakeholder empathy and prioritize user needs. Recent research studies by Kreger (2022) and Huo et al. (2023) used the term "pain point," where Kreger (2022) implicitly used the term to refer to specific moments in the user experience that lead to frustration, confusion, or difficulty when interacting with digital banking services and Huo (2023) used the term to refer to the specific interaction touchpoints within the in-vehicle human-machine interfaces (HMI) where users experience lower levels of emotional satisfaction. Giroux-Huppé et al. (2019) work differentiated explicit and implicit pain points. Where explicit pain points are consciously acknowledged negative emotions reported during or after a task, and implicit pain points, also termed *psychophysiological pain points*, are automatic physiological responses characterized by high emotional arousal and negative emotional valence in reaction to an event during the interaction (Giroux-Huppé et al., 2019). Kreger (2022) suggested failure mapping to identify and resolve user pain points, and Huo (2023) combines Kansei Engineering, which quantifies emotional reactions to design, with user experience mapping to identify and improve areas of user dissatisfaction. However, neither work used any psychophysiological measures. Meanwhile, the seminal work of Giroux-Huppé et al. (2019) used psychophysiological measures to identify psychophysiological pain points in online grocery shopping. Giroux-Huppé et al. (2019) distinguished between explicit pain points, consciously acknowledged negative emotions, and implicit pain points, characterized by automatic physiological responses. Their work demonstrated that psychophysiological measures, such as heightened arousal and negative valence, offer a real-time, objective approach to capturing user frustration, surpassing the limitations of traditional self-reported methods.

2.5 Prior research on measuring multi aspects of user’s response

Although Giroux-Huppé et al. (2019) research introduced a novel approach to identifying pain points accurately in real-time, the approach concentrated on one aspect of the user's response: the emotion the users felt. Prior research has used different psychophysiological methods to assess various aspects of user response separately. Using a multimodal approach to measure two or more constructs has been gaining traction in

UX research recently. Prior study by Léger et al. (2014) combined eye tracking with electroencephalography (EEG) to improve temporal precision in measuring attentional, cognitive, and motor processes of participants who interacted with a system involving email notifications during a primary task. Charland et al. (2015) used psychophysiological tools to measure the key dimensions of engagement, behavioral, cognitive, and emotional, during learning tasks, suggesting that the combination of these psychophysiological tools allowed differentiation of low, medium, high levels of engagement, providing a comprehensive understanding of how learners interact with tasks. Korosec-Serfaty et al. (2022) investigated how technostress and financial stress in digital financial technology usage impacts users' emotional and cognitive responses by analyzing psychophysiological, perceptual, and behavioural data. Parsons et al. (2023) used psychophysiological measures to assess, in real time, users' cognitive and emotional states in virtual environments. Also, Mithun et al. (2023) introduced *Mind Indriya*, a composite system using psychophysiological measures to assess the cognitive load, anxiety, and visual attention in real-time. These are a few examples of research studies that used a multimodal approach and assessed two or more factors using psychophysiological measures.

This literature underscores the limitations of self-reported usability assessments, including biases and their inability to capture implicit responses during interaction. While psychophysiological measures offer promising complement to self-reported assessments, prior research has often focused narrowly on isolated constructs like emotional arousal in assessing usability challenges encountered by users in a digital environment. Moreover, the potential of these psychophysiological responses to identify usability challenges across tasks must be further addressed. This study bridged these gaps by employing a multimodal approach to simultaneously capture emotional, cognitive, and attentional dimensions of users' responses to usability challenges and developing predictive models to evaluate their reliability across tasks. This comprehensive framework advances usability research by offering a more nuanced and actionable understanding of user experiences.

For this research study, the term "*usability pain point (UPP)*" referred to the usability challenges experienced by users during their interaction with a digital interface, which gives rise to automatic psychophysiological responses characterized by abnormal changes in emotional arousal, valence, visual attention, and cognitive load.

References

Alexandros, L., & Michalis, X. (2013). The physiological measurements as a critical indicator in users' experience evaluation. In *Proceedings of the 17th Panhellenic Conference on Informatics* (pp. 258-263).

Alexandrovsky, D., Putze, S., Bonfert, M., Höffner, S., Michelmann, P., Wenig, D., Malaka, R., & Smeddinck, J. D. (2020). Examining Design Choices of Questionnaires in VR User Studies. *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, 1–21. https://doi.org/10.1145/3313831.3376260

Apraiz Iriarte, A., Lasa, G., & Mazmela, M. (2021). Evaluating User Experience with physiological monitoring: A Systematic Literature Review. *Dyna (Bilbao)*, 8, 21. https://doi.org/10.6036/NT10072

Aston-Jones, G., & Cohen, J. D. (2005). AN INTEGRATIVE THEORY OF LOCUS COERULEUS-NOREPINEPHRINE FUNCTION: Adaptive gain and optimal performance. *Annual Review of Neuroscience*, 28(1), 403–450. https://doi.org/10.1146/annurev.neuro.28.061604.135709

Bargas-Avila, J. A., & Hornbæk, K. (2011). Old wine in new bottles or novel challenges: A critical analysis of empirical studies of user experience. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 2689–2698. https://doi.org/10.1145/1978942.1979336

Barreto, A., Zhai, J., & Adjouadi, M. (2007). Non-intrusive physiological monitoring for Automated stress detection in Human-Computer Interaction. In *Springer eBooks* (pp. 29–38). https://doi.org/10.1007/978-3-540-75773-3_4

Bruun, A., Law, E. L.-C., Heintz, M., & Alkly, L. H. A. (2016). Understanding the Relationship between Frustration and the Severity of Usability Problems: What can Psychophysiological Data (Not) Tell Us? *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, 3975–3987. https://doi.org/10.1145/2858036.2858511

Carmichael, L., Poirier, S.-M., Coursaris, C., Léger, P.-M., & Senecal, S. (2022). Users' Information Disclosure Behaviors during Interactions with Chatbots: The Effect of Information Disclosure Nudges. *Applied Sciences*, 12, 12660. https://doi.org/10.3390/app122412660

Charland, P., Léger, P.-M., Senecal, S., Courtemanche, F., Mercier, J., Skelling, Y., & L. LeMoyne, E. (2015). Assessing the Multiple Dimensions of Engagement to Characterize Learning: A Neurophysiological Perspective. *Journal of Visualized Experiments*, 101. https://doi.org/10.3791/52627

Cohen, J. D., McClure, S. M., & Yu, A. J. (2007). Should I stay or should I go? How the human brain manages the trade-off between exploitation and exploration. *Philosophical Transactions of the Royal Society B Biological Sciences*, 362(1481), 933–942. https://doi.org/10.1098/rstb.2007.2098

Dair, Z., Dockray, S., & O'Reilly, R. (2023). Complex Adaptive Systems and Psychophysiological Data - An Exploratory Approach. *2023 31st Irish Conference on Artificial Intelligence and Cognitive Science (AICS)*., 1-4. https://doi.org/10.1109/AICS60730.2023.10470799.

Dirican, A. C., & Göktürk, M. (2011). Psychophysiological measures of human cognitive states applied in human computer interaction. *Procedia Computer Science*, 3, 1361–1367. https://doi.org/10.1016/j.procs.2011.01.016

Ferreira, E., Ferreira, D., Kim, S., Siirtola, P., Röning, J., Forlizzi, J. F., & Dey, A. K. (2014). Assessing real-time cognitive load based on psycho-physiological measure for younger and older adults. 2014 *IEEE Symposium on Computational Intelligenc Cognitive Algorithms, Mind, and Brain* (CCMB), 39–48. https://doi.org/10.1109/CCMB.2014.7020692

Gartner. (n.d.) *Fueling the future of business*. (n.d.). Gartner. https://www.gartner.com/document/4023333?ref=solrAll&refval=420472053

Gilzenrat, M. S., Nieuwenhuis, S., Jepma, M., & Cohen, J. D. (2010). Pupil diameter tracks changes in control state predicted by the adaptive gain theory of locus coeruleus function. *Cognitive Affective & Behavioral Neuroscience*, 10(2), 252–269. https://doi.org/10.3758/cabn.10.2.252

Giroux-Huppé, C., Sénécal, S., Fredette, M., Chen, S. L., Demolin, B., & Léger, P. (2019). Identifying psychophysiological pain points in the online user journey: the case of online grocery. In *Lecture notes in computer science* (pp. 459–473). https://doi.org/10.1007/978-3-030-23535-2_34

Haselmann, T., & Vossen, G. (2011). Software-as-a-Service in Small and Medium Enterprises: An Empirical Attitude assessment. *In Lecture notes in computer science* (pp. 43–56). https://doi.org/10.1007/978-3-642-24434-6_4

Hopstaken, J. F., Van Der Linden, D., Bakker, A. B., & Kompier, M. A. (2015). The window of my eyes: Task disengagement and mental fatigue covary with pupil dynamics. *Biological Psychology*, 110, 100–106. https://doi.org/10.1016/j.biopsycho.2015.06.013

Horwitz, A. G., McCarthy, K., & Sen, S. (2024). A review of the peak-end rule in mental health contexts. *Current Opinion in Psychology*, 101845. https://doi.org/10.1016/j.copsyc.2024.101845

Hudon, A., Demazure, T., Karran, A., Léger, P.-M., & Senecal, S. (2021). *Explainable Artificial Intelligence (XAI): How the Visualization of AI Predictions Affects User Cognitive Load and Confidence* (pp. 237–246). https://doi.org/10.1007/978-3-030-88900-5_27

Huo, F., Zhao, Y., Chai, C., & Fang, F. (2023). A user experience map design method based on emotional quantification of in-vehicle HMI. *Humanities and Social Sciences Communications*, 10. https://doi.org/10.1057/s41599-023-01761-4

Inan Nur, A., B. Santoso, H., & O. Hadi Putra, P. (2021). The Method and Metric of User Experience Evaluation: A Systematic Literature Review. *Proceedings of the 2021 10th International Conference on Software and Computer Applications*, 307–317. https://doi.org/10.1145/3457784.3457832

ISO 9241-210:2019(en), Ergonomics of human-system interaction — *Part 210: Human-centred design for interactive systems.* (n.d.). https://www.iso.org/obp/ui/en/#iso:std:iso:9241:-210:ed-2:v1:en

Kahneman, D., Fredrickson, B. L., Schreiber, C. A., & Redelmeier, D. A. (1993). When more pain is preferred to less: adding a better end. *Psychological Science*, 4(6), 401–405. https://doi.org/10.1111/j.1467-9280.1993.tb00589.x

Klaus, H., Rosemann, M. & Gable, G.G. (2000). What is ERP?. *Information Systems Frontiers 2*, 141–162. https://doi.org/10.1023/A:1026543906354

Korosec-Serfaty, M., Riedl, R., Senecal, S., & Léger, P.-M. (2022). Attentional and Behavioral Disengagement as Coping Responses to Technostress and Financial Stress: An Experiment Based on Psychophysiological, Perceptual, and Behavioral Data. *Frontiers in Neuroscience*, 16, 883431. https://doi.org/10.3389/fnins.2022.883431

Kreger, A. (2022). *Digital banking user experience: Solve user Pain Points through information architecture*. https://doi.org/10.13140/RG.2.2.34542.08000

Krejtz, K., Duchowski, A., Krejtz, I., Szarkowska, A., & Kopacz, A. (2016). Discerning Ambient/Focal Attention with CoefficientK. *ACM Transactions on Applied Perception*, 13(3), 1–20. https://doi.org/10.1145/2896452

Kucewicz, M. T., Dolezal, J., Kremen, V., Berry, B. M., Miller, L. R., Magee, A. L., Fabian, V., & Worrell, G. A. (2018). Pupil size reflects successful encoding and recall of memory in humans. *Scientific Reports*, 8(1). https://doi.org/10.1038/s41598-018-23197-6

Kwak, D., Ma, X., & Kim, S. (2021). When does social desirability become a problem? Detection and reduction of social desirability bias in information systems research. *Information & Management*, 58(7), 103500. https://doi.org/10.1016/j.im.2021.103500

Lamme, V. A. (2003). Why visual attention and awareness are different. *Trends in Cognitive Sciences*, 7(1), 12–18. https://doi.org/10.1016/s1364-6613(02)00013-x

Law, E. L., Van Schaik, P., & Roto, V. (2013). Attitudes towards user experience (UX) measurement. *International Journal of Human-Computer Studies*, 72(6), 526–541. https://doi.org/10.1016/j.ijhcs.2013.09.006

Léger, P.-M., Senecal, S., Courtemanche, F., Guinea, A., Titah, R., Fredette, M., & L. LeMoyne, E. (2014). Precision is in the Eye of the Beholder: Application of Eye Fixation-Related Potentials to Information Systems Research. *Journal of the Association for Information Systems*, 15. https://doi.org/10.17705/1jais.00376

Lounis, C. A., Hassoumi, A., Lefrancois, O., Peysakhovich, V., & Causse, M. (2020). Detecting ambient/focal visual attention in professional airline pilots with a modified Coefficient K: a full flight simulator study. *ACM Symposium on Eye Tracking Research and Applications*, 10, 1–6. https://doi.org/10.1145/3379157.3391412

Lovallo, W. (2013). Psychophysiology: Theory and methods. In *Springer eBooks*, 1569–1572. https://doi.org/10.1007/978-1-4419-1005-9_484

Maia, C. L. B., & Furtado, E. S. (2019). An Approach to Analyze User's Emotion in HCI Experiments Using Psychophysiological Measures. *IEEE Access*, 7, 36471–36480. IEEE Access. https://doi.org/10.1109/ACCESS.2019.2904977

Minzenberg, M. J., Watrous, A. J., Yoon, J. H., Ursu, S., & Carter, C. S. (2008). Modafinil shifts human locus coeruleus to Low-Tonic, High-Phasic activity during functional MRI. *Science*, 322(5908), 1700–1702. https://doi.org/10.1126/science.1164908

Mithun, M. B., Karmakar, S., Varghese, T., Jaiswal, D., Chatterjee, D., Gavas, R. D., Ramakrishnan, R. K., & Pal, A. (2023). Mind Indriya: A System for Simultaneous Assessment of Cognitive Load, Anxiety and Visual Attention. *2023 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, 3234–3240. https://doi.org/10.1109/SMC53992.2023.10394104

Murphy, P. R., O'Connell, R. G., O'Sullivan, M., Robertson, I. H., & Balsters, J. H. (2014). Pupil diameter covaries with BOLD activity in human locus coeruleus. *Human Brain Mapping*, 35(8), 4140–4154. https://doi.org/10.1002/hbm.22466

Nederhof, A. J. (1985). Methods of coping with social desirability bias: A review. *European Journal of Social Psychology*, 15(3), 263–280. https://doi.org/10.1002/ejsp.2420150303

Parsons, T., Asbee, J., & Courtney, C. (2023). Interaction of Cognitive and Affective Load Within a Virtual City | *2023 IEEE Transactions on Affective Computing*, vol. 14, no. 4, pp. 2768-2775. https://doi.org/10.1109/TAFFC.2022.3220953

Partala, T., & Surakka, V. (2003). Pupil size variation as an indication of affective processing. *International Journal of Human-Computer Studies*, 59(1–2), 185–198. https://doi.org/10.1016/s1071-5819(03)00017-x

Perrig, S. A. C., Aeschbach, L. F., Scharowski, N., von Felten, N., Opwis, K., & Brühlmann, F. (2024). Measurement practices in user experience (UX) research: A systematic quantitative literature review. *Frontiers in Computer Science*, 6. https://doi.org/10.3389/fcomp.2024.1368860

Pirolli, P., & Card, S. (1999). Information foraging. *Psychological Review*, 106(4), 643–675. https://doi.org/10.1037/0033-295x.106.4.643

Pittet, P., & Barthélémy, J. (2015). Experience of formal application ontology development to enhance user understanding in a Geo Business Intelligence SAAS platform. In *Lecture notes in business information processing*, 51–62. https://doi.org/10.1007/978-3-319-21545-7_5

Platzer, D. (2018). Regarding the Pain of Users: Towards a Genealogy of the "Pain Point." *Ethnographic Praxis in Industry Conference Proceedings*, 2018(1), 301–315. https://doi.org/10.1111/1559-8918.2018.01209

Putze, S., Alexandrovsky, D., Putze, F., Höffner, S., Smeddinck, J. D., & Malaka, R. (2020). Breaking The Experience: Effects of Questionnaires in VR User Studies. *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, 1–15. https://doi.org/10.1145/3313831.3376144

Roto, V., & Kaasinen, E. (2008, September). The second international workshop on mobile internet user experience. In *Proceedings of the 10th international conference on Human computer interaction with mobile devices and services*, 571-573. https://doi.org/10.1145/1409240.140935

Russell, J. A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology*, 39(6), 1161–1178. https://doi.org/10.1037/h0077714

Statista. (n.d.). *Software as a Service - Canada | Statista Market forecast.* https://www.statista.com/outlook/tmo/public-cloud/software-as-a-service/canada

Sweller, J., van Merrienboer, J. J. G., & Paas, F. G. W. C. (1998). Cognitive Architecture and Instructional Design. *Educational Psychology Review*, 10(3), 251–296. https://doi.org/10.1023/A:1022193728205

Swoboda, D., Boasen, J., Léger, P., Pourchon, R., & Sénécal, S. (2022). Comparing the Effectiveness of Speech and Physiological Features in Explaining Emotional Responses during Voice User Interface Interactions. *Applied Sciences*, 12(3), 1269. https://doi.org/10.3390/app12031269

Van Acker, B. B., Parmentier, D. D., Vlerick, P., & Saldien, J. (2018). Understanding mental workload: from a clarifying concept analysis toward an implementable framework. *Cognition Technology & Work*, 20(3), 351–365. https://doi.org/10.1007/s10111-018-0481-3

Van Der Wel, P., & Van Steenbergen, H. (2018). Pupil dilation as an index of effort in cognitive control tasks: A review. *Psychonomic Bulletin & Review*, 25(6), 2005–2015. https://doi.org/10.3758/s13423-018-1432-y

Vanneste, P., Raes, A., Morton, J., Bombeke, K., Van Acker, B. B., Larmuseau, C., Depaepe, F., & Van Den Noortgate, W. (2020). Towards measuring cognitive load through multimodal physiological data. *Cognition Technology & Work*, 23(3), 567–585. https://doi.org/10.1007/s10111-020-00641-0

Vignaux, M., Léger, P.-M., Charland, P., Salame, Y., Durand, E., Bouillot, N., Pardoen, M., & Senecal, S. (2021). An Exploratory Study on the Impact of Collective Immersion on Learning and Learning Experience. *Multimodal Technologies and Interaction*, 5, 17. https://doi.org/10.3390/mti5040017

Vrijheid, M., Armstrong, B., Bedard, D., Brown, J., Deltour, I., Iavarone, I., Krewski, D., Lagorio, S., Moore, S., Richardson, L., Giles, G., McBride, M., Parent, M.-E., Siemiatycki, J., & Cardis, E. (2008). Recall bias in the assessment of exposure to mobile phones. *Journal of Exposure Science & Environmental Epidemiology*, 19, 369–381. https://doi.org/10.1038/jes.2008.27

Wals, S. F., & Wichary, S. (2022). Under Pressure: Cognitive Effort During Website-Based Task Performance is Associated with Pupil Size, Visual Exploration, and Users' Intention to Recommend. *International Journal of Human-Computer Interaction*, 39(18), 3504–3515. https://doi.org/10.1080/10447318.2022.2098576

Yang, Y. & Sun, Y.(2017). *Facial expression recognition based on Arousal-Valence Emotion Model and Deep Learning method*. IEEE Conference Publication | IEEE Xplore. https://ieeexplore.ieee.org/document/8789024

Yusuf, M., Silas, F. A., & Haruna, S. (2018). Implementing Personnel Management System as SaaS. *Circulation in Computer Science*, 3(5), 1-6. https://doi.org/10.22632/ccs-2018-252-86

Chapter 3
A Multimodal Approach to Identifying and Predicting Usability Pain Points: An Experimental Study in User Experience Research

Kelvin Jacinto, Sylvain Sénécal, Constantinos Coursaris, Pierre-Majorique Léger,
Shang-Lin Chen, Thaddé Rolon-Mérette, Alexander Karran
HEC Montreal

Abstract

The increasing reliance on self-reported measures in usability assessments has highlighted biases undermining their reliability, particularly in complex digital environments. This study addressed the need for more objective methods by examining biosignals which offers precise measurements of bodily reactions to identify and predict usability challenges in the context of digital interfaces. It explored distinct psychophysiological signatures of users interacting with a digital application and evaluated their reliability in predicting usability issues in business enterprise environments.

The study introduced a multimodal approach using combined psychophysiological measures of emotional arousal, valence, cognitive load and visual attention from 86 participants who performed tasks in three enterprise systems. Pain points were introduced artificially in controlled tasks to elicit psychophysiological responses. Cluster analysis revealed four distinct user profiles reactions to these artificially induced pain points. The study used logistic regression to train predictive models to identify when users encounter usability pain points on a natural task.

The study's key findings included the identification of unique psychophysiological signatures and the moderate predictive success of models using pupil dilation and k-coefficient as significant indicators to usability pain points. Despite individual variability

and moderate precision challenges, these results demonstrated the feasibility of using psychophysiological measures for real-time usability assessment.

This research advanced the understanding of user responses to usability pain points in digital enterprise environments. It underscored the potential for psychophysiological data in real-time usability evaluation and addressed the challenges when using self-reported assessment.

Highlights

- Developed and validated a novel multimodal methodology combining psychophysiological measures (emotional arousal, valence, cognitive load, and visual attention) to detect and predict real-time usability pain points.
- Cluster analysis revealed four unique psychophysiological response profiles to usability pain points, indicating that users exhibit diverse emotional, attentional, and cognitive responses to usability pain points.
- Shifts in pupil dilation and the k-coefficient emerged as predictors of usability pain points.
- Predictive models trained on psychophysiological signatures demonstrated moderate success in detecting spontaneous usability pain points in natural task settings.

Keywords

psychophysiological signatures, usability, pain points, predictive model, cluster analysis

3.1 Introduction

The definition of usability has evolved from Shackel's (2009) early characterization, describing it as "the capability to be used by humans easily and effectively," to the International Organization for Standardization (1998, pg. 22), defining it as "the extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use." These definitions highlight the importance of user-centered design principles in creating functional but also intuitive and pleasant systems (ISO, 1999). One way to improve usability is through usability assessment/testing by identifying user pain points. Most methods used in usability assessment are based on self-reported and observational data (Brunn et al., 2016).

The relationship between usability assessment and user pain points is a critical area of study in human-computer interaction. Usability assessment systematically evaluates how effectively, efficiently, and satisfactorily users can interact with a system (ISO, 1998). Pain points can manifest as obstacles users encounter during interactions, leading to frustration or inefficiency (Platzer, 2018). By pinpointing these pain points, designers and developers can prioritize improvements that enhance the overall user experience (Platzer, 2018; Costa et al., 2016). For instance, usability assessment can reveal common pain points encountered by the users, allowing teams to address these issues directly, thereby improving user satisfaction and performance (Costa et al., 2016). Ibarra-Noriega et al. (2024) work on a mobile health platform for assessing postoperative dental pain illustrated how formal usability evaluations can uncover specific challenges users face, thereby guiding subsequent design improvements (Ibarra-Noriega, 2024).

The concepts of usability assessment and pain points are widely used in UX research to ensure that design solutions are practical and centered around user needs. This user-centered design philosophy emphasizes the importance of understanding user experiences and expectations, which can lead to more effective and engaging digital products (Klotins et al., 2018; Berni et al., 2023).

Usability assessment in UX research predominantly relies on self-reported measures to gauge users' perceptions, emotions, and reactions (Bargas-Avila & Hornbæk, 2011; Inan Nur et al., 2021; Perrig et al., 2024). However, these methods are susceptible to various biases, including social desirability and recall bias (Kwak et al., 2021; Vrijheid et al., 2008; Horwitz et al., 2024), which can distort the accuracy of the data. This problem is particularly pronounced in complex digital environments where moment-by-moment experiences may not be accurately captured through post-task questionnaires or interviews. Therefore, one of the primary challenges is finding reliable and objective ways to assess usability pain points beyond self-reporting.

Psychophysiological signatures hold significant potential in complementing self-reported assessments, which are often susceptible to biases and inaccuracies stemming from subjective interpretation and social desirability effects (Ajenaghughrure et al., 2020; Wager et al., 2013). Psychophysiological signatures are objective biological signals that reflect an individual's emotional, cognitive, and psychological states, derived from physiological responses such as heart rate variability, skin conductance, and brain activity (Chang et al., 2015; Ajenaghughrure et al., 2020). For instance, while self-reports can offer insights into an individual's conscious experience and emotional states, psychophysiological measures provide a more objective perspective, allowing researchers to observe the underlying biological processes that correspond to these experiences (Dirican & Göktürk, 2011; Chang et al., 2015; Visser et al., 2017). While psychophysiological measures may not always precisely capture bodily reactions, they provide a valuable complementary perspective. For instance, Barreto et al. (2007) demonstrated the effectiveness of these measures in detecting stress levels through heart rate variability. Similarly, Maia and Furtado (2019) used galvanic skin response to monitor emotional responses, while Ferreira et al. (2014) assessed cognitive load during interactions using electroencephalography (EEG). However, many of these studies have primarily focused on assessing these aspects independently, which may limit a more comprehensive understanding of the holistic, multimodal responses to usability challenges.

Despite advancements in psychophysiological methods, limited research integrates multiple psychophysiological measures to comprehensively capture a user's response to usability pain points in real time. Furthermore, to our knowledge, no prior research has examined the potential of using various psychophysiological methods to assess different aspects of the user's response – emotional and cognitive response and visual attention behaviour - when users encounter a pain point and develop predictive models that could anticipate pain points in future tasks. Therefore, integrating psychophysiological methods and predictive models of pain points would be a promising avenue for enhancing real-time usability monitoring and system optimization.

This study built upon prior work (Giroux-Huppé et al., 2019) by adopting a multimodal approach to psychophysiological measurement, combining metrics of emotional arousal, valence, cognitive load and visual attention. This study aimed to answer the following:

Research Question #1 (RQ1): *What are the distinct psychophysiological signature patterns exhibited by users when encountering usability issues in a digital interface?*

Research Question #2 (RQ2): *To what extent does these psychophysiological signatures of a user experiencing usability issues reliably identify pain points in other tasks within the same system?*

By capturing these responses simultaneously, the study aimed to identify distinct "psychophysiological signatures" associated with usability pain points. These signatures were then used to train predictive models that could reliably forecast the occurrence of usability pain points in similar tasks, an innovative approach that has not been extensively explored in existing literature. The study also aimed to evaluate the predictive model's extent to reliably identify usability pain points in another similar task.

The study revealed unique psychophysiological response patterns, which classified into distinct user profiles based on the users' reactions to usability pain points. Through cluster analysis, the study identified four unique profiles indicative of variability in users' psychophysiological responses during usability disruptions. Additionally, the trained predictive models, though moderately successful, demonstrated the potential of psychophysiological measures to detect usability issues in real-time, offering a new avenue for continuous usability assessment.

This article starts with a comprehensive introduction to the research topic, including the problem statement, importance, and research gap. A literature analysis examining earlier research using traditional usability assessment and recent psychophysiological approaches in usability assessment comes next. This is followed by a methods section, which discusses how the experiment was conducted. This includes participant recruitment, experimental design, data collection, and analytical approach. Next is the data analysis and results section, highlighting the findings from cluster analysis and model performance evaluation. The discussion section contextualizes the findings within the existing literature and discusses implications, limitations, and recommendations for future research. The article concludes with a summary of the critical insights and contributions to the field of UX research.

3.2 Background

User Experience (UX) research has historically relied heavily on self-reported assessments, such as surveys, interviews, and questionnaires (Bargas-Avila & Hornbæk, 2011; Inan Nur et al.,2021; Perrig et al., 2024), to understand users' perceptions, emotions, and interactions with products and services (Brunn et al., 2016). Studies like those by Bargas-Avila and Hornbæk (2011) and Inan Nur et al. (2021) have demonstrated the dominance of these methods, with over 50% of UX studies employing questionnaires. However, while self-reported assessments are perceived as practical and easy to administer (Law et al., 2013), they are prone to various biases that undermine the insights' accuracy.

3.2.1 Challenges in Self-reported Usability Assessments

One major challenge in self-reported assessments is social desirability bias, where participants provide responses they perceive as socially acceptable rather than truthful reflections of their experiences (Nederhof, 1985). For example, Kwak et al. (2021)

observed this bias in studies on mobile internet addiction, with respondents underreporting behaviours perceived as unfavourable. Similarly, recall bias poses another significant challenge. Users often need help remembering past interactions accurately, leading to distorted data, as evidenced by Vrijheid et al. (2008) in their study on mobile phone usage. Additionally, the peak-end rule further complicated self-report assessments by emphasizing users' most intense and final moments of an experience over the entire interaction (Kahneman et al., 1993), as Horwitz et al. (2024) highlighted in their work on retrospective mental health assessments.

Moreover, the timing of data collection introduces another layer of complexity. Studies reviewed by Bargas-Avila and Hornbæk (2011) and Inan Nur et al. (2021) reveal that most UX evaluations occur post-interaction, potentially omitting critical real-time responses. Virtual reality (VR) studies further underscore this limitation; transitions between immersive and physical environments often disrupt users' presence, leading to skewed feedback (Putze et al., 2020; Alexandrovsky et al., 2020). Together, these challenges highlight the need for more reliable, objective methods in UX research.

3.2.2 Emergence of Psychophysiological Approach to Usability Assessment

Psychophysiological methods have emerged as promising alternatives to address the biases of self-reported assessments. These techniques, which include tracking physiological responses like electrodermal activity (EDA), heart rate variability, pupil dilation, and brain activity, offer non-invasive, real-time insights into users' emotional and cognitive states (Dirican & Göktürk, 2011). While historically underutilized—accounting for only 5% of studies from 2005 to 2009 (Bargas-Avila & Hornbæk, 2011)—their adoption has increased, with 14% of studies between 2000 and 2019 incorporating such measures (Inan Nur et al., 2021). This growth reflects a broader recognition of their potential, as confirmed by a systematic review by Apraiz Iriarte et al. (2021), which highlighted a surge in publications employing psychophysiological methods post-2016.

Research has demonstrated the effectiveness of psychophysiological measures in assessing emotional responses, cognitive processes, and attention. For instance, studies have used the arousal-valence model (Russell, 1980) to map emotions based on psychophysiological data. Alexandros and Michalis (2013) used psychophysiological measures to assess emotional transitions during interactions. Maia and Furtado (2019) found significant correlations between emotional dimensions and biosignals like EEG, EDA, and heart rate during pleasure-driven tasks. Vignaux et al. (2021) showed that EDA and ECG effectively measured heightened emotional engagement in immersive, collective learning environments. Similarly, Swoboda et al. (2022) highlighted the role of speech and physiological measures in detecting emotional responses during voice interface interactions.

Meanwhile, advancements in cognitive load assessment have shown that measures like EDA and pupil dilation can provide nuanced insights into task complexity and user engagement (Vanneste et al., 2020). Cognitive load theory (Sweller et al., 1998) highlighted the limitations of working memory when faced with excessive information or complexity. Research by Ferreira et al. (2014) demonstrated the feasibility of real-time cognitive load assessment using non-invasive sensors. Vanneste et al. (2020) showed that multimodal measures, including EDA, EEG, and eye tracking, provide nuanced insights into cognitive load by capturing both arousal and mental components. Hudon et al. (2021) utilized pupillary dilation, specifically task-evoked pupillary response (TEPR), as a proxy for cognitive effort, highlighting the link between pupil size and task demands (Kucewicz et al., 2018; Van Der Wel & Van Steenbergen, 2018). The Adaptive Gain Theory (Aston-Jones & Cohen, 2005) explains how the locus coeruleus-norepinephrine (LC-NE) system regulates task engagement, with pupil size fluctuations reflecting its phasic (task-focused) or tonic (disengaged) modes (Gilzenrat et al., 2010). These findings underscore psychophysiological measures' value in capturing cognitive load and task engagement dynamics.

Visual attention, another aspect of user behaviour, involves selectively focusing on relevant stimuli while filtering out distractions (Lamme, 2003). Rooted in Pirolli and Card's (1999) information foraging theory, users maximize information gain by adapting their strategies or environment. This process is essential in computer-based tasks, where users must prioritize key display elements (Wals & Wichary, 2022). Krejtz et al. (2016) introduced non-invasive measures to distinguish ambient and focal attention modes, which Lounis et al. (2020) applied to assess pilots' visual attention across flight phases using eye-tracking technology. Similarly, Carmichael et al. (2022) demonstrated that eye-tracking could measure the effectiveness of information disclosure nudges in chatbot interactions, validating their influence on user behaviour. These studies highlighted the role of psychophysiological methods in advancing our understanding of visual attention in diverse user experiences.

3.2.3 Application of Psychophysiological Methods to Usability Pain Point Detection

Integrating psychophysiological methods into UX research has revolutionized the detection of usability pain points—user frustrations or challenges encountered during interactions (Platzer, 2018). Giroux-Huppé et al. (2019) distinguished between explicit pain points and implicit pain points. Where explicit pain points are consciously acknowledged negative emotions reported during or after a task, and implicit pain points, also termed *psychophysiological pain points*, are automatic physiological responses characterized by high emotional arousal and negative emotional valence in reaction to an event during the interaction (Giroux-Huppé et al., 2019). Their work demonstrated that psychophysiological measures, such as heightened arousal and negative valence, offer a real-time, objective approach to capturing user frustration, surpassing the limitations of traditional self-reported methods.

3.2.4 The proposed novel approach in identifying pain points

Although Giroux-Huppé et al. (2019) research introduced a novel approach to identifying pain points accurately in real-time, the approach concentrated on one aspect of the user's response: the emotion the users felt. Prior and recent UX research increasingly adopts multimodal approaches to measure multiple aspects of user’s response simultaneously. Léger et al. (2014) combined eye tracking and EEG to improve the temporal precision of attentional, cognitive, and motor process measurements. Charland et al. (2015) used psychophysiological tools to differentiate engagement levels across behavioural, cognitive, and emotional dimensions during learning tasks. Korosec-Serfaty et al. (2022) examined the impact of technostress and financial stress on users' emotional and cognitive responses by integrating psychophysiological, perceptual, and behavioural data. Parsons et al. (2023) assessed users' cognitive and emotional states in virtual environments in real time, while Mithun et al. (2023) introduced the *Mind Indriya* system to measure cognitive load, anxiety, and visual attention simultaneously. These studies highlighted the growing emphasis on multimodal methods to capture a comprehensive picture of user responses.

To expand on Giroux-Huppé's work, this study applied multimodal approach, using various psychophysiological methods to assess the different aspects of the user's response, which are characterized by changes in emotional arousal, valence, visual attention, and cognitive load when users encounter a pain point. This aimed to provide a more accurate representation of pain points beyond isolated emotional data. This proposed that when users encounter a pain point, there will be a distinct psychophysiological signature such as by changes in emotional arousal, valence, visual attention, and cognitive load.

Additionally, by using the captured psychophysiological signatures, this study aimed to develop predictive models and to examine the reliability of these signatures in identifying pain points across similar tasks. This proposed that the psychophysiological signature associated with a pain point experienced during a task can be used to identify pain points on other similar tasks reliably.

For this research study, the term "*usability pain point (UPP)*" referred to the usability challenges experienced by users during their interaction with a digital interface, which gives rise to automatic psychophysiological responses characterized by abnormal changes in emotional arousal, valence, visual attention, and cognitive load.

3.3 Methods

This research builds on the novel approach on identifying pain points introduced by Giroux-Huppé et al. (2019) by assessing multi-aspect of user’s psychophysiological responses. The aim of this experiment is to evoke psychophysiological responses in users when they encounter a pain point to be able to identify distinct psychophysiological signatures characterized by changes in emotional arousal, valence, visual attention and cognitive load, providing a foundation for developing predictive models to identify pain points on other similar tasks reliably.

3.3.1 *Study Design*

The study design comprises three phases: Phase I employs a mixed-subject design to collect psychophysiological data by artificially manipulating the occurrence of pain points in the tasks that the users are to perform, Phase II focuses on using the data collected on Phase I to train a predictive model to identify pain points, and Phase III evaluates the predictive model’s performance and its reliability to identify usability pain points on a natural task. The proposed process framework is detailed in **Figure 1**.

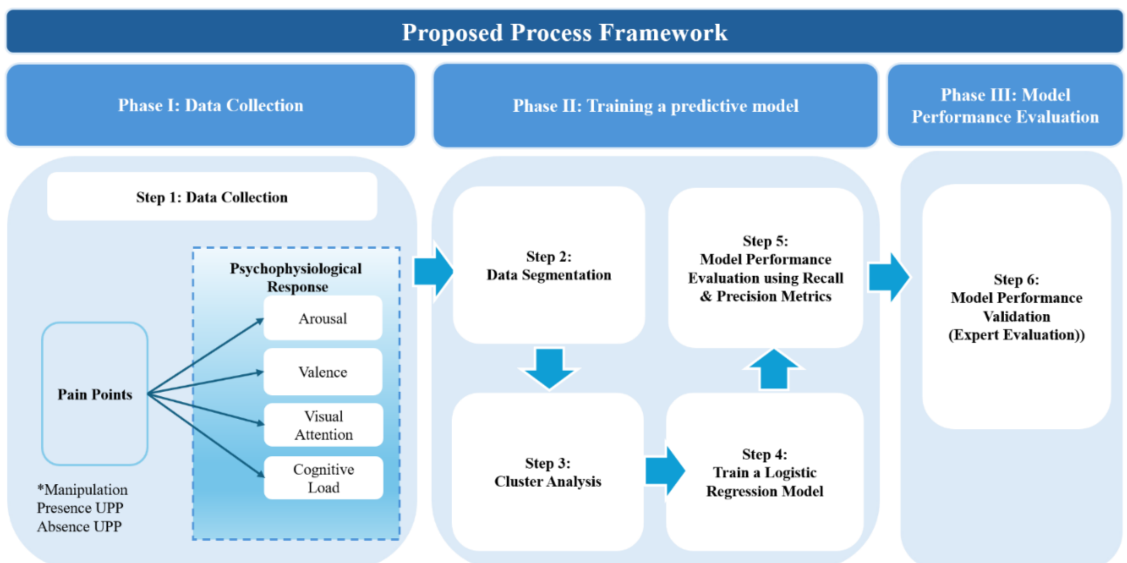

Figure 1. *The proposed process framework*

3.3.2 *Participants*

Eighty-six (86) participants were recruited via our institution's panel to participate in the study. Recruited participants had an advanced level of French, did not have skin allergies or sensitivity, had no astigmatism, did not suffer from epilepsy and had no current or prior experience working with the following SaaS enterprise systems: ServiceNow, Microsoft Dynamics 365 CRM, and Salesforce Cloud Service. All participants provided signed consent in line with the HEC Montreal research ethics committee [Certificate No.: 2024-5933]. Each participant received a compensation of $20.

The 41 participants were randomly assigned to Condition A, 24 men and 17 women, ranging from 18 to 45 years old ($M=26.5$; $SD = 6.10$). Forty-five (45) participants were randomly assigned to Condition B, 24 men and 21 women, ranging from 18 to 59 years old ($M=27.07$; $SD=8.97$). A t-test for age revealed no statistically significant difference between the mean ages of Condition A and Condition B ($t(84) = -0.332$, $p = .741$). Additionally, a chi-square test for gender indicated no significant association between gender and condition ($\chi^2(1, N = 86) = 0.072$, $p = .789$). Other details on Condition A and B were elaborated in section “***3.3.4 Study Conditions and Tasks***”.

3.3.3 Stimuli

The SaaS software delivery model is becoming popular for enterprises of all types and sizes (Haselmann & Vossen, 2011). Software as a Service (SaaS) is a cloud service where consumers can access software applications over the internet or "the cloud" (Yusuf et al., 2018). According to Gartner, the shift to SaaS fundamentally changes the organization's ownership models from owning a software license to paying a third party for software usage. In 2022, SaaS spending constituted an average of 11% of total IT spending, up from 5% in 2018 (*Gartner: Fueling the Future of Business*). Canada's Software as a Service (SaaS) market is experiencing significant growth due to trends in collaboration-focused markets, CRM and HRMS software adoption, customer preferences for cloud-based solutions, government support for digital transformation and a diverse business landscape such as remote work (Statista). For this research study, three SaaS platforms were selected to be the experimental stimuli: Microsoft Dynamics 365 CRM, Salesforce, and ServiceNow.

3.3.3.1 Microsoft Dynamics 365 CRM. According to Statista's Market Insights Financial Statements of Key Players (updated March 2024), one of the vital players in the SaaS market in Canada in 2022 is Microsoft Cloud, making up 17% of the market. Microsoft Dynamics 365 CRM is a comprehensive, integrated system designed to streamline various business processes related to customer management (Microsoft). The trial version of the customer service instance offered by Microsoft was used for this experiment stimuli. **Figure 2** presents the homepage the participants see before performing the task in Microsoft Dynamics 365 – CRM.

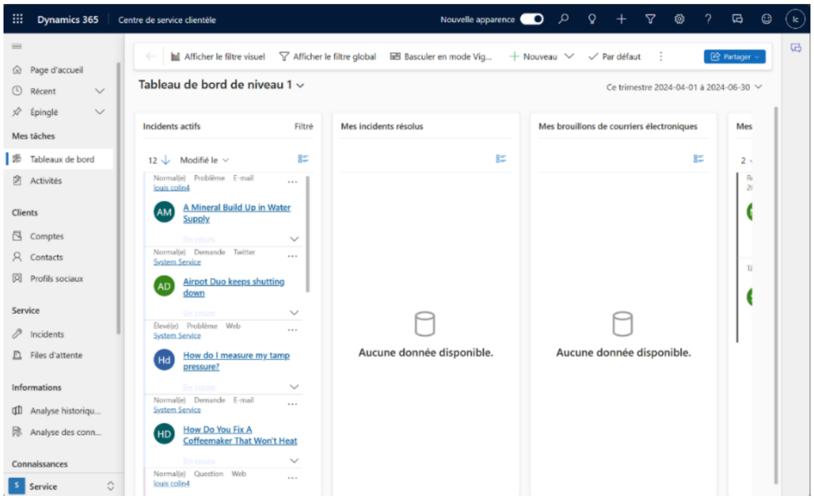

Figure 2. *Microsoft Dynamics 365 – CRM Homepage*

3.3.3.2 Salesforce. Following Microsoft's lead is Salesforce, comprising 13% of the SaaS market in Canada in 2022 (Statista). Salesforce is a leading cloud-based CRM platform that provides various applications and services to assist business operations, explicitly managing customer relationships and interactions (Salesforce). The institution's Salesforce account was used for this stimulus. **Figure 3** presents the homepage that the participants see before performing the task in Salesforce.

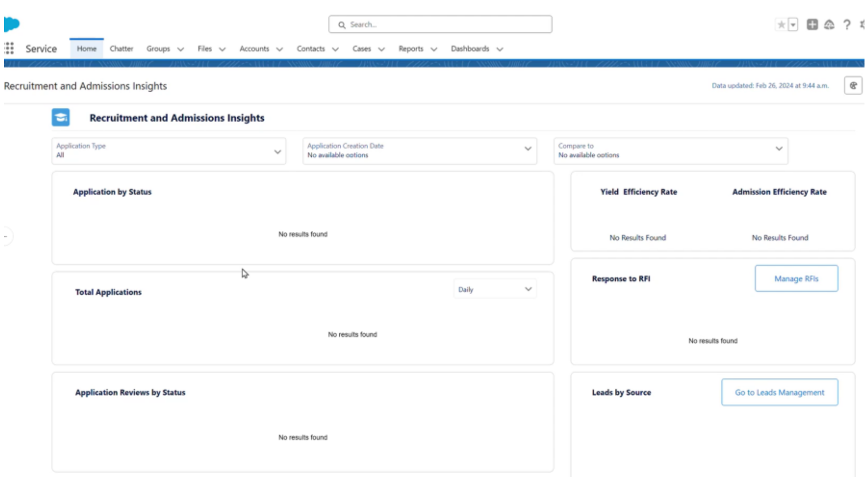

Figure 3. *Salesforce Homepage*

3.3.3.3 ServiceNow. Another vital player in the SaaS market is ServiceNow, comprising 4% of the SaaS market in Canada in 2022 (Statista). ServiceNow is a cloud-based platform that offers a group of applications designed to help streamline and automate business processes, particularly in IT Service Management (ITSM), IT Operations Management (ITOM), and IT Business Management (ITBM) (ServiceNow). The sponsoring client provided a developer version of the enterprise system. **Figure 4** presents the homepage that the participants see before performing the task in ServiceNow.

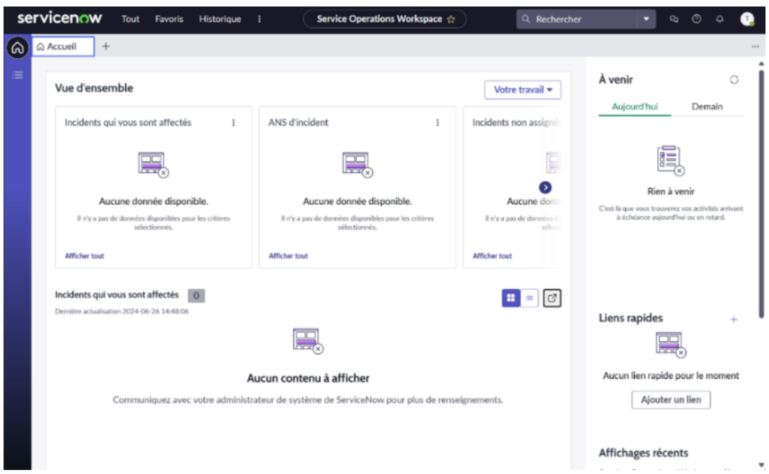

Figure 4. *ServiceNow Homepage*

3.3.4 Study Conditions and Tasks

3.3.4.1 Study Conditions. Participants were randomly assigned to Condition A or B. The two conditions entailed the same three tasks and merely varied in the sequence of the systems being used to perform those tasks. For Condition A, participants performed Task 1 and 2 using Microsoft Dynamics and Task 3 using ServiceNow. For Condition B, participants performed Task 1 and 2 using Salesforce and Task 3 using ServiceNow. The goal of Phase I is to collect psychophysiological data by artificially manipulating the occurrence of pain points in either Task 1 or Task 2. Assigning each participant to either Condition A or B helped generalize the psychophysiological data collected from different systems during Task 1 and Task 2. Also, randomizing the manipulated pain point to either Task 1 or 2 minimized any potential bias that could arise from participants anticipating or adapting to the pain point in a specific task.

3.3.4.2 Task 1: Create a Contact. Participants were given the task of creating a new contact file for a new client. **Figure 5** explains the step-by-step process that the user will perform to accomplish the task. The participants will begin on the assigned stimulus homepage. They need to locate and click the "Contacts" tab, which will direct them to the contacts form page, where they will fill in the information required for the task. Once done, the participants must click the "Save" or "Submit" button.

Figure 5. *Process flow for completing Task 1*

3.3.4.3 Task 2: Create a Case. Participants were tasked with creating a new case file for a client. **Figure 6** explains the step-by-step process that the user will perform to accomplish the task. It is similar to Task 1, but instead of clicking the "Contact" tab, participants will have to click the "Case" or "Incidents" tab, which will direct them to the case/incident page form. Once the required information is completed, the participants must click the "Save" or "Submit" button.

Figure 6. *Process flow for completing Task 2 and Task 3*

3.3.4.4 Task 3: Create a Case. The same steps were used for Task 3 (*see **Figure 6***) when creating a case. However, participants will be exposed to ServiceNow when performing Task 3.

In summary, the use of Conditions A and B and randomly assigning the manipulated pain points to either Task 1 or Task 2 enhanced the generalizability of the psychophysiological data used in Phase II, which focused on training a predictive model to identify pain points. Moreover, Task 3 was the same for Conditions A and B, and no artificial pain points were induced. Task 3 data were used in Phase III to evaluate the predictive model's performance and reliability to identify usability pain points on a natural task.

3.3.5 Stimuli Manipulation

Configuration errors were randomly induced in either Task 1 or Task 2 in MS Dynamics 365 CRM and Salesforce to manipulate the presence of pain points. The manipulated errors were placed on the Contacts tab for Task 1 and the Cases/Incidents tab for Task 2. Having a similar number of clicks (*see **Figure 5** and **Figure 6** in the previous section*) before reaching the manipulated section ensured that participants had consistent experience across tasks, which helped in isolating the effect of the manipulated pain point rather than the differences in task complexity. Also, providing an onboarding video that showed the steps to accomplish the tasks standardized the participants' approach, which reduced variability in how participants performed the tasks, ensuring that any psychophysiological responses were due to the pain point and not differences in task understanding or execution. Placing the pain point at a specific interaction (*clicking the contact or case tab*) allowed the study to precisely measure the psychophysiological responses to that interaction by extracting the psychophysiological data encompassing 10-second intervals centered on the manipulated pain point.

Once the participants clicked the Contacts tab or the Case/Incidents tab where the configuration errors were planted, an error message will appear (*see **Figure 7** & **Figure 8***). Although these manipulated errors are aimed at eliciting psychophysiological responses, the outcomes may vary among participants.

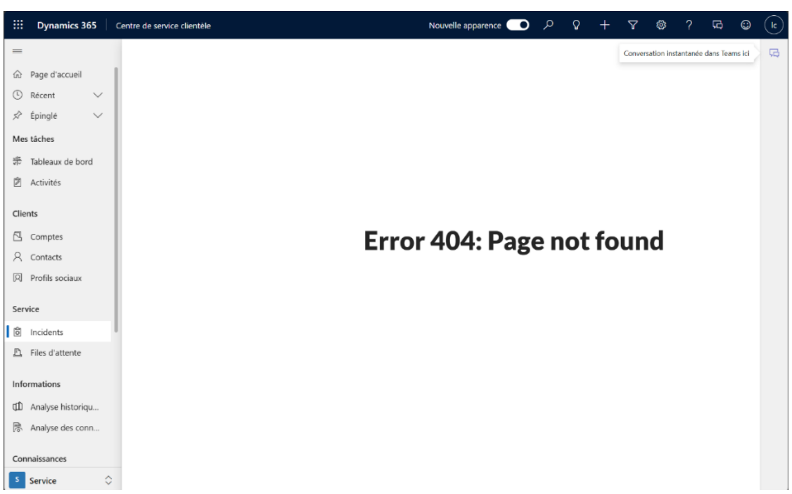

Figure 7. *Error message in Microsoft Dynamics 365*

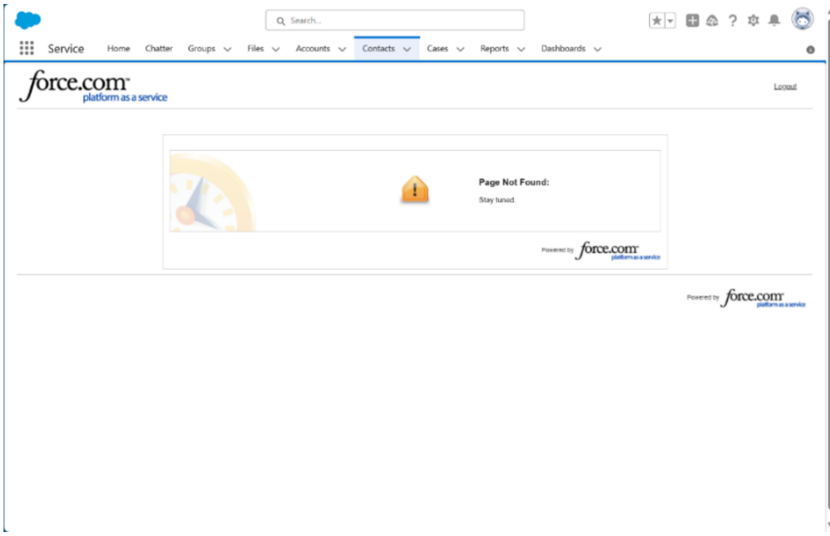

Figure 8. *Error message in Salesforce*

There will be no manipulated error in Task 3. Task 3 will serve as the natural task to test the reliability of the predictive models in identifying usability pain points in Phase III (refer to **Figure 1**).

3.3.6 Measures

This study combined eye tracking, EDA, and automated facial expression recognition data to measure the psychophysiological response patterns exhibited by users when encountering a pain point.

3.3.6.1 Emotional Arousal. To measure emotional arousal, Electrodermal activity (EDA) was used. EDA is a physiological measure relevant to emotion research and is commonly used to indicate physiological arousal (Braithwaite et al., 2013). By capturing skin conductance, which refers to the electrical properties of the skin changes in response to sweat secretion, EDA provides a moment-by-moment measure of arousal (Boucsein, 2012). Arousal is the "state of being physiologically alert, awake, and attentive, associated with sensory stimulation and activation of fibres from the reticular activating system" (Beri & K, 2019). A study by Caruella et al., (2019) reviews the use of EDA to assess consumer emotions in marketing and consumer research, where the review finds that EDA is a valuable tool for understanding consumer emotions, though the paper also recommends combining EDA with other methods to capture both arousal and valence.

3.3.6.2 Valence. The study utilized a Facial Expression Recognition (FER) software to measure the participant's facial expressions during the experience. The hedonic tone of the feeling, referred to as valence (Garbas et al., 2013), is measured through facial expression recognition and has been extensively studied in HCI (Yang & Sun, 2017). According to Mehrabian's (2017) criterion for emotional expression, humans express their emotions through facial expressions by 55%, 7% by language, and 38% by voice. Facial Expression Recognition (FER) advancement was supported by Li and Deng's (2022) comprehensive review of deep learning techniques applied to FER, which discussed the transition of FER from controlled laboratory settings to more challenging real-world environments and emphasizing the increasing use of deep neural networks to address environmental complexities.

Facial expression and electrodermal activity effectively measure the two dimensions of emotion and comprehensively describe a person's emotional change (Russell, 1980; Yang & Sun, 2017).

3.3.6.3 Visual Attention. To measure visual attention, the study used eye tracking and a visual search measure, k-coefficient. Eye-tracking (or oculography) is a research method that uses an eye-tracking device to track the point of gaze or a user's eye movement during task execution (Borys et al., 2017). A visual search measure has been proposed by Krejtz et al. (2016) to characterize two modes of attention – ambient vs. focal attention. Coefficient K (k-coefficient) measures visual behaviour fluctuating between focal and ambient viewing modes, combining the length of saccades with the duration of fixations (Krejtz et al., 2016).

3.3.6.4 Cognitive Load. Another psychophysiological measure captured by an eye-tracking device is pupil dilation. Pupil dilation measures pupil size changes, providing insights into attention, arousal, emotion, and mental workload (Bergstrom et al., 2014). Pupil size changes in response to emotionally arousing stimuli (Bradley et al., 2008). A study by Partala and Surakka (2003) showed that pupil size was significantly larger during emotionally negative and positive stimuli than during neutral stimuli. Pupil diameter is valuable for real-time cognitive assessment, distinguishing task difficulty with larger pupil dilations under challenging tasks (Kreijtz et al., 2018). Biometrics quantify task-related cognitive effort and provide temporally specific and non-intrusive measurements of cognitive dynamics throughout a task (Wals & Wichary, 2022). This is further supported by Gilzentrat et al. (2010) findings that pupil diameter can serve as a proxy to the locus coeruleus-norepinephrine (LC-NE) system, which has a vital function in regulating cognitive control and provides evidence that the LC-NE system's regulation of cognitive control can be monitored non-invasively through pupillometry.

3.3.7 Instrumentation

The study used non-intrusive tools to capture the user's psychophysiological responses.

3.3.7.1 Tobii Pro Eye Tracker & Tobii Pro Lab software. Tobii Pro eye tracker was used to capture the pupil dilation and gaze entropy of the participant while performing the task, and the psychophysiological responses were recorded by the Tobii Pro Lab software version 1.217 (Tobii AB, Danderyd, Sweden).

3.3.7.2 Face Reader v.9. Face Reader (version 9) was used to analyze facial expressions and emotions felt by the participant while performing the task. Facial expression analysis with FaceReader can recognize several specific properties in facial images, including the six universal expressions that infer emotional valence (Noldus FaceReader, Wageningen, The Netherlands).

3.3.7.3 Media Recorder. In conjunction with the Tobii Pro eye-tracker and FaceReader, the Media Recorder is a software application used to capture and synchronize various types of data streams, which includes video, audio and biometric data to provide a comprehensive view of the participant's interactions and responses (Noldus Media Recorder, Wageningen, The Netherlands).

3.3.7.4 Biopac MP-150 & Acqknowledge software. Using the EDA sensors securely placed on the participant's non-dominant hand, mostly the participant's left hand, the EDA signal responses were relayed to the Biopac MP-150. BIOPAC MP-150 is a data acquisition system which provides intuitive analysis and visualization tools for capturing physiological signals, specifically EDA, which is the measure to capture the participant's arousal level experienced while performing the tasks (BIOPAC Systems, Inc.). The BIOPAC MP-150 works with Acqknowledge software (BIOPAC Systems, Inc.).

3.3.7.5 Observer XT. Observer XT is a software tool designed to collect, analyze, and present observational data from various sources (Tobii Pro Eye-tracker, FaceReader, Media Recorder, and Biopac MP-150) (Noldus Observer XT, Wageningen, The Netherlands).

3.3.8 Laboratory Setup

3.3.8.1 Experimental Room Setup. As shown in **Figure 9**, a dedicated computer workstation was provided for participant task completion with a microphone that facilitated clear communication with the moderator and a webcam that captured and recorded facial expressions. A Tobii Pro eye-tracker was positioned beneath the screen to monitor eye movements, and an iPad was used to display the consent form, compensation details, onboarding videos, and essential task instructions.

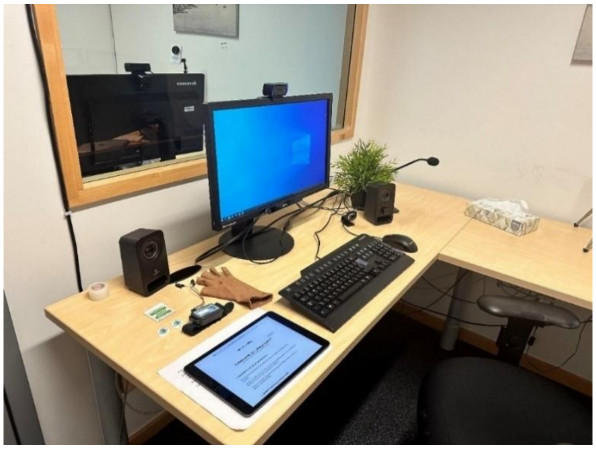

Figure 9. *Lab setup in the experimental room*

3.3.8.2 Observation Room setup. The moderator's workstation shown in **Figure 10**, which is composed of three system units and five monitors that display the Acknowledge Biopac software (1), Observer XT (2), Media Recorder (3), mirrored participant's screen (4), and Tobii Pro (5).

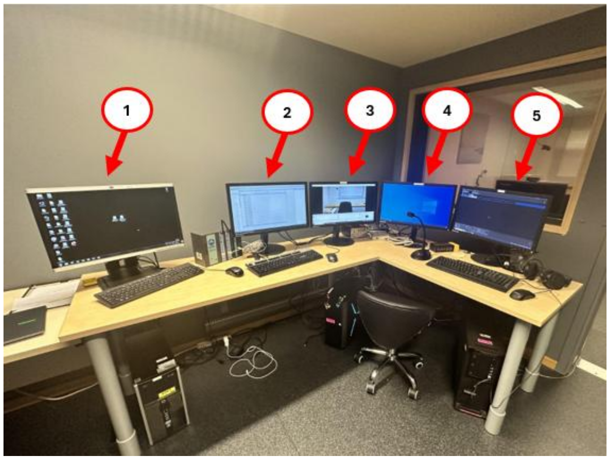

Figure 10. *Lab setup in the observation room*

3.3.8.3 Instrument and Equipment Synchronization setup. Presented in **Figure 11** is the instrument and equipment synchronization setup designed to capture physiological and behavioural response data simultaneously. The SyncBox (Noldus) synchronized data from the EDA, eye-tracking, and facial expression systems, ensuring temporal alignment across all collected measurements.

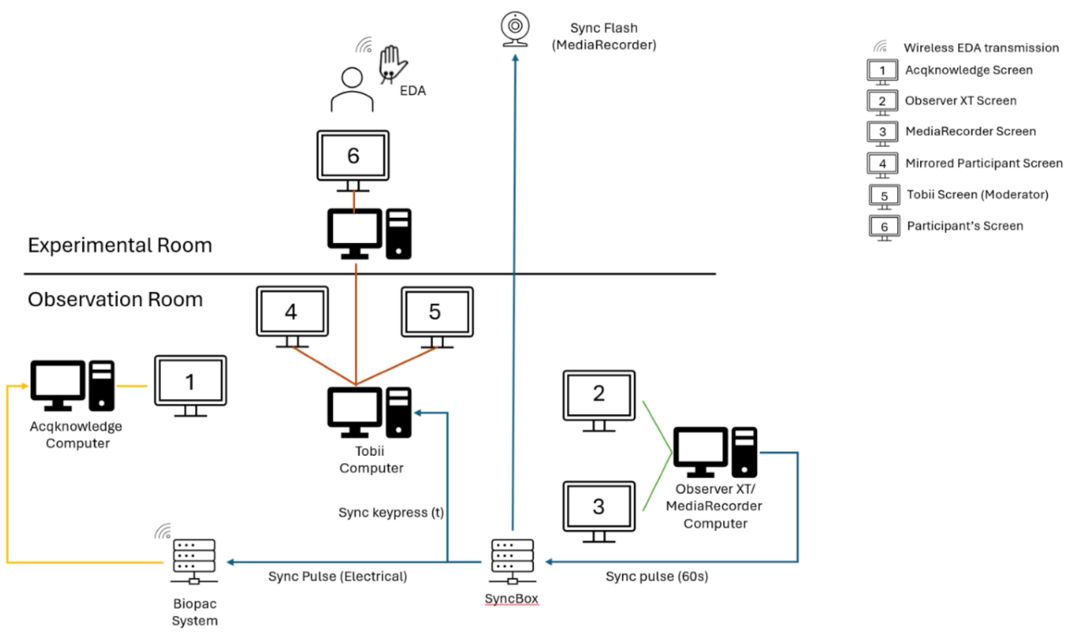

Figure 11. *Instrument and Equipment Synchronization Setup*

3.3.9 Experimental Procedure

Following the proposed process framework presented in the earlier section of the study (shown in **Figure 1**), details of each step are elaborated below.

3.3.9.1 Phase I: Collection of psychophysiological data

Step 1: Data Collection. As shown in **Figure 12**, the study experience begins with welcoming the participants and providing them the consent form. The participants were recruited based on the inclusion and exclusion criteria stated in section ***3.3.2 Participants***. Following the signing of the consent form, participants provided socio-demographic information, EDA sensors were then attached to the non-dominant hand, and an eye calibration test was conducted.

Three tasks were administered. To ensure that the participants responses and interactions with the SaaS enterprise systems were unbiased and uninfluenced by previous experience, participants who have current and prior experience with ServiceNow, Microsoft Dynamics 365 CRM, and Salesforce Cloud Service were not recruited for this study. Prior to task commencement, participants viewed an onboarding video on the iPad, familiarizing themselves with the assigned system (either MS Dynamics or Salesforce for Tasks 1 and 2 and ServiceNow for Task 3) (see Appendix B). Task instructions (see Appendix A) were displayed on the monitor, while relevant information was provided on the iPad (see Appendix C). The SaaS system instances were presented to the participants' screens, where they would perform the tasks. Upon completing all three tasks, participants received a compensation form and expressed gratitude for their involvement.

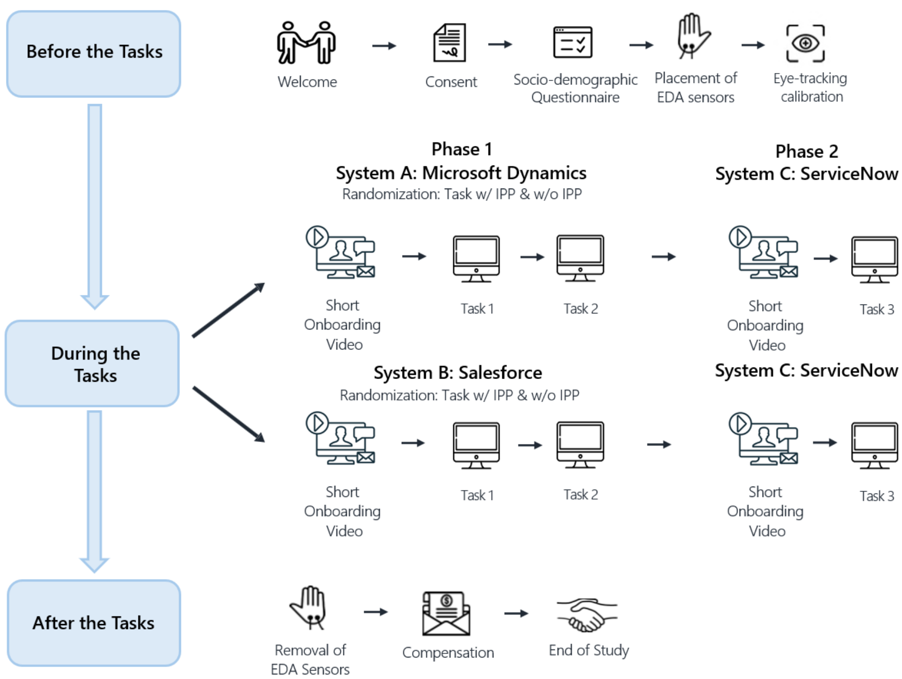

Figure 12. *Step-by-step procedure experienced by the participant during the study*

3.3.9.2 Phase II: Training a predictive model

Step 2: Data Segmentation. Following the experiment and data collection, UPP markers were integrated into the physiological data to correspond with identified usability issues. Data segments surrounding UPP events were extracted, encompassing 10-second intervals centered on the UPP with a 1-second offset. This temporal window captures immediate physiological changes in response to the UPP. For instance, a separate study done by Brown et al. (2011) and Kim et al. (2013) used a brief pre- and post-stimulus window to accurately evaluate immediate physiological changes which correlates to pain such as changes in heart rate, skin conductance, and other autonomic responses. Additionally, the study by Brown et al. (2011) utilized a 20-second baseline period before each stimulus to establish a reference point for measuring changes in brain activity, ensuring that the observed responses were directly attributable to the stimuli. For the purpose of this study, 10-second intervals before and after the UPP marker was chosen based on the intuition that participants' reactions reach their peak within a few seconds and remain strong for a few seconds. The 10-second interval after the stimulus captures this peak reaction, while the 10-second interval before the stimulus serves as a comparison to accurately assess the changes induced by the usability pain point.

Step 3: Data Preparation for Cluster Analysis. To prepare the data for cluster analysis, the increase in physiological measures (x_aug) following a UPP was calculated by subtracting the baseline value (x_before) from the post-UPP value (x_after).

Hence, the study used the formula x_aug = x_after – x_before for cluster analysis, where x = variable.

Each participant's mean values (x_aug) were calculated and subsequently employed as input for cluster analysis using the K-Means algorithm. K-means clustering was utilized to identify groups of participants exhibiting similar physiological responses. The resulting clusters were then categorized into strong and weak reaction groups.

Step 4: Training a Logistic Regression Model. After identifying the clusters, the subsequent step was to train a logistic regression model by calculating the cumulative averages of the psychophysiological variables (x_ca), which act as baseline "norm" values for each participant. The data was then centered by subtracting these cumulative averages (x-ca) from the raw data (x), resulting in x_b. This centered data captured individual deviations from the "norm."

Hence, the study used the formula x_b = x – x_ca to train the models, where x = variable.

Logistic regression models with random intercepts were subsequently trained using the centered physiological data (x_b) to detect the likelihood of a pain point. Separate models were developed for clusters exhibiting strong reactions to identify significant predictors. The dependent variable (DV) was the time before versus after the UPP event.

Step 5: Model Performance Evaluation tested on Task 1 and 2. The trained models were subsequently applied to identify pain points within the Task 1 and Task 2 data, and the performance was evaluated using recall and precision to assess the detection of UPPs.

3.3.9.3 Phase III: Model performance evaluation

Step 6: Model Performance Evaluation tested on Task 3. In the final step, the trained model was applied to Task 3 data to identify pain points. Expert evaluations were conducted by analyzing video recordings of participant interactions. Each predicted pain point was rated on a 0-3 scale, with 0 indicating "absolutely not a pain point" and 3 indicating "absolutely a pain point."

3.4 Data Analysis and Results

This section presents the analyses conducted to examine the psychophysiological responses to usability pain points and assess the predictive model's reliability in detecting pain points in a natural task.

3.4.1 Manipulation check

The manipulation check aimed to verify whether the induced configuration errors in Task 1 and 2 successfully elicited distinct psychophysiological responses in participants characterized by emotional arousal, valence, visual attention and cognitive load changes. Results indicated that 61 out of 86 participants showed significant changes in these indicators following a usability pain point exposure, confirming that the manipulations effectively induced psychophysiological responses characteristic of encountering usability obstacles.

Thus, a final sample of 61 participants with sufficient psychophysiological data was used for cluster analysis. By focusing on the psychophysiological data gathered from these 61 participants, the study ensured that the cluster analysis was based on reliable and meaningful data, which is crucial in Phase II in training a predictive model in identifying distinct patterns in psychophysiological responses when users encounter a usability pain point.

3.4.2 Cluster Analysis Results

K-means clustering was performed to identify groups of participants based on their psychophysiological responses to usability pain points. The purpose of this analysis was to identify patterns in user reactions. As a result, four distinct clusters were identified. **Table 2** presents the mean (*M*) increases in four psychophysiological measures—valence, EDA phasic, pupil dilation, and k-coefficient—following a usability pain point.

Table 2

Mean Increase in Physiological Measures Following a UPP per Cluster

Cluster	**Valence**	**EDA phasic**	**Pupil Dilation**	**K-coefficient**
	M	*M*	*M*	*M*
Cluster 1 (N=25)	0.080	0.372	0.722	-0.393
Cluster 2 (N=25)	-0.558	0.271	-0.728	-0.022
Cluster 3 (N=6)	-0.022	-2.772	0.458	0.549
Cluster 4 (N=5)	2.413	-0.131	-0.519	1.414

Note. Total N=61; M = Mean

Each cluster showed varied patterns of psychophysiological reactions to the usability pain point. For example, in Cluster 1, participants responded with a moderate increase in arousal (EDA) (0.372), a significant increase in pupil dilation (0.722) and a moderate decrease in k-coefficient (-0.393). In contrast, participants in Cluster 2 showed significant decrease in valence (-0.558) and pupil dilation (-0.728) when encountering the usability pain point. Cluster 3 participants demonstrated a significant decrease in arousal (-2.772), along with a moderate increase in pupil dilation (0.458) and k-coefficient (0.549). Meanwhile, participants in Cluster 4 showed a significant increase in valence (2.413) and k-coefficient (1.414), accompanied by a decrease in arousal (-0.131) and pupil dilation (-0.519).

Notably, most participants displayed responses aligned with the patterns found in Clusters 1 (N= 25) and 2 (N= 25). The similarity in response patterns across these participants indicates that the majority experienced the UPP as significant disruptions. Conversely, only a small subset of participants exhibited different response patterns found in Clusters 3 (N= 6) and 4 (N= 5), potentially indicating participants' differences in sensitivity to usability disruptions or varied task engagement levels. Clusters 3 and 4 participants might represent users who are less affected by minor interface issues or who adopt different coping mechanisms, such as maintaining focus without exhibiting high arousal.

Overall, the different response patterns support the concept of a "psychophysiological signature" when faced with a usability pain point, triggering distinct psychophysiological responses consistent with shifts in emotional arousal, valence, visual attention and cognitive load. However, the prominence of patterns found in Clusters 1 and 2 among participants underscores the robustness of psychophysiological responses as indicators of usability pain points. The concentration of participants within these clusters strengthens the proposition that a predictable and recognizable response signature exists to usability issues, validating the use of these indicators in training the predictive model.

3.4.3 Logistic Regression Model Training Results

The following analysis step involved training logistic regression models to identify the likelihood of a pain point occurrence based on the observed psychophysiological responses. Separate models were created for strong-reaction clusters (Clusters 1 and 2) to capture the primary predictors for each cluster. Participants without EDA phasic data were excluded from the model training for this analysis. Thus, a final sample of 45 participants with complete psychophysiological data was used for logistic regression. As discussed in the Procedure section, Step 4 of this study, the centered data (x_b) for variables valence, pupil dilation, EDA phasic, and k-coefficient were used for this analysis, where centered data for valence represents shifts in emotional valence, centered data for pupil dilation represents shifts in cognitive load, centered data for EDA phasic represents changes in emotional arousal, and centered data for k-coefficient represents shifts in visual attention.

3.4.3.1 Predictive Model A (PM-A). Predictive Model A (PM-A) used the centered data (x_b) for variables valence, pupil dilation, EDA phasic, and k-coefficient from Cluster 1 participants ($N=23$). The results indicated that the centered data for the k-coefficient (Estimate = -0.659, SE = 0.244, $t(335) = -2.700$, $p = .007$) was significant. This suggest that changes in visual scanning behaviour are a strong predictor of usability pain points for participants in this cluster. Other predictors, including centered data for valence (Estimate = 0.029, SE = 1.294, $t(335) = 0.020$, $p = .981$), centered data for pupil dilation (Estimate = -0.542, SE = 0.767, $t(335) = -0.710$, $p = .481$), and centered data for EDA phasic (Estimate = 0.135, SE = 0.254, $t(335) = 0.530$, $p = .595$), were not statistically significant.

Given the lack of significance for centered data for valence, an alternative model excluding this variable was tested. The results for the refined model remained similar, with centered data for k-coefficient continuing to be a significant predictor (Estimate = -0.666, SE = 0.237, t(335) = -2.810; p = .005).

Hence, the final predictive model A (PM-A) based on Cluster 1 centered data:

- **PM-A:** *prob_UPP= 1 / (1 + exp(-1 * (0.3252 - 0.572 * pupil_b + 0.1214 * phasic_b - 0.6664 * k_b)))*

Where prob_UPP = probability of usability pain point, x = variable (i.e. valence), and x_b = centered data of the variable x

3.4.3.2 Predictive Model B (PM-B). Another model, Predictive Model B (PM-B), was trained for detecting usability pain point using the centered data from Cluster 2 (N=22). In this model, the centered data for pupil dilation (Estimate = -25.291, SE = 2.769, t(345) = -9.130, p<0.001) emerged as a highly significant predictor of usability pain point, underscoring the shift in cognitive workload. Other variables, including the centered data for valence (Estimate = -3.655, SE = 2.832, t(345) = -1.290, p = .198), the centered data for EDA phasic (Estimate = 1.649, SE = 1.395, t(345) = 1.180, p = .238), and the centered data for k-coefficient (Estimate = 0.163, SE = 0.528, t(345) = 0.310, p = .758), were not significant.

Hence, the final predictive model B (PM-B) based on Cluster 2 centered data:

- **PM-B:** *prob_UPP= 1 / (1 + exp(-1 * (-0.3404 - 3.6547 * valence_b - 25.2911 * pupil_b + 1.6491 * phasic_b + 0.1626 * k_b)))*

Where prob_UPP = probability of usability pain point, x = variable (i.e. valence), and x_b = centered data of the variable x

In summary, the results partially support the study's second proposal by identifying shifts in visual attention and cognitive load, as measured by k-coefficient and pupil dilation, respectively, as solid indicators of usability pain points, which are then applied to the predictive model.

3.4.4 Performance evaluation: PM-A and PM-B tested on Task 1 and Task 2

In **Table 3**, the performance of PM-A and PM-B were evaluated based on recall and precision metrics for detecting usability pain points on Task 1 and 2 across different configurations of cut points, gap durations, and assumed durations.

Table 3

Performance evaluation of predictive models (PM-A & PM-B) for identifying usability pain points across different configurations

Obs	Cut point		Gap (ms)	Minimum Duration (s)	Assumed Duration (s)	Number of True Positives among the predicted UPP		Number of UPP predicted by the model		Number of induced UPP	*Recall*	*Precision*
	PM-A	PM-B				PM-A	PM-B	PM-A	PM-B			
1	0.7	0.95	8000	3	20	6	17	104	232	66	0.348	0.068
1	0.75	0.98	8000	3	20	3	14	25	191	66	0.258	0.079
1	0.73	0.98	8000	3	20	4	14	46	191	66	0.273	0.076
1	0.75	0.99	8000	3	20	3	10	25	164	66	0.197	0.069
1	0.75	0.999	8000	3	20	3	3	25	78	66	0.091	0.058
1	0.75	0.993	10000	8	15	1	7	7	99	66	0.121	0.075
1	0.75	0.993	10000	5	15	3	8	17	132	66	0.167	0.074
1	0.75	0.995	10000	5	15	3	6	17	115	66	0.136	0.068

Several configurations were tested to evaluate both model's performance. Cut points were adjusted across different configurations, ranging from 0.7 to 0.75 for PM-A and 0.95 to 0.995 for PM-B. A cut point indicates the threshold above which the predicted probability is considered a positive prediction (UPP detected). The gap duration, set between 8,000 and 10,000 milliseconds (ms), was used to merge neighbouring data points within a short time interval, considering them a single pain point. The minimum duration for a predicted pain point to be considered valid was either 3 or 5 seconds (s). To specify the time interval within which a UPP is detected, the assumed duration was fixed at 20 seconds for most configurations.

Table 4 presents the selected configuration to optimize the prediction performance regarding recall and precision, particularly for identifying around 100 distinct UPPs without necessarily maximizing these metrics. The selected configuration indicated that the models' performance had limited success in identifying true UPPs, with only 10.6% of the actual positive cases of UPPs detected (recall = 0.106), and among all predicted pain points, only 8% were actual UPPs (precision = 0.080).

Table 4

The selected configuration aims to predict approximately 100 distinct UPP

Obs	Cut point		Gap (ms)	Minimum Duration (s)		Assumed Duration (s)	Number of True Positives among the predicted UPP		Number of UPP predicted by the model		Number of Induced UPP	*Recall*	*Precision*
	PM-A	PM-B		PM-A	PM-B		PM-A	PM-B	PM-A	PM-B			
1	0.75	0.995	10000	5	10	15	3	4	17	71	66	0.106	0.080

This finding partially supports the study's second proposal. While the precision rate indicated limited accuracy, it suggested that the models have some capacity to identify actual pain points but may generate a substantial number of false positives. The relatively low recall implies that while some UPPs are detected, the models may not identify many actual pain points under the current configuration. However, these initial results highlighted the model's potential in detecting usability pain points. The ability to identify even a subset of UPPs represented progress toward the study's objective, as it showed the feasibility of using psychophysiological measures for real-time UPP detection.

3.4.5 Performance evaluation: PM-A and PM-B tested on Task 3

PM-A and PM-B trained models were applied to Task 3, where no manipulated pain points were induced, to test its ability to detect any spontaneous pain points that might naturally occur during user interaction. The models identified 116 pain point events, with 42 classified as UPP with k-coefficient as the significant predictor and 74 as UPP with pupil dilation as the significant predictor. While recall and precision are typically used to measure model performance, the recall rate could not be calculated because the ground truth for the spontaneous pain points is unknown. In this case, precision was assessed through expert evaluation, which analyzed each predicted instance to determine the likelihood of representing a true pain point.

Table 5 presents the number of UPP events categorized by the expert as likely to be a usability pain point.

Table 5

Expert's assessment of the predicted pain point

Assessment Scale	Number of UPP events
0 = Absolutely not a UPP	21
1 = Somewhat Like a UPP	39
2 = Most likely a UPP	33
3 = Absolutely a UPP	23

Following the expert assessment, a frequency analysis was conducted by creating a binary variable indicating whether the expert's evaluation (on a scale of 0–3) represented a higher likelihood of an actual pain point (>1) or not (≤1).

Table 6 presents the precision of the predicted UPP based on the expert assessment. The expert evaluation provided a precision of 47.6% of the instances predicted as UPP with k-coefficient predictor and 48.6% of the UPP with pupil dilation predictor, indicating that the expert ratings corroborated nearly half of the model’s predicted pain points.

Table 6

Precision of predicted UPP based on expert assessment

UPP based on significant predictor	Number of UPP events	Likelihood of an actual UPP (>1) based on expert assessment	*Precision*
UPP (k-coefficient)	42	20	0.476
UPP (pupil dilation)	74	36	0.486

*If expert assessment is >1, then the likelihood of an actual UPP is true (1), otherwise (0)

The findings further support the study’s second proposal that, while the predictive models have limitations, they can detect spontaneous UPP with reasonable precision, even in an unmanipulated task. The moderate precision suggests that the models capture genuine instances of user friction that were likely subtle but impactful enough to elicit changes in psychophysiological responses when encountering a usability pain point.

3.5 Discussion

The study identified psychophysiological signature patterns, defined by changes in emotional arousal, valence, visual attention, and cognitive load when users encountered usability pain points in enterprise systems. Moreover, using the captured psychophysiological data, this study developed predictive models and examined their reliability in identifying pain points across similar tasks. Three SaaS platforms were selected: Microsoft Dynamics 365 – CRM, Salesforce, and ServiceNow. An artificially manipulated pain point was placed randomly in the first or second task to evoke reactions, with data collected for model training. The models' reliability was then evaluated on a third task, free of manipulation.

Given the limited research on user behaviour in response to usability pain points, this study introduced a novel multimodal approach that uses psychophysiological data to identify usability pain points during interaction by assessing users' emotional, cognitive, and attentional responses.

3.5.1 Profiling users’ psychophysiological response to usability pain points

The study aimed to explore and identify users' distinct psychophysiological response patterns when encountering usability issues (RQ1). It successfully revealed characteristic profiles characterized by changes in emotional arousal, valence, visual attention, and cognitive load, revealing different user experience profiles. The cluster analysis identified four distinct user profiles with varied psychophysiological reactions to usability pain points (UPPs) during interaction to a digital interface. Two of these user profiles, provisionally named "Conventional Reactor A" (Cluster 1) and "Conventional Reactor B" (Cluster 2), indicated that the majority of the participants experienced the usability pain point as a significant disruption to their interaction with the SaaS interface.

Conversely, a small subset of participants responded differently, as found in the other two user profiles, provisionally named "Unconventional Reactor A" (Cluster 3) and "Unconventional Reactor B" (Cluster 4). However, why do participants exposed to identical UPP react differently?

The differentiation between these user profiles can potentially be interpreted through the lens of the fight, flight, freeze, and fawn responses. The fight or flight theory describes the physiological reaction that arises when an organism perceives danger or a threat to survival (Cannon, 1929). Humans adapt to stressful situations by engaging in "fight or flight" behaviour when encountering an impending danger or threat. The fight response involves confronting the threat, while the flight response entails escaping from it (Cannon, 1929).

Conventional Reactor A showed a slight positive increase in an emotional state after encountering a usability obstacle with heightened physiological arousal, placing the emotion felt by the participants in the high arousal/positive valence quadrant (Russell, 1980); this may suggest that participants were surprised when encountering the usability issue. There is a drastic increase in pupil size, which may indicate that participants' LC-NE system was in tonic mode and may suggest a disengagement in the current task, where attention is no longer primarily focused on task-relevant stimuli but also responds to irrelevant stimuli and shifting to an exploration strategy (Aston-Jones & Cohen., 2005; Gilzenrat et al., 2010). There was a decrease in the k-coefficient, which may indicate that the participants switched to ambient visual scanning behaviour (Kreijtz et al., 2016) after encountering the usability issue when exploring the SaaS enterprise environment. Conventional Reactor A profile possibly reflected a "flight" response when encountering the usability obstacle.

Whereas Conventional Reactor B showed an extremely significant association with the drastic decrease in pupil dilation. The drastic decline in pupil dilation may indicate that participants' LC-NE system was in phasic mode, which may be associated with high task engagement on the current task, where attention is concentrated on task-relevant stimuli to optimize performance, which may suggest that participants shifted to exploitation strategy upon encountering the usability issue (Aston-Jones & Cohen., 2005; Gilzenrat et al., 2010). Participants showed a strong negative emotional response and increased arousal, suggesting that the participants may have felt frustrated, distressed or annoyed (high arousal/negative valence quadrant) (Russell, 1980) upon encountering the usability obstruction. This may also indicate that participants focused on the stimulus before switching to ambient scanning behaviour (Kreijtz et al., 2016), where the k-coefficient exhibited by participants was slightly higher than that of Conventional Reactor A participants. Conventional Reactor B's response may be interpreted as a "fight" response when encountering a usability issue.

Additional responses, such as freeze and fawn, have been added to expand the fight or flight theory. The freeze response temporarily suspends the fight-or-flight response characterized by hyper-focused attention on a perceived threat (Kozlowska et al., 2015). The less commonly discussed fawn response involves appeasement behaviours aimed at reducing the danger from an aggressor (Owca, 2020).

Unconventional Reactor A profile exhibited a minimal decrease in emotional valence and a significant drop in EDA phasic, which may place the emotion felt by the participants in the low arousal/negative emotions quadrant (Russell, 1980), suggesting participants may have felt bored or tired. The pupil dilation increased, which may indicate that the participants' LC-NE system was in tonic mode and may present an exploration strategy (Aston-Jones & Cohen., 2005; Gilzenrat et al., 2010). Despite the exploration strategy, the participants exhibited focused attention (increased k-coefficient) (Kreijtz et al., 2016) while visually scanning the SaaS enterprise environment. This behaviour can potentially be a "fawn" response due to the indifferent emotion felt by the participants and disengagement from the current task, quickly diverting their focus to a different stimulus that they think might solve the current issue.

Lastly, the Unconventional Reactor B user profile presented a substantial increase in emotional valence, which may indicate a strong positive emotion experienced when encountering the usability challenge. However, there was a slight decrease in arousal, which may place the emotions felt by the participants in the low arousal/positive valence quadrant (Russell, 1980). This may suggest that the participants felt calm when encountering the usability issue. The pupil dilation decreased, which may indicate that the participants' LC-NE system was in phasic mode and may present an exploitation strategy (Aston-Jones & Cohen., 2005; Gilzenrat et al., 2010) supplemented by a notable increase in the k-coefficient, which may indicate focused attention (Kreijtz et al., 2016). This response by Unconventional Reactor B potentially represents a "freeze" response due to the calm emotions, exploitation strategy, and focused attention on the stimulus.

It is important to note that this study does not directly focus on the fight-flight-freeze-fawn responses but on the psychophysiological signature patterns exhibited by the participants upon encountering a UPP. The fight-flight-freeze-fawn responses (Cannon, 1929) were interpreted based on the existing research on information gathering theory (Pirolli & Card, 1999), adaptive gain theory (Aston-Jones & Cohen., 2005), arousal-valence model (Russell, 1980) and psychophysiological theories.

3.5.2 Assessing predictive models' performance in usability pain point detection

The study aimed to evaluate the extent to which a psychophysiological signature of a user experiencing usability issues can reliably identify usability pain points in another similar task (RQ2). Predictive models were trained using robust indicators—specifically, shifts in visual attention and cognitive load, as measured by the k-coefficient and pupil dilation found in Conventional Reactor A and Conventional Reactor B user profiles, respectively. These models were applied to detect the occurrence of usability pain points, with their performance assessed through precision and recall metrics. Although the initial performance evaluation results where the predictive models are applied to the pain-point-induced tasks (Task 1 and 2) showed a low recall of 10.6% and precision of 8%, the predictive models highlighted its potential in detecting usability pain points. The ability to identify even a subset of UPPs represents progress toward the study's objective, as it showed the feasibility of using psychophysiological measures for real-time UPP detection.

To further address RQ2, on the second performance evaluation, the predictive models were applied to a natural task – no artificially manipulated pain point – to test its ability to detect any spontaneous pain point that may naturally occur during the interaction. The models detected 116 UPP events, which an expert then validated. The expert evaluation provided a precision of 47.6% of the instances predicted as UPP with k-coefficient predictor and 48.6% of the UPP with pupil dilation predictor, indicating that the expert ratings corroborated nearly half of the model's predicted pain points. The results showed that while the model is not perfect, it can detect spontaneous usability pain points with reasonable precision, even in an unmanipulated task. This capability supported the model's utility in real-world scenarios where usability issues arise organically.

However, the inability to calculate recall due to unknown ground truth for spontaneous pain points limits a complete evaluation of the model's performance. Without recall, it is difficult to determine how many actual UPPs the models missed. Thus, the models should be used cautiously, and an inter-rater expert evaluation is recommended to mitigate subjectivity bias. Also, the models may require further refinement to improve reliability in detecting true UPPs in similar tasks.

These mixed results suggest that while psychophysiological data holds promise for real-time identification of pain points, individual differences in response intensity complicate the model's accuracy. This variation raises questions about the reliability of using psychophysiological measures alone for usability assessment, highlighting the need to account for response variability in predictive models. From a UX research perspective, these findings suggest that psychophysiological data can be a valuable tool for identifying pain points but may be more effective when combined with other data sources, such as behavioral observations or self-reports. This approach could enhance the interpretability of psychophysiological responses and allow for a more comprehensive understanding of user experiences. As a result, psychophysiological measures could serve as a complementary method in UX research, providing objective insights that help capture real-time reactions, but requiring contextualization to ensure accuracy and relevance across diverse user profiles and tasks.

3.5.3 Contributions

3.5.3.1 Theoretical Contributions. The study challenged the traditional notion that all users react uniformly to usability issues. The study enriched existing UX and HCI research theories by identifying distinct psychophysiological response patterns across user profiles. It highlighted the importance of individual differences in emotional, cognitive, and attentional responses. By demonstrating that specific psychophysiological measures, such as pupil dilation and visual attention shifts (k-coefficient), can reliably signal usability issues, the study advanced the theoretical understanding of objective, non-verbal cues as indicators of user experience disruptions. The study provided a comprehensive framework for user behaviour analysis regarding usability issues and user interactions by integrating psychophysiological measures of emotional arousal, valence, cognitive load, and visual attention. This multimodal approach added depth to theoretical models, emphasizing the value of combining diverse data streams for richer insights into user interactions. The study's application of psychophysiological principles to usability assessment linked the fields of psychophysiology and UX, contributing to a novel interdisciplinary perspective that lays the groundwork for future theoretical explorations that connect physiological responses to user experience metrics. The findings also provided empirical evidence linking emotional, cognitive, and visual attention responses to usability issues, enriching theoretical discussions on how these factors influence user behaviour during complex tasks.

3.5.3.2 Practical Contributions. This study provided actionable insights for improving the usability of enterprise systems by leveraging psychophysiological data, predictive modelling, and adaptive design strategies. Incorporating tools like eye tracking, EDA sensors, and facial expression recognition software offers more profound insights into users' emotional and cognitive states, revealing implicit usability issues. The predictive models developed in this study enabled proactive usability management by identifying usability challenges in real-time and allowing immediate interventions to enhance user experiences. Given the variability in user responses, adaptive interfaces with customizable features like task shortcuts and personalized guidance are recommended to accommodate diverse user-profiles and minimize user frustration. The study highlighted that shifts in pupil dilation and k-coefficient are reliable indicators of usability issues and suggested that high-impact usability adjustments should focus on these indicators. These strategies not only improve usability but also enhance user satisfaction, productivity, and system adoption.

In practice, UX designers could use a data-driven approach to enhance decision-making by using real-time user insights, complimenting subjective feedback with a more objective one. System developers could integrate the predictive models into the backend of enterprise platforms to proactively address user challenges, enhancing user experience and system robustness. Human factors specialists could utilize the predictive models to deeply analyze user-system interactions, leading to evidence-based recommendations for improving task efficiency and reducing errors. Human resources and learning development experts could incorporate predictive models into their training modules and platforms to mitigate users' learning challenges, hence improving employees' training experience, particularly in mastering new digital interfaces. Management could leverage the predictive model's insights to improve employee productivity, satisfaction, and retention, directly supporting organizational success.

The study highlighted its practical relevance by outlining actionable findings for certain professional positions, thereby facilitating the transformation of enterprise systems into more user-friendly, efficient, and adaptive environments. In addition, the study could empower professionals across disciplines to leverage the findings for meaningful user experience enhancements.

3.5.3.3 Methodological Contributions. This study also contributes methodologically by introducing a novel multimodal approach to identifying usability pain points by using various psychophysiological methods to assess the different aspects of the user's response, which are characterized by changes in emotional arousal, valence, cognitive load, and visual attention when users encounter a usability obstruction. In addition, by integrating cluster analysis and logistic regression models to predict usability challenges, this study demonstrated a robust multimodal approach for analyzing psychophysiological data in usability studies. The clustering method enabled us to capture meaningful psychophysiological signatures, revealing different user experiences, a factor that enhances the interpretability of psychophysiological data. In addition, through logistic regression the study was able to identify significant indicators to usability pain points (pupil dilation and k-coefficient) and developed models that were able to detect usability pain points on a natural task.

3.5.4 Limitations and Future Work

This study adopted a controlled laboratory setting and specific task parameters to focus on the feasibility of identifying usability pain points (UPPs) using psychophysiological measures. While this approach allowed for rigorous testing and analysis of the proposed predictive models, certain aspects were necessarily constrained, creating opportunities for future research to extend and refine the findings.

The modest recall (10.6%) and precision (8%) rates observed in training the predictive models reflect the complexity of reliably identifying user pain points. These rates highlighted the inherent challenge of capturing the nuanced variability in user responses. While recall and precision are typically used to measure model performance, the recall rate could not be calculated because the ground truth for the spontaneous pain points is unknown. In this case, the decision to use a single expert evaluation for validating model precision was a practical choice for this study. However, the reliance on a single expert introduced subjectivity that future studies could address by incorporating multi-rater assessments or alternative validation techniques.

Although the study introduced a novel perspective by interpreting the link between psychophysiological responses to existing theories like the arousal-valence model, adaptive gain theory, and theory of visual scanning behaviour, the actual classification of user responses into the fight-flight-freeze-fawn categories was not measured and validated. The reliance on existing theories to infer these behavioural responses might oversimplify the complexity of human reactions, which are influenced by multiple factors beyond the psychophysiological measures considered. This study opens avenues for future work to validate the interpretation of fight-flight-freeze-fawn response categorization by incorporating more direct behavioural or self-reported measures and more advanced psychophysiological measures, providing a richer, nuanced understanding of user reactions.

The controlled environment and specific sample size enabled precise measurement of psychophysiological responses, but these factors limit the generalizability of findings to broader, real-world contexts. User interactions could be influenced by a variety of contextual factors, including workflow complexities, stress levels, and multitasking demands. Future studies should extend this work by applying the predictive models in real-world scenarios, considering different tasks, other enterprise systems, and diverse user groups. This expansion would validate the models' applicability and enhance their robustness in identifying usability challenges across varied scenarios.

Additionally, while this study primarily used k-coefficient and pupil dilation as significant indicators in developing the predictive models, future studies could explore and employ various indicators, machine learning techniques and other advanced signal processing methods that could significantly improve the predictive model's performance. The combination of these technical advancements with multi-dimensional psychophysiological data would support more accurate, real-time detection of usability pain points, enabling more profound insights into user behaviour and enhancing the utility of predictive models in UX research.

Moreover, future studies could explore whether other psychological or behavioural frameworks align more closely with the psychophysiological signatures observed. Expanding the study to explore additional psychophysiological measures, alongside developing the theoretical framework to include more dimensions of emotional and cognitive responses, could provide a more comprehensive understanding of users' interaction with usability challenges.

By addressing these avenues, future work can build on the foundation established by this study, advancing our understanding of psychophysiological responses to usability challenges and improving the tools available for real-time usability assessment.

3.6 Conclusion

This research explored psychophysiological responses to usability pain points in digital enterprise environments. It identified distinct psychophysiological signatures when users encounter usability issues (RQ1) and determined the extent to which these patterns reliably predict pain points in other tasks (RQ2). By integrating a novel multimodal approach to psychophysiological measures of emotional arousal, valence, visual attention, and cognitive load, the study provided objective, comprehensive insights into user experiences when encountering usability challenges.

Cluster analysis revealed four unique psychophysiological response profiles, which indicated users' diverse emotional, attentional, and cognitive responses to usability pain points. The study identified shifts in pupil dilation and k-coefficient as reliable indicators of usability challenges and utilized these indicators to train predictive models. The models demonstrated moderate success in detecting spontaneous usability pain points in non-manipulated tasks. The results highlighted the predictive models' real-time capability to detect usability challenges.

This research opened opportunities for expanding the scope of psychophysiological applications. Future studies could explore cross-industry implementation, refine predictive algorithms for real-world variability, and integrate these insights with machine learning to further advance human-computer interaction. By linking technology and user experience, this work contributed to a more user-centered and practical approach to identifying and predicting usability challenges experienced by users in real-time, improving users' experience and enhancing system design in enterprise environments.

In conclusion, this research underscores the value of psychophysiological measures in identifying usability challenges beyond traditional methods. By advancing predictive models and providing actionable insights, it paves the way for more precise and user-centered design in digital enterprise environments, contributing to a future of enhanced user satisfaction and system performance.

References

Ajenaghughrure, I., Sousa, S., & Lamas, D. (2020). Measuring trust with psychophysiological signals: a systematic mapping study of approaches used. *Multimodal Technologies and Interaction*, 4(3), 63. https://doi.org/10.3390/mti4030063

Alexandros, L., & Michalis, X. (2013). The physiological measurements as a critical indicator in users' experience evaluation. In *Proceedings of the 17th Panhellenic Conference on Informatics* (pp. 258-263).

Alexandrovsky, D., Putze, S., Bonfert, M., Höffner, S., Michelmann, P., Wenig, D., Malaka, R., & Smeddinck, J. D. (2020). Examining Design Choices of Questionnaires in VR User Studies. *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, 1–21. https://doi.org/10.1145/3313831.3376260

Apraiz Iriarte, A., Lasa, G., & Mazmela, M. (2021). Evaluating User Experience with physiological monitoring: A Systematic Literature Review. *Dyna (Bilbao)*, 8, 21. https://doi.org/10.6036/NT10072

Aston-Jones, G., & Cohen, J. D. (2005). AN INTEGRATIVE THEORY OF LOCUS COERULEUS-NOREPINEPHRINE FUNCTION: Adaptive gain and optimal performance. *Annual Review of Neuroscience*, 28(1), 403–450. https://doi.org/10.1146/annurev.neuro.28.061604.135709

Bargas-Avila, J. A., & Hornbæk, K. (2011). Old wine in new bottles or novel challenges: A critical analysis of empirical studies of user experience. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 2689–2698. https://doi.org/10.1145/1978942.1979336

Barreto, A., Zhai, J., & Adjouadi, M. (2007). Non-intrusive physiological monitoring for Automated stress detection in Human-Computer Interaction. In *Springer eBooks* (pp. 29–38). https://doi.org/10.1007/978-3-540-75773-3_4

Bergstrom, J. R., Duda, S., Hawkins, D., & McGill, M. (2014). Physiological response measurements. In *Elsevier eBooks* (pp. 81–108). https://doi.org/10.1016/b978-0-12-408138-3.00004-2

Beri, D., & K, J. R. (2019). Physiological Correlates of Arousal: A Metaanalytic Review. *Journal of Neurology and Neuroscience*, 10(04). https://doi.org/10.36648/2171-6625.10.4.302

Berni, A., Borgianni, Y., Basso, D., & Carbon, C. (2023). Fundamentals and issues of user experience in the process of designing consumer products. *Design Science*, 9. https://doi.org/10.1017/dsj.2023.8

BIOPAC Systems, Inc. (n.d.). AcqKnowledge software [Software]. Goleta, CA.

BIOPAC Systems, Inc. (n.d.). MP-150 [Product brochure or manual]. Goleta, CA.

Borys, M., & Plechawska-Wójcik, M. (2017). Eye-tracking metrics in perception and visual attention research. *EJMT*, 3, 11-23.

Boucsein, W. (2012). Electrodermal activity. *Springer Science & Business Media*.

Bradley, M. M., Miccoli, L., Escrig, M. A., & Lang, P. J. (2008). The pupil as a measure of emotional arousal and autonomic activation. *Psychophysiology*, 45(4), 602–607. https://doi.org/10.1111/j.1469-8986.2008.00654.x

Braithwaite, J. J., Watson, D. G., Jones, R., & Rowe, M. (2013). A guide for analysing electrodermal activity (EDA) & skin conductance responses (SCRs) for psychological experiments. *Psychophysiology*, 49(1), 1017-1034.

Brown, J., Chatterjee, N., Younger, J., & Mackey, S. (2011). Towards a physiology-based measure of pain: patterns of human brain activity distinguish painful from non-painful thermal stimulation. *PLoS ONE*, 6(9), e24124. https://doi.org/10.1371/journal.pone.0024124

Bruun, A., Law, E. L.-C., Heintz, M., & Alkly, L. H. A. (2016). Understanding the Relationship between Frustration and the Severity of Usability Problems: What can Psychophysiological Data (Not) Tell Us? *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, 3975–3987. https://doi.org/10.1145/2858036.2858511

Carmichael, L., Poirier, S.-M., Coursaris, C., Léger, P.-M., & Senecal, S. (2022). Users' Information Disclosure Behaviors during Interactions with Chatbots: The Effect of Information Disclosure Nudges. *Applied Sciences*, 12, 12660. https://doi.org/10.3390/app122412660

Cannon, W. B. (1929). Organization for Physiological Homeostasis. *Physiological Reviews*, 9(3), 399–431. https://doi.org/10.1152/physrev.1929.9.3.399

Chang, L., Gianaros, P., Manuck, S., Krishnan, A., & Wager, T. (2015). A sensitive and specific neural signature for picture-induced negative affect. *Plos Biology*, 13(6), e1002180. https://doi.org/10.1371/journal.pbio.1002180

Charland, P., Léger, P.-M., Senecal, S., Courtemanche, F., Mercier, J., Skelling, Y., & L. LeMoyne, E. (2015). Assessing the Multiple Dimensions of Engagement to Characterize Learning: A Neurophysiological Perspective. *Journal of Visualized Experiments*, 101. https://doi.org/10.3791/52627

Cohen, J. D., McClure, S. M., & Yu, A. J. (2007). Should I stay or should I go? How the human brain manages the trade-off between exploitation and exploration. *Philosophical Transactions of the Royal Society B Biological Sciences*, 362(1481), 933–942. https://doi.org/10.1098/rstb.2007.2098

Costa, I., Silva, W., Damian, A., Rivero, L., Gadelha, B., Teixeira de Oliveira, E., & Conte, T. (2016). *An Empirical Study to Evaluate the Feasibility of a UX and Usability Inspection Technique for Mobile Applications* (p. 599). https://doi.org/10.18293/SEKE2016-127

Dair, Z., Dockray, S., & O'Reilly, R. (2023). Complex Adaptive Systems and Psychophysiological Data - An Exploratory Approach. *2023 31st Irish Conference on Artificial Intelligence and Cognitive Science (AICS)*., 1-4. https://doi.org/10.1109/AICS60730.2023.10470799.

Dirican, A. C., & Göktürk, M. (2011). Psychophysiological measures of human cognitive states applied in human computer interaction. *Procedia Computer Science*, 3, 1361–1367. https://doi.org/10.1016/j.procs.2011.01.016

Ferreira, E., Ferreira, D., Kim, S., Siirtola, P., Röning, J., Forlizzi, J. F., & Dey, A. K. (2014). Assessing real-time cognitive load based on psycho-physiological measures for younger and older adults. 2014 *IEEE Symposium on Computational Intelligence, Cognitive Algorithms, Mind, and Brain* (CCMB), 39–48. https://doi.org/10.1109/CCMB.2014.7020692

Garbas, J.U., Ruf, T., Unfried, M. and Dieckmann, A., (2013). Towards Robust Real-Time Valence Recognition from Facial Expressions for Market Research Applications. *2013 Humaine Association Conference on Affective Computing and Intelligent Interaction.*, 570-575. https://doi.org/10.1109/ACII.2013.100.

Gartner. (n.d.) *Fueling the future of business*. (n.d.). Gartner. https://www.gartner.com/document/4023333?ref=solrAll&refval=420472053

Gilzenrat, M. S., Nieuwenhuis, S., Jepma, M., & Cohen, J. D. (2010). Pupil diameter tracks changes in control state predicted by the adaptive gain theory of locus coeruleus function. *Cognitive Affective & Behavioral Neuroscience*, 10(2), 252–269. https://doi.org/10.3758/cabn.10.2.252

Giroux-Huppé, C., Sénécal, S., Fredette, M., Chen, S. L., Demolin, B., & Léger, P. (2019). Identifying psychophysiological pain points in the online user journey: the case of online grocery. In *Lecture notes in computer science* (pp. 459–473). https://doi.org/10.1007/978-3-030-23535-2_34

Haselmann, T., & Vossen, G. (2011). Software-as-a-Service in Small and Medium Enterprises: An Empirical Attitude assessment. *In Lecture notes in computer science* (pp. 43–56). https://doi.org/10.1007/978-3-642-24434-6_4

Hopstaken, J. F., Van Der Linden, D., Bakker, A. B., & Kompier, M. A. (2015). The window of my eyes: Task disengagement and mental fatigue covary with pupil dynamics. *Biological Psychology*, 110, 100–106. https://doi.org/10.1016/j.biopsycho.2015.06.013

Horwitz, A. G., McCarthy, K., & Sen, S. (2024). A review of the peak-end rule in mental health contexts. *Current Opinion in Psychology*, 101845. https://doi.org/10.1016/j.copsyc.2024.101845

Hudon, A., Demazure, T., Karran, A., Léger, P.-M., & Senecal, S. (2021). *Explainable Artificial Intelligence (XAI): How the Visualization of AI Predictions Affects User Cognitive Load and Confidence* (pp. 237–246). https://doi.org/10.1007/978-3-030-88900-5_27

Huo, F., Zhao, Y., Chai, C., & Fang, F. (2023). A user experience map design method based on emotional quantification of in-vehicle HMI. *Humanities and Social Sciences Communications*, 10. https://doi.org/10.1057/s41599-023-01761-4

Ibarra-Noriega, A. M., Yansane, A., Mullins, J., Simmons, K., Skourtes, N., Holmes, D., White, J., Kalenderian, E., & Walji, M. F. (2024). Evaluating and improving the usability of a mHealth platform to assess postoperative dental pain. *JAMIA Open*, 7(1). https://doi.org/10.1093/jamiaopen/ooae018

Inan Nur, A., B. Santoso, H., & O. Hadi Putra, P. (2021). The Method and Metric of User Experience Evaluation: A Systematic Literature Review. *Proceedings of the 2021 10th International Conference on Software and Computer Applications*, 307–317. https://doi.org/10.1145/3457784.3457832

ISO 9241-210:2019(en), Ergonomics of human-system interaction — *Part 210: Human-centred design for interactive systems.* (n.d.). https://www.iso.org/obp/ui/en/#iso:std:iso:9241:-210:ed-2:v1:en

ISO, W. (1998). 9241-11. Ergonomic requirements for office work with visual display terminals (VDTs). The international organization for standardization, 45(9), 22.

ISO, I. (1999). 13407: Human-centred design processes for interactive systems. Geneva: ISO.

Kahneman, D., Fredrickson, B. L., Schreiber, C. A., & Redelmeier, D. A. (1993). When more pain is preferred to less: adding a better end. *Psychological Science*, 4(6), 401–405. https://doi.org/10.1111/j.1467-9280.1993.tb00589.x

Kim, J., Lee, S., Min, K. S., Jung, H. H., Lee, J. E., Kim, S. J., … & Chang, J. W. (2013). Ventral posterolateral deep brain stimulation treatment for neuropathic pain shortens pain response after cold stimuli. *Journal of Neuroscience Research*, 91(7), 997-1004. https://doi.org/10.1002/jnr.23222

Klaus, H., Rosemann, M. & Gable, G.G. (2000). What is ERP?. *Information Systems Frontiers 2*, 141–162. https://doi.org/10.1023/A:1026543906354

Klotins, E., Unterkalmsteiner, M., & Gorschek, T. (2018). Software engineering in start-up companies: An analysis of 88 experience reports. *Empirical Software Engineering*, 24(1), 68–102. https://doi.org/10.1007/s10664-018-9620-y

Korosec-Serfaty, M., Riedl, R., Senecal, S., & Léger, P.-M. (2022). Attentional and Behavioral Disengagement as Coping Responses to Technostress and Financial Stress: An Experiment Based on Psychophysiological, Perceptual, and Behavioral Data. *Frontiers in Neuroscience*, 16, 883431. https://doi.org/10.3389/fnins.2022.883431

Kozlowska, K., Walker, P., McLean, L., & Carrive, P. (2015). Fear and the defense cascade. *Harvard Review of Psychiatry*, 23(4), 263–287. https://doi.org/10.1097/hrp.0000000000000065

Kreger, A. (2022). *Digital banking user experience: Solve user Pain Points through information architecture*. https://doi.org/10.13140/RG.2.2.34542.08000

Krejtz, K., Duchowski, A., Krejtz, I., Szarkowska, A., & Kopacz, A. (2016). Discerning Ambient/Focal Attention with CoefficientK. *ACM Transactions on Applied Perception*, 13(3), 1–20. https://doi.org/10.1145/2896452

Kucewicz, M. T., Dolezal, J., Kremen, V., Berry, B. M., Miller, L. R., Magee, A. L., Fabian, V., & Worrell, G. A. (2018). Pupil size reflects successful encoding and recall of memory in humans. *Scientific Reports*, 8(1). https://doi.org/10.1038/s41598-018-23197-6

Kwak, D., Ma, X., & Kim, S. (2021). When does social desirability become a problem? Detection and reduction of social desirability bias in information systems research. *Information & Management*, 58(7), 103500. https://doi.org/10.1016/j.im.2021.103500

Lamme, V. A. (2003). Why visual attention and awareness are different. *Trends in Cognitive Sciences*, 7(1), 12–18. https://doi.org/10.1016/s1364-6613(02)00013-x

Law, E. L., Van Schaik, P., & Roto, V. (2013). Attitudes towards user experience (UX) measurement. *International Journal of Human-Computer Studies*, 72(6), 526–541. https://doi.org/10.1016/j.ijhcs.2013.09.006

Léger, P.-M., Senecal, S., Courtemanche, F., Guinea, A., Titah, R., Fredette, M., & L. LeMoyne, E. (2014). Precision is in the Eye of the Beholder: Application of Eye Fixation-Related Potentials to Information Systems Research. *Journal of the Association for Information Systems*, 15. https://doi.org/10.17705/1jais.00376

Li, S. & Deng, W. (2022). Deep Facial Expression Recognition: A Survey. *IEEE Transactions on Affective Computing*, 13(3), 1195-1215. https://doi.org/10.1109/TAFFC.2020.2981446.

Lounis, C. A., Hassoumi, A., Lefrancois, O., Peysakhovich, V., & Causse, M. (2020). Detecting ambient/focal visual attention in professional airline pilots with a modified Coefficient K: a full flight simulator study. *ACM Symposium on Eye Tracking Research and Applications*, 10, 1–6. https://doi.org/10.1145/3379157.3391412

Lovallo, W. (2013). Psychophysiology: Theory and methods. In *Springer eBooks*, 1569–1572. https://doi.org/10.1007/978-1-4419-1005-9_484

Maia, C. L. B., & Furtado, E. S. (2019). An Approach to Analyze User's Emotion in HCI Experiments Using Psychophysiological Measures. *IEEE Access*, 7, 36471–36480. IEEE Access. https://doi.org/10.1109/ACCESS.2019.2904977

Mehrabian, A. (2017). Communication without words. In *Routledge eBooks* (pp. 193–200). https://doi.org/10.4324/9781315080918-15

Microsoft. Dynamics 365 Customer Service. Retrieved from https://dynamics.microsoft.com/en-us/customer-service/

Minzenberg, M. J., Watrous, A. J., Yoon, J. H., Ursu, S., & Carter, C. S. (2008). Modafinil shifts human locus coeruleus to Low-Tonic, High-Phasic activity during functional MRI. *Science*, 322(5908), 1700–1702. https://doi.org/10.1126/science.1164908

Mithun, M. B., Karmakar, S., Varghese, T., Jaiswal, D., Chatterjee, D., Gavas, R. D., Ramakrishnan, R. K., & Pal, A. (2023). Mind Indriya: A System for Simultaneous Assessment of Cognitive Load, Anxiety and Visual Attention. *2023 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, 3234–3240. https://doi.org/10.1109/SMC53992.2023.10394104

Murphy, P. R., O'Connell, R. G., O'Sullivan, M., Robertson, I. H., & Balsters, J. H. (2014). Pupil diameter covaries with BOLD activity in human locus coeruleus. *Human Brain Mapping*, 35(8), 4140–4154. https://doi.org/10.1002/hbm.22466

Nederhof, A. J. (1985). Methods of coping with social desirability bias: A review. *European Journal of Social Psychology*, 15(3), 263–280. https://doi.org/10.1002/ejsp.2420150303

Noldus (2021). FaceReader: Tool for automatic analysis of facial expressions: Version 9 [Software]. Wageningen, The Netherlands: Noldus Information Technology B.V.

Noldus. MediaRecorder - Synchronous video recordings. https://www.noldus.com/mediarecorder-human

Noldus. The Observer XT. Behavioral coding - Event logging software | Behavioral Coding - Event Logging Software | the Observer XT. https://www.noldus.com/observer-xt

Owca, J. (2020). *The association between a Psychotherapist's theoretical orientation and perception of complex trauma and repressed anger in the fawn response* (Order No. 28086275). ProQuest Dissertations & Theses Global Closed Collection. (2447256147). Retrieved from https://login.proxy2.hec.ca/login?url=https://www.proquest.com/dissertations-theses/association-between-psychotherapist-s-theoretical/docview/2447256147/se-2

Parsons, T., Asbee, J., & Courtney, C. (2023). Interaction of Cognitive and Affective Load Within a Virtual City | *2023 IEEE Transactions on Affective Computing*, vol. 14, no. 4, pp. 2768-2775. https://doi.org/10.1109/TAFFC.2022.3220953

Partala, T., & Surakka, V. (2003). Pupil size variation as an indication of affective processing. *International Journal of Human-Computer Studies*, 59(1–2), 185–198. https://doi.org/10.1016/s1071-5819(03)00017-x

Partala, T., & Surakka, V. (2003). Pupil size variation as an indication of affective processing. *International Journal of Human-Computer Studies*, 59(1–2), 185–198. https://doi.org/10.1016/s1071-5819(03)00017-x

Perrig, S. A. C., Aeschbach, L. F., Scharowski, N., von Felten, N., Opwis, K., & Brühlmann, F. (2024). Measurement practices in user experience (UX) research: A systematic quantitative literature review. *Frontiers in Computer Science*, 6. https://doi.org/10.3389/fcomp.2024.1368860

Pirolli, P., & Card, S. (1999). Information foraging. *Psychological Review*, 106(4), 643–675. https://doi.org/10.1037/0033-295x.106.4.643

Pittet, P., & Barthélémy, J. (2015). Experience of formal application ontology development to enhance user understanding in a Geo Business Intelligence SAAS platform. In *Lecture notes in business information processing*, 51–62. https://doi.org/10.1007/978-3-319-21545-7_5

Platzer, D. (2018). Regarding the Pain of Users: Towards a Genealogy of the "Pain Point." *Ethnographic Praxis in Industry Conference Proceedings*, 2018(1), 301–315. https://doi.org/10.1111/1559-8918.2018.01209

Putze, S., Alexandrovsky, D., Putze, F., Höffner, S., Smeddinck, J. D., & Malaka, R. (2020). Breaking The Experience: Effects of Questionnaires in VR User Studies. *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, 1–15. https://doi.org/10.1145/3313831.3376144

Roto, V., & Kaasinen, E. (2008, September). The second international workshop on mobile internet user experience. In *Proceedings of the 10th international* *conference on Human computer interaction with mobile devices and services*, 571-573. https://doi.org/10.1145/1409240.140935

Russell, J. A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology*, 39(6), 1161–1178. https://doi.org/10.1037/h0077714

Salesforce. What is Salesforce? Retrieved from https://www.salesforce.com/products/what-is-salesforce/

ServiceNow. (n.d.). What is ServiceNow? Retrieved from https://www.servicenow.com/what-is-servicenow.html

Shackel, B. (2009). Usability – Context, framework, definition, design and evaluation. *Interacting With Computers*, 21(5–6), 339–346. https://doi.org/10.1016/j.intcom.2009.04.007

Statista. (n.d.). *Software as a Service - Canada | Statista Market forecast.* https://www.statista.com/outlook/tmo/public-cloud/software-as-a-service/canada

Sweller, J., van Merrienboer, J. J. G., & Paas, F. G. W. C. (1998). Cognitive Architecture and Instructional Design. *Educational Psychology Review*, 10(3), 251–296. https://doi.org/10.1023/A:1022193728205

Swoboda, D., Boasen, J., Léger, P., Pourchon, R., & Sénécal, S. (2022). Comparing the Effectiveness of Speech and Physiological Features in Explaining Emotional Responses during Voice User Interface Interactions. *Applied Sciences*, 12(3), 1269. https://doi.org/10.3390/app12031269

Tobii AB (2024). Tobii Pro Lab (Version v 1.217) [Computer software]. Danderyd, Sweden: Tobii AB.

Tobii AB (2024). Tobii Pro Lab User Manual (Version v 1.217). Tobii AB, Danderyd, Sweden.

Van Acker, B. B., Parmentier, D. D., Vlerick, P., & Saldien, J. (2018). Understanding mental workload: from a clarifying concept analysis toward an implementable framework. *Cognition Technology & Work*, 20(3), 351–365. https://doi.org/10.1007/s10111-018-0481-3

Van Der Wel, P., & Van Steenbergen, H. (2018). Pupil dilation as an index of effort in cognitive control tasks: A review. *Psychonomic Bulletin & Review*, 25(6), 2005–2015. https://doi.org/10.3758/s13423-018-1432-y

Vanneste, P., Raes, A., Morton, J., Bombeke, K., Van Acker, B. B., Larmuseau, C., Depaepe, F., & Van Den Noortgate, W. (2020). Towards measuring cognitive load through multimodal physiological data. *Cognition Technology & Work*, 23(3), 567–585. https://doi.org/10.1007/s10111-020-00641-0

Vignaux, M., Léger, P.-M., Charland, P., Salame, Y., Durand, E., Bouillot, N., Pardoen, M., & Senecal, S. (2021). An Exploratory Study on the Impact of Collective Immersion on Learning and Learning Experience. *Multimodal Technologies and Interaction*, 5, 17. https://doi.org/10.3390/mti5040017

Visser, L., Tollenaar, M., Bosch, J., Doornen, L., Haes, H., & Smets, E. (2017). Are psychophysiological arousal and self-reported emotional stress during an oncological consultation related to memory of medical information? an experimental study. Stress, 20(1), 103-111. https://doi.org/10.1080/10253890.2017.1286323

Vrijheid, M., Armstrong, B., Bedard, D., Brown, J., Deltour, I., Iavarone, I., Krewski, D., Lagorio, S., Moore, S., Richardson, L., Giles, G., McBride, M., Parent, M.-E., Siemiatycki, J., & Cardis, E. (2008). Recall bias in the assessment of exposure to mobile phones. *Journal of Exposure Science & Environmental Epidemiology*, 19, 369–381. https://doi.org/10.1038/jes.2008.27

Wager, T., Atlas, L., Lindquist, M., Roy, M., Woo, C., & Kross, E. (2013). An fmri-based neurologic signature of physical pain. New England Journal of Medicine, 368(15), 1388-1397. https://doi.org/10.1056/nejmoa1204471

Wals, S. F., & Wichary, S. (2022). Under Pressure: Cognitive Effort During Website-Based Task Performance is Associated with Pupil Size, Visual Exploration, and Users' Intention to Recommend. *International Journal of Human-Computer Interaction*, 39(18), 3504–3515. https://doi.org/10.1080/10447318.2022.2098576

Yang, Y. & Sun, Y.(2017). *Facial expression recognition based on Arousal-Valence Emotion Model and Deep Learning method*. IEEE Conference Publication | IEEE Xplore. https://ieeexplore.ieee.org/document/8789024

Yusuf, M., Silas, F. A., & Haruna, S. (2018). Implementing Personnel Management System as SaaS. *Circulation in Computer Science*, 3(5), 1-6. https://doi.org/10.22632/ccs-2018-252-86

Chapter 4
Can We Predict User Frustration? A Novel Approach to Identifying Usability Pain Points in Real-Time

Kelvin Jacinto, Sylvain Sénécal, Constantinos Coursaris, Pierre-Majorique Léger, Shang-Lin Chen, Thaddé Rolon-Mérette, Alexander Karran

HEC Montreal

4.1 Introduction

What if we could understand user frustration in real time and predict usability issues before they arise? With the rise of Software as a Service (SaaS) in the digital business landscape (Statista; Gartner; Haselmann & Vossen, 2011), the usability of enterprise software is crucial. From customer relationship management (CRM) and Human Resource Management (HRM) systems to cloud-based platforms, these tools play a vital role in an organization’s daily operations (Klaus et al., 2000). However, the complexity of these enterprise systems can create significant usability issues, leading to frustrated users, reduced productivity, and even higher employee turnover. Traditional usability assessment methods, such as post-task questionnaires and interviews, often rely on subjective feedback (Bargas-Avila & Hornbæk, 2011; Law et al., 2013; Inan Nur et al., 2021; Perrig et al., 2024), which can be limited by social desirability and recall bias (Kwak et al., 2021; Vrijheid et al., 2008; Horwitz et al., 2024).

This research addressed the limitations of self-reported assessment by adopting a novel, multimodal, data-driven approach to usability assessment. This study provided a real-time, objective view of user pain points in SaaS enterprise environments by leveraging psychophysiological data from eye tracking metrics, electrodermal activity (EDA), and facial expressions. The insights gained from this approach can empower managers, system designers and UX professionals to design systems that better meet user needs (Klotins et al., 2018; Roto & Kaasinen, 2008), ultimately improving user satisfaction, productivity (Costa et al., 2016), and the organization's return on investment (ROI).

4.2 Key Research Question

The study was guided a primary research question:

- **How can users' psychophysiological response patterns be identified and leveraged to reliably detect usability challenges in similar tasks?**

This research aimed to provide a nuanced understanding of how users react to usability challenges and how these insights can be used to train predictive models in identifying usability pain points in a similar task in real time.

4.3 Research Methodology

This experimental study involved 86 participants, each tasked with completing interactions on three selected SaaS platforms: Microsoft Dynamics 365 CRM, Salesforce, and ServiceNow. Participants were exposed to manipulated usability disruptions designed to evoke natural responses to usability pain points, with data collected on their emotional arousal, valence, cognitive load, and visual attention. Using a combination of non-invasive tools—such as an eye tracker, EDA sensors, and Facial Expression Recognition (FER)—the study tracked users' psychophysiological responses in real-time.

The research followed a multimodal approach, integrating several psychophysiological measures to develop predictive models that could accurately detect and forecast usability pain points. A cluster analysis was performed to identify group of participants based on their psychological responses. Then, the predictive models were trained using logistic regression and evaluated using recall and precision metrics. Lastly, the predictive models' performance was validated through expert evaluation.

4.4 Findings

The study uncovered several vital insights that managers can leverage to improve the user experience in enterprise software.

4.4.1 Diverse User Response Patterns

The research revealed four unique user profiles, each exhibiting distinct psychophysiological responses to usability pain points. For instance, some users showed high cognitive load and strong negative emotions, while others remained largely unaffected by minor disruptions. These profiles suggest that users react differently based on factors like differences in sensitivity to usability disruptions or varied task engagement levels. Recognizing that users react to usability pain points in unique ways highlights the importance of designing adaptable and customizable interfaces. Systems that offer flexibility and adaptability can better accommodate varied user preferences and comfort levels.

4.4.2 Reliable Predictors of Usability Pain Points

The study found that shifts in pupil dilation and visual scanning behaviour (measured by the k-coefficient) were reliable indicators of usability pain points. In particular, increased pupil dilation correlated with higher cognitive load, a clear sign of user difficulty (Sweller et al., 1998; Kucewicz et al., 2018; Van Der Wel & Van Steenbergen, 2018). By monitoring these psychophysiological indicators, managers and UX professionals can identify and address specific usability issues that may go unnoticed with traditional feedback methods. This allows for proactive problem-solving, reducing frustration and enhancing task efficiency.

4.4.3 Moderate Predictive Model Accuracy

The predictive models developed in this study demonstrated moderate success in detecting usability pain points, achieving a precision rate of approximately 48% in identifying usability pain points in natural (unmanipulated) tasks. While not perfect, these models represented a step forward in real-time usability pain point detection. The developed predictive models can be integrated as an additional layer of usability assessment during interaction with the digital enterprise system. Its capability helps detect potential usability issues before they escalate, saving time and improving the user experience.

4.5 Best Practices and Recommendations

Based on the findings, here are some best practices and recommendations for managers and UX professionals seeking to improve the usability of enterprise systems.

4.5.1 Integrate Psychophysiological Data in Usability Assessment

While traditional feedback methods are helpful, they may only partially capture users' real-time experiences (Law et al., 2013). By incorporating psychophysiological data into UX assessments, implicit usability challenges experienced by the users, which they may not articulate themselves, can be disclosed. Using non-invasive tools such as eye tracking, EDA sensors, and facial recognition software during usability assessment, managers and UX professionals can comprehensively view the user's emotional and cognitive states during interaction with the enterprise system (Dirican & Goéktiirk, 2011; Dair et al., 2023).

4.5.2 Employ Predictive Models for Proactive Usability Management

The newly developed predictive models based on psychophysiological data offer an objective way to anticipate and address usability pain points proactively. Integrating these predictive models into usability evaluations allows for real-time identification of usability issues that will enable immediate interventions to improve the user experience.

4.5.3 Develop Adaptive Interfaces to Accommodate Diverse User Profiles

The study highlights that not all users react to usability pain points, similarly, emphasizing the need for flexible and customizable interfaces. This opens an avenue to designing systems with adaptability in mind. Integrating features like task shortcuts and personalized guidance can cater to diverse user profiles, thus reducing the likelihood of frustration when interacting with the digital enterprise system.

4.5.4 Enhance UX Training for Psychophysiological Data Interpretation

To better integrate the use of the psychophysiological data on UX evaluations and to use advanced non-invasive tools effectively, it is essential to train UX teams on interpreting and applying psychophysiological data insights. By offering workshops or training sessions on analyzing psychophysiological data, UX teams can familiarize themselves with these measures, allowing them to extract more meaningful insights and effectively make data-driven design improvements.

4.5.5 Focus on High-Impact Usability Adjustments

The findings revealed that specific metrics, such as pupil dilation (measure for cognitive load) and k-coefficient (measure for visual attention), are exceptionally reliable in identifying usability pain points. In coordination with the system designers and UX professionals, managers should prioritize design adjustments based on these indicators to maximize UX improvements. Use insights from psychophysiological data to make targeted adjustments to the most problematic areas within enterprise systems, focusing on tasks with high cognitive load or intense user engagement.

4.6 Discussion

This research underscored the power of psychophysiological measures to transform the way usability is assessed and managed in enterprise systems. For managers and UX professionals, integrating these advanced psychophysiological measures into usability testing and design processes opens the door to a more accurate, objective understanding of user experiences. By adopting these practices, companies can enhance productivity, reduce employee turnover, and improve satisfaction with enterprise systems.

Furthermore, these findings highlighted the potential for developing adaptive, responsive systems that proactively meet user needs. As enterprise software continues to play a pivotal role in organizational success, the insights from this research offer a competitive edge, positioning companies to optimize their digital tools for maximum usability and impact. Embracing these innovations can provide organizations with a strategic advantage, ensuring that their technology investments yield substantial returns and contribute positively to employee experience and operational efficiency.

Chapter 5

Thesis Conclusion

This research's primary objective was to investigate users' psychophysiological signatures when they encounter usability pain points in digital enterprise environments. Specifically, the study aimed to identify users' distinct psychophysiological response patterns when encountering usability issues (RQ1) and to what extent does these psychophysiological signatures identify usability pain points in another similar task reliably (RQ2). By exploring these questions, the study enhanced our understanding of how psychophysiological measures can be used to offer objective insights into user experiences, particularly in the context of usability assessment.

5.1 Key Findings

The research revealed several significant findings related to the psychophysiological responses to usability pain points. Through cluster analysis, four distinct user profiles were identified, each exhibiting unique patterns of emotional arousal, valence, visual attention, and cognitive load when exposed to usability challenges. The profiles highlighted variability in user responses, suggesting that not all users react uniformly to usability issues. Moreover, the study successfully developed predictive models that utilized these psychophysiological signatures to identify usability pain points. This proved that specific psychophysiological measures particularly shift in pupil dilation and k-coefficient, can serve as reliable indicators of disruptions in user experience. Despite some limitations in model performance, the results showed that the predictive models could find usability pain points with moderate precision, even in tasks that were not directly manipulated with usability pain points, highlighting the predictive models' capabilities in detecting usability challenges.

5.2 Implications for Research and Practice

The findings of this research have significant implications in UX and HCI research. This study challenged the notion that all users respond similarly to usability issues, thus enriching the existing theories in UX research. By demonstrating the diversity in psychophysiological responses, the study opened avenues for deeper investigation into the psychological and situational factors that may influence individual differences in user experiences. From a practical standpoint, the ability to monitor and identify usability pain points in real time using psychophysiological measures provides a powerful tool for system designers and UX researchers. This could lead to improved software design that prioritizes user satisfaction and engagement, ultimately enhancing productivity in enterprise settings. The multimodal approach employed in this study contributes to the methodological toolkit available for UX researchers. By integrating various psychophysiological measures, the research presents a robust framework for capturing and analyzing user responses, paving the way for future studies to explore the complexities of user behaviour in digital environments further.

5.3 Limitations and Future Directions

While this study has made significant strides in understanding psychophysiological responses to usability challenges, it has limitations. Given the moderate precision and recall achieved, future research could explore advanced machine-learning algorithms and signal-processing techniques to enhance model accuracy. Techniques such as deep learning could provide better generalization and higher accuracy in detecting subtle psychophysiological cues. To address the limitations introduced by a single expert for validation, future studies should incorporate multi-rater assessments or alternative validation techniques, reducing subjectivity and increasing reliability. Also, this study was conducted in a controlled laboratory setting, recommending future studies to apply the predictive model in real-world settings, with diverse user groups and enterprise systems, that would help generalize findings and assess the robustness of the model in detecting natural usability pain points under varied conditions. Expanding the scope to include other psychophysiological methods, such as heart rate variability or brainwave patterns, could offer a more comprehensive understanding of users' cognitive and emotional states.

Additionally, integrating direct behavioural or self-reported measures could help validate the inferred fight-flight-freeze-fawn responses, enhancing the interpretability of psychophysiological data. Investigating whether other psychological frameworks align more closely with the observed psychophysiological signatures could provide deeper insights. Expanding the theoretical framework to incorporate a broader range of emotional and cognitive responses could enrich our understanding of user interactions during challenging tasks.

5.4 Practical Recommendations

Furthermore, the study highlighted the significant potential of incorporating psychophysiological data, predictive modelling, and adaptive design strategies to enhance the usability of enterprise systems. By integrating tools such as eye tracking, EDA sensors, and facial recognition software, organizations can gain deeper insights into users' emotional and cognitive responses, uncovering implicit usability challenges. The use of predictive models enables proactive identification and resolution of usability pain points, shifting usability management from reactive to anticipatory. Furthermore, designing adaptive interfaces tailored to diverse user profiles and training UX teams to interpret psychophysiological data ensures targeted and effective system improvements. Prioritizing high-impact usability adjustments based on reliable metrics, such as pupil dilation and the k-coefficient, further optimizes system design and user satisfaction. Collectively, these strategies provide a data-driven approach to usability enhancement, offering practical value for improving user experiences and driving system adoption in enterprise environments.

The study offered practical insights to improve the usability of enterprise systems for different professional roles. By utilizing the predictive models, UX designers and system developers could enhance decision-making based on real-time user insights and address usability issues preemptively. Human Resources and learning development professionals could utilize these models to improve training programs, facilitating employees' acclimatization to new digital tools. Management could leverage these insights as a strategic advancement in the competitive digital market to enhance productivity, satisfaction, and retention, increasing the return on investment. This research advocates for user-centered, efficient, and adaptable enterprise systems while encouraging interdisciplinary collaboration.

In conclusion, this study demonstrated that psychophysiological measures can provide valuable insights into identifying usability pain points that traditional self-reporting methods may overlook. By developing and validating a multimodal approach that combines various physiological signals and assessing different aspects of user's response, this research advanced both the theoretical understanding and practical application of psychophysiology in UX research. While limitations exist, the findings paved the way for future work to refine predictive models and broaden the application of psychophysiological data in real-time usability assessment. Ultimately, this research contributed to a future where user experiences can be optimized with greater precision, transforming enterprise system design and enhancing user satisfaction across digital environments.

Bibliography

Ajenaghughrure, I., Sousa, S., & Lamas, D. (2020). Measuring trust with psychophysiological signals: a systematic mapping study of approaches used. *Multimodal Technologies and Interaction*, 4(3), 63. https://doi.org/10.3390/mti4030063

Alexandros, L., & Michalis, X. (2013). The physiological measurements as a critical indicator in users' experience evaluation. In *Proceedings of the 17th Panhellenic Conference on Informatics* (pp. 258-263).

Alexandrovsky, D., Putze, S., Bonfert, M., Höffner, S., Michelmann, P., Wenig, D., Malaka, R., & Smeddinck, J. D. (2020). Examining Design Choices of Questionnaires in VR User Studies. *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, 1–21. https://doi.org/10.1145/3313831.3376260

Apraiz Iriarte, A., Lasa, G., & Mazmela, M. (2021). Evaluating User Experience with physiological monitoring: A Systematic Literature Review. *Dyna (Bilbao)*, 8, 21. https://doi.org/10.6036/NT10072

Aston-Jones, G., & Cohen, J. D. (2005). AN INTEGRATIVE THEORY OF LOCUS COERULEUS-NOREPINEPHRINE FUNCTION: Adaptive gain and optimal performance. *Annual Review of Neuroscience*, 28(1), 403–450. https://doi.org/10.1146/annurev.neuro.28.061604.135709

Bargas-Avila, J. A., & Hornbæk, K. (2011). Old wine in new bottles or novel challenges: A critical analysis of empirical studies of user experience. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 2689–2698. https://doi.org/10.1145/1978942.1979336

Barreto, A., Zhai, J., & Adjouadi, M. (2007). Non-intrusive physiological monitoring for Automated stress detection in Human-Computer Interaction. In *Springer eBooks* (pp. 29–38). https://doi.org/10.1007/978-3-540-75773-3_4

Bergstrom, J. R., Duda, S., Hawkins, D., & McGill, M. (2014). Physiological response measurements. In *Elsevier eBooks* (pp. 81–108). https://doi.org/10.1016/b978-0-12-408138-3.00004-2

Beri, D., & K, J. R. (2019). Physiological Correlates of Arousal: A Metaanalytic Review. *Journal of Neurology and Neuroscience*, 10(04). https://doi.org/10.36648/2171-6625.10.4.302

Berni, A., Borgianni, Y., Basso, D., & Carbon, C. (2023). Fundamentals and issues of user experience in the process of designing consumer products. *Design Science*, 9. https://doi.org/10.1017/dsj.2023.8

BIOPAC Systems, Inc. (n.d.). AcqKnowledge software [Software]. Goleta, CA.

BIOPAC Systems, Inc. (n.d.). MP-150 [Product brochure or manual]. Goleta, CA.

Borys, M., & Plechawska-Wójcik, M. (2017). Eye-tracking metrics in perception and visual attention research. *EJMT*, 3, 11-23.

Boucsein, W. (2012). Electrodermal activity. *Springer Science & Business Media*.

Bradley, M. M., Miccoli, L., Escrig, M. A., & Lang, P. J. (2008). The pupil as a measure of emotional arousal and autonomic activation. *Psychophysiology*, 45(4), 602–607. https://doi.org/10.1111/j.1469-8986.2008.00654.x

Braithwaite, J. J., Watson, D. G., Jones, R., & Rowe, M. (2013). A guide for analysing electrodermal activity (EDA) & skin conductance responses (SCRs) for psychological experiments. *Psychophysiology*, 49(1), 1017-1034.

Brown, J., Chatterjee, N., Younger, J., & Mackey, S. (2011). Towards a physiology-based measure of pain: patterns of human brain activity distinguish painful from non-painful thermal stimulation. *PLoS ONE*, 6(9), e24124. https://doi.org/10.1371/journal.pone.0024124

Bruun, A., Law, E. L.-C., Heintz, M., & Alkly, L. H. A. (2016). Understanding the Relationship between Frustration and the Severity of Usability Problems: What can Psychophysiological Data (Not) Tell Us? *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, 3975–3987. https://doi.org/10.1145/2858036.2858511

Carmichael, L., Poirier, S.-M., Coursaris, C., Léger, P.-M., & Senecal, S. (2022). Users' Information Disclosure Behaviors during Interactions with Chatbots: The Effect of Information Disclosure Nudges. *Applied Sciences*, 12, 12660. https://doi.org/10.3390/app122412660

Cannon, W. B. (1929). Organization for Physiological Homeostasis. *Physiological Reviews*, 9(3), 399–431. https://doi.org/10.1152/physrev.1929.9.3.399

Chang, L., Gianaros, P., Manuck, S., Krishnan, A., & Wager, T. (2015). A sensitive and specific neural signature for picture-induced negative affect. *Plos Biology*, 13(6), e1002180. https://doi.org/10.1371/journal.pbio.1002180

Charland, P., Léger, P.-M., Senecal, S., Courtemanche, F., Mercier, J., Skelling, Y., & L. LeMoyne, E. (2015). Assessing the Multiple Dimensions of Engagement to Characterize Learning: A Neurophysiological Perspective. *Journal of Visualized Experiments*, 101. https://doi.org/10.3791/52627

Cohen, J. D., McClure, S. M., & Yu, A. J. (2007). Should I stay or should I go? How the human brain manages the trade-off between exploitation and exploration. *Philosophical Transactions of the Royal Society B Biological Sciences*, 362(1481), 933–942. https://doi.org/10.1098/rstb.2007.2098

Costa, I., Silva, W., Damian, A., Rivero, L., Gadelha, B., Teixeira de Oliveira, E., & Conte, T. (2016). *An Empirical Study to Evaluate the Feasibility of a UX and Usability Inspection Technique for Mobile Applications* (p. 599). https://doi.org/10.18293/SEKE2016-127

Dair, Z., Dockray, S., & O'Reilly, R. (2023). Complex Adaptive Systems and Psychophysiological Data - An Exploratory Approach. *2023 31st Irish Conference on Artificial Intelligence and Cognitive Science (AICS)*., 1-4. https://doi.org/10.1109/AICS60730.2023.10470799.

Dirican, A. C., & Göktürk, M. (2011). Psychophysiological measures of human cognitive states applied in human computer interaction. *Procedia Computer Science*, 3, 1361–1367. https://doi.org/10.1016/j.procs.2011.01.016

Ferreira, E., Ferreira, D., Kim, S., Siirtola, P., Röning, J., Forlizzi, J. F., & Dey, A. K. (2014). Assessing real-time cognitive load based on psycho-physiological measures for younger and older adults. 2014 *IEEE Symposium on Computational Intelligence, Cognitive Algorithms, Mind, and Brain* (CCMB), 39–48. https://doi.org/10.1109/CCMB.2014.7020692

Garbas, J.U., Ruf, T., Unfried, M. and Dieckmann, A., (2013). Towards Robust Real-Time Valence Recognition from Facial Expressions for Market Research Applications. *2013 Humaine Association Conference on Affective Computing and Intelligent Interaction.*, 570-575. https://doi.org/10.1109/ACII.2013.100.

Gartner. (n.d.) *Fueling the future of business*. (n.d.). Gartner. https://www.gartner.com/document/4023333?ref=solrAll&refval=420472053

Gilzenrat, M. S., Nieuwenhuis, S., Jepma, M., & Cohen, J. D. (2010). Pupil diameter tracks changes in control state predicted by the adaptive gain theory of locus coeruleus function. *Cognitive Affective & Behavioral Neuroscience*, 10(2), 252–269. https://doi.org/10.3758/cabn.10.2.252

Giroux-Huppé, C., Sénécal, S., Fredette, M., Chen, S. L., Demolin, B., & Léger, P. (2019). Identifying psychophysiological pain points in the online user journey: the case of online grocery. In *Lecture notes in computer science* (pp. 459–473). https://doi.org/10.1007/978-3-030-23535-2_34

Haselmann, T., & Vossen, G. (2011). Software-as-a-Service in Small and Medium Enterprises: An Empirical Attitude assessment. *In Lecture notes in computer science* (pp. 43–56). https://doi.org/10.1007/978-3-642-24434-6_4

Hopstaken, J. F., Van Der Linden, D., Bakker, A. B., & Kompier, M. A. (2015). The window of my eyes: Task disengagement and mental fatigue covary with pupil dynamics. *Biological Psychology*, 110, 100–106. https://doi.org/10.1016/j.biopsycho.2015.06.013

Horwitz, A. G., McCarthy, K., & Sen, S. (2024). A review of the peak-end rule in mental health contexts. *Current Opinion in Psychology*, 101845. https://doi.org/10.1016/j.copsyc.2024.101845

Hudon, A., Demazure, T., Karran, A., Léger, P.-M., & Senecal, S. (2021). *Explainable Artificial Intelligence (XAI): How the Visualization of AI Predictions Affects User Cognitive Load and Confidence* (pp. 237–246). https://doi.org/10.1007/978-3-030-88900-5_27

Huo, F., Zhao, Y., Chai, C., & Fang, F. (2023). A user experience map design method based on emotional quantification of in-vehicle HMI. *Humanities and Social Sciences Communications*, 10. https://doi.org/10.1057/s41599-023-01761-4

Ibarra-Noriega, A. M., Yansane, A., Mullins, J., Simmons, K., Skourtes, N., Holmes, D., White, J., Kalenderian, E., & Walji, M. F. (2024). Evaluating and improving the usability of a mHealth platform to assess postoperative dental pain. *JAMIA Open*, 7(1). https://doi.org/10.1093/jamiaopen/ooae018

Inan Nur, A., B. Santoso, H., & O. Hadi Putra, P. (2021). The Method and Metric of User Experience Evaluation: A Systematic Literature Review. *Proceedings of the 2021 10th International Conference on Software and Computer Applications*, 307–317. https://doi.org/10.1145/3457784.3457832

ISO 9241-210:2019(en), Ergonomics of human-system interaction — *Part 210: Human-centred design for interactive systems.* (n.d.). https://www.iso.org/obp/ui/en/#iso:std:iso:9241:-210:ed-2:v1:en

ISO, W. (1998). 9241-11. Ergonomic requirements for office work with visual display terminals (VDTs). The international organization for standardization, 45(9), 22.

ISO, I. (1999). 13407: Human-centred design processes for interactive systems. Geneva: ISO.

Kahneman, D., Fredrickson, B. L., Schreiber, C. A., & Redelmeier, D. A. (1993). When more pain is preferred to less: adding a better end. *Psychological Science*, 4(6), 401–405. https://doi.org/10.1111/j.1467-9280.1993.tb00589.x

Kim, J., Lee, S., Min, K. S., Jung, H. H., Lee, J. E., Kim, S. J., … & Chang, J. W. (2013). Ventral posterolateral deep brain stimulation treatment for neuropathic pain shortens pain response after cold stimuli. *Journal of Neuroscience Research*, 91(7), 997-1004. https://doi.org/10.1002/jnr.23222

Klaus, H., Rosemann, M. & Gable, G.G. (2000). What is ERP?. *Information Systems Frontiers 2*, 141–162. https://doi.org/10.1023/A:1026543906354

Klotins, E., Unterkalmsteiner, M., & Gorschek, T. (2018). Software engineering in start-up companies: An analysis of 88 experience reports. *Empirical Software Engineering*, 24(1), 68–102. https://doi.org/10.1007/s10664-018-9620-y

Korosec-Serfaty, M., Riedl, R., Senecal, S., & Léger, P.-M. (2022). Attentional and Behavioral Disengagement as Coping Responses to Technostress and Financial Stress: An Experiment Based on Psychophysiological, Perceptual, and Behavioral Data. *Frontiers in Neuroscience*, 16, 883431. https://doi.org/10.3389/fnins.2022.883431

Kozlowska, K., Walker, P., McLean, L., & Carrive, P. (2015). Fear and the defense cascade. *Harvard Review of Psychiatry*, 23(4), 263–287. https://doi.org/10.1097/hrp.0000000000000065

Kreger, A. (2022). *Digital banking user experience: Solve user Pain Points through information architecture*. https://doi.org/10.13140/RG.2.2.34542.08000

Krejtz, K., Duchowski, A., Krejtz, I., Szarkowska, A., & Kopacz, A. (2016). Discerning Ambient/Focal Attention with CoefficientK. *ACM Transactions on Applied Perception*, 13(3), 1–20. https://doi.org/10.1145/2896452

Kucewicz, M. T., Dolezal, J., Kremen, V., Berry, B. M., Miller, L. R., Magee, A. L., Fabian, V., & Worrell, G. A. (2018). Pupil size reflects successful encoding and recall of memory in humans. *Scientific Reports*, 8(1). https://doi.org/10.1038/s41598-018-23197-6

Kwak, D., Ma, X., & Kim, S. (2021). When does social desirability become a problem? Detection and reduction of social desirability bias in information systems research. *Information & Management*, 58(7), 103500. https://doi.org/10.1016/j.im.2021.103500

Lamme, V. A. (2003). Why visual attention and awareness are different. *Trends in Cognitive Sciences*, 7(1), 12–18. https://doi.org/10.1016/s1364-6613(02)00013-x

Law, E. L., Van Schaik, P., & Roto, V. (2013). Attitudes towards user experience (UX) measurement. *International Journal of Human-Computer Studies*, 72(6), 526–541. https://doi.org/10.1016/j.ijhcs.2013.09.006

Léger, P.-M., Senecal, S., Courtemanche, F., Guinea, A., Titah, R., Fredette, M., & L. LeMoyne, E. (2014). Precision is in the Eye of the Beholder: Application of Eye Fixation-Related Potentials to Information Systems Research. *Journal of the Association for Information Systems*, 15. https://doi.org/10.17705/1jais.00376

Li, S. & Deng, W. (2022). Deep Facial Expression Recognition: A Survey. *IEEE Transactions on Affective Computing*, 13(3), 1195-1215. https://doi.org/10.1109/TAFFC.2020.2981446.

Lounis, C. A., Hassoumi, A., Lefrancois, O., Peysakhovich, V., & Causse, M. (2020). Detecting ambient/focal visual attention in professional airline pilots with a modified Coefficient K: a full flight simulator study. *ACM Symposium on Eye Tracking Research and Applications*, 10, 1–6. https://doi.org/10.1145/3379157.3391412

Lovallo, W. (2013). Psychophysiology: Theory and methods. In *Springer eBooks*, 1569–1572. https://doi.org/10.1007/978-1-4419-1005-9_484

Maia, C. L. B., & Furtado, E. S. (2019). An Approach to Analyze User's Emotion in HCI Experiments Using Psychophysiological Measures. *IEEE Access*, 7, 36471–36480. IEEE Access. https://doi.org/10.1109/ACCESS.2019.2904977

Mehrabian, A. (2017). Communication without words. In *Routledge eBooks* (pp. 193–200). https://doi.org/10.4324/9781315080918-15

Microsoft. Dynamics 365 Customer Service. Retrieved from https://dynamics.microsoft.com/en-us/customer-service/

Minzenberg, M. J., Watrous, A. J., Yoon, J. H., Ursu, S., & Carter, C. S. (2008). Modafinil shifts human locus coeruleus to Low-Tonic, High-Phasic activity during functional MRI. *Science*, 322(5908), 1700–1702. https://doi.org/10.1126/science.1164908

Mithun, M. B., Karmakar, S., Varghese, T., Jaiswal, D., Chatterjee, D., Gavas, R. D., Ramakrishnan, R. K., & Pal, A. (2023). Mind Indriya: A System for Simultaneous Assessment of Cognitive Load, Anxiety and Visual Attention. *2023 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, 3234–3240. https://doi.org/10.1109/SMC53992.2023.10394104

Murphy, P. R., O'Connell, R. G., O'Sullivan, M., Robertson, I. H., & Balsters, J. H. (2014). Pupil diameter covaries with BOLD activity in human locus coeruleus. *Human Brain Mapping*, 35(8), 4140–4154. https://doi.org/10.1002/hbm.22466

Nederhof, A. J. (1985). Methods of coping with social desirability bias: A review. *European Journal of Social Psychology*, 15(3), 263–280. https://doi.org/10.1002/ejsp.2420150303

Noldus (2021). FaceReader: Tool for automatic analysis of facial expressions: Version 9 [Software]. Wageningen, The Netherlands: Noldus Information Technology B.V.

Noldus. MediaRecorder - Synchronous video recordings. https://www.noldus.com/mediarecorder-human

Noldus. The Observer XT. Behavioral coding - Event logging software | Behavioral Coding - Event Logging Software | the Observer XT. https://www.noldus.com/observer-xt

Owca, J. (2020). *The association between a Psychotherapist's theoretical orientation and perception of complex trauma and repressed anger in the fawn response* (Order No. 28086275). ProQuest Dissertations & Theses Global Closed Collection. (2447256147). Retrieved from https://login.proxy2.hec.ca/login?url=https://www.proquest.com/dissertations-theses/association-between-psychotherapist-s-theoretical/docview/2447256147/se-2

Parsons, T., Asbee, J., & Courtney, C. (2023). Interaction of Cognitive and Affective Load Within a Virtual City | *2023 IEEE Transactions on Affective Computing*, vol. 14, no. 4, pp. 2768-2775. https://doi.org/10.1109/TAFFC.2022.3220953

Partala, T., & Surakka, V. (2003). Pupil size variation as an indication of affective processing. *International Journal of Human-Computer Studies*, 59(1–2), 185–198. https://doi.org/10.1016/s1071-5819(03)00017-x

Partala, T., & Surakka, V. (2003). Pupil size variation as an indication of affective processing. *International Journal of Human-Computer Studies*, 59(1–2), 185–198. https://doi.org/10.1016/s1071-5819(03)00017-x

Perrig, S. A. C., Aeschbach, L. F., Scharowski, N., von Felten, N., Opwis, K., & Brühlmann, F. (2024). Measurement practices in user experience (UX) research: A systematic quantitative literature review. *Frontiers in Computer Science*, 6. https://doi.org/10.3389/fcomp.2024.1368860

Pirolli, P., & Card, S. (1999). Information foraging. *Psychological Review*, 106(4), 643–675. https://doi.org/10.1037/0033-295x.106.4.643

Pittet, P., & Barthélémy, J. (2015). Experience of formal application ontology development to enhance user understanding in a Geo Business Intelligence SAAS platform. In *Lecture notes in business information processing*, 51–62. https://doi.org/10.1007/978-3-319-21545-7_5

Platzer, D. (2018). Regarding the Pain of Users: Towards a Genealogy of the "Pain Point." *Ethnographic Praxis in Industry Conference Proceedings*, 2018(1), 301–315. https://doi.org/10.1111/1559-8918.2018.01209

Putze, S., Alexandrovsky, D., Putze, F., Höffner, S., Smeddinck, J. D., & Malaka, R. (2020). Breaking The Experience: Effects of Questionnaires in VR User Studies. *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, 1–15. https://doi.org/10.1145/3313831.3376144

Roto, V., & Kaasinen, E. (2008, September). The second international workshop on mobile internet user experience. In *Proceedings of the 10th international conference on Human computer interaction with mobile devices and services*, 571-573. https://doi.org/10.1145/1409240.140935

Russell, J. A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology*, 39(6), 1161–1178. https://doi.org/10.1037/h0077714

Saleh, A., Abuaddous, H., Enaizan, O., & Ghabban, F. (2021). User experience assessment of a covid-19 tracking mobile application (aman) in jordan. Indonesian Journal of Electrical Engineering and Computer Science, 23(2), 1120. https://doi.org/10.11591/ijeecs.v23.i2.pp1120-1127

Salesforce. What is Salesforce? Retrieved from https://www.salesforce.com/products/what-is-salesforce/

ServiceNow. (n.d.). What is ServiceNow? Retrieved from https://www.servicenow.com/what-is-servicenow.html

Shackel, B. (2009). Usability – Context, framework, definition, design and evaluation. *Interacting With Computers*, 21(5–6), 339–346. https://doi.org/10.1016/j.intcom.2009.04.007

Statista. (n.d.). *Software as a Service - Canada | Statista Market forecast.* https://www.statista.com/outlook/tmo/public-cloud/software-as-a-service/canada

Sweller, J., van Merrienboer, J. J. G., & Paas, F. G. W. C. (1998). Cognitive Architecture and Instructional Design. *Educational Psychology Review*, 10(3), 251–296. https://doi.org/10.1023/A:1022193728205

Swoboda, D., Boasen, J., Léger, P., Pourchon, R., & Sénécal, S. (2022). Comparing the Effectiveness of Speech and Physiological Features in Explaining Emotional Responses during Voice User Interface Interactions. *Applied Sciences*, 12(3), 1269. https://doi.org/10.3390/app12031269

Tobii AB (2024). Tobii Pro Lab (Version v 1.217) [Computer software]. Danderyd, Sweden: Tobii AB.

Tobii AB (2024). Tobii Pro Lab User Manual (Version v 1.217). Tobii AB, Danderyd, Sweden.

Van Acker, B. B., Parmentier, D. D., Vlerick, P., & Saldien, J. (2018). Understanding mental workload: from a clarifying concept analysis toward an implementable framework. *Cognition Technology & Work*, 20(3), 351–365. https://doi.org/10.1007/s10111-018-0481-3

Van Der Wel, P., & Van Steenbergen, H. (2018). Pupil dilation as an index of effort in cognitive control tasks: A review. *Psychonomic Bulletin & Review*, 25(6), 2005–2015. https://doi.org/10.3758/s13423-018-1432-y

Vanneste, P., Raes, A., Morton, J., Bombeke, K., Van Acker, B. B., Larmuseau, C., Depaepe, F., & Van Den Noortgate, W. (2020). Towards measuring cognitive load through multimodal physiological data. *Cognition Technology & Work*, 23(3), 567–585. https://doi.org/10.1007/s10111-020-00641-0

Vignaux, M., Léger, P.-M., Charland, P., Salame, Y., Durand, E., Bouillot, N., Pardoen, M., & Senecal, S. (2021). An Exploratory Study on the Impact of Collective Immersion on Learning and Learning Experience. *Multimodal Technologies and Interaction*, 5, 17. https://doi.org/10.3390/mti5040017

Visser, L., Tollenaar, M., Bosch, J., Doornen, L., Haes, H., & Smets, E. (2017). Are psychophysiological arousal and self-reported emotional stress during an oncological consultation related to memory of medical information? an experimental study. Stress, 20(1), 103-111. https://doi.org/10.1080/10253890.2017.1286323

Vrijheid, M., Armstrong, B., Bedard, D., Brown, J., Deltour, I., Iavarone, I., Krewski, D., Lagorio, S., Moore, S., Richardson, L., Giles, G., McBride, M., Parent, M.-E., Siemiatycki, J., & Cardis, E. (2008). Recall bias in the assessment of exposure to mobile phones. *Journal of Exposure Science & Environmental Epidemiology*, 19, 369–381. https://doi.org/10.1038/jes.2008.27

Wager, T., Atlas, L., Lindquist, M., Roy, M., Woo, C., & Kross, E. (2013). An fmri-based neurologic signature of physical pain. New England Journal of Medicine, 368(15), 1388-1397. https://doi.org/10.1056/nejmoa1204471

Wals, S. F., & Wichary, S. (2022). Under Pressure: Cognitive Effort During Website-Based Task Performance is Associated with Pupil Size, Visual Exploration, and Users' Intention to Recommend. *International Journal of Human-Computer Interaction*, 39(18), 3504–3515. https://doi.org/10.1080/10447318.2022.2098576

Yang, Y. & Sun, Y.(2017). *Facial expression recognition based on Arousal-Valence Emotion Model and Deep Learning method*. IEEE Conference Publication | IEEE Xplore. https://ieeexplore.ieee.org/document/8789024

Yusuf, M., Silas, F. A., & Haruna, S. (2018). Implementing Personnel Management System as SaaS. *Circulation in Computer Science*, 3(5), 1-6. https://doi.org/10.22632/ccs-2018-252-86

Appendices

Appendix A – Task Instructions

Task 1 instructions for both Microsoft Dynamics and Salesforce

Tâche 1:

M. John Smith est un nouveau client de l'entreprise. **Votre tâche consiste à ajouter M. Smith au système en tant que contact client.**

Vous trouverez les informations concernant M. Smith sur l'iPad qui vous a été fourni. Utilisez ces informations pour compléter le dossier de contact du client.

Il est fortement recommandé de suivre la procédure standard de création d'un contact, comme indiqué dans la courte vidéo d'accueil. Il est bon de noter que le système a été conçu pour accomplir la tâche par des voies alternatives.

Veuillez informer le modérateur lorsque vous avez fini de lire les instructions ci-dessus.

Veuillez attendre les instructions suivantes.

Task 2 instructions for Salesforce

Tâche 2:

Vous avez reçu un courriel de Mme Jane Doe concernant une demande de renseignements sur la machine à café qu'elle vient d'acheter. La machine à café n'a pas pu s'allumer. Mme Doe cherche à obtenir de l'aide pour résoudre le problème. **Votre tâche consiste à créer un dossier requêtes pour Mme Jane Doe.**

Vous trouverez les informations concernant le cas de Mme Doe sur l'iPad qui vous a été fourni. Utilisez ces informations pour compléter le dossier requêtes du client.

Il est fortement recommandé de suivre la procédure opérationnelle standard pour création d'un dossier, comme indiqué dans la courte vidéo d'introduction. Il est bon de noter que le système a été conçu pour accomplir la tâche par des voies alternatives.

Veuillez informer le modérateur lorsque vous avez fini de lire les instructions ci-dessus.

Veuillez attendre les instructions suivantes.

Task 2 instructions for Microsoft Dynamics

Tâche 2:

Vous avez reçu un courriel de Mme Marlène Dumoulin concernant une demande de renseignements sur la machine à café qu'elle vient d'acheter. La machine à café n'a pas pu s'allumer. Mme Doe cherche à obtenir de l'aide pour résoudre le problème. **Votre tâche consiste à créer un dossier requêtes pour Mme Marlène Dumoulin.**

Vous trouverez les informations concernant le cas de Mme Dumoulin sur l'iPad qui vous a été fourni. Utilisez ces informations pour compléter le dossier requêtes du client.

Il est fortement recommandé de suivre la procédure opérationnelle standard pour création d'un dossier, comme indiqué dans la courte vidéo d'introduction. Il est bon de noter que le système a été conçu pour accomplir la tâche par des voies alternatives.

Veuillez informer le modérateur lorsque vous avez fini de lire les instructions ci-dessus.

Veuillez attendre les instructions suivantes.

Task 3 instructions for ServiceNow

Tâche 3:

Vous avez reçu un courriel d'une cliente, M. Hector Currie, concernant une demande de mise à jour du logiciel de son ordinateur. Le logiciel est obsolète et a besoin d'une mise à jour. M. Currie cherche de l'aide pour résoudre le problème. **Votre tâche consiste à créer un dossier pour M. Currie.**

Vous trouverez les informations relatives au dossier de M. Currie sur l'iPad qui vous a été fourni. Utilisez ces informations pour compléter le dossier cas du client.

Il est fortement recommandé de suivre la procédure opérationnelle standard pour créer un dossier, comme indiqué dans la courte vidéo d'introduction. Il est bon de noter que le système a été conçu pour accomplir cette tâche par d'autres voies.

Veuillez informer le modérateur lorsque vous avez fini de lire les instructions ci-dessus.

Veuillez attendre les instructions suivantes.

Appendix B – Short Training Video before each Task

The participants are shown a short training video of the enterprise system they are assigned to before they perform the task.

Microsoft Dynamics Training (French): https://youtu.be/cfYfzPnwUk4

Salesforce Training Video (French): https://youtu.be/PXLuaznoo-A

ServiceNow Training Video (French): https://youtu.be/2Auz7ysTSuM

Appendix C – Required information to accomplish the task

The training videos (Appendix B) and the information required (below) to accomplish the task are shown on an iPad.

Group A (Task 1 & Task 2 – Microsoft Dynamics; Task 3 – ServiceNow):
https://sway.cloud.microsoft/KWgfEQ8FANGSTJXD?ref=Link&loc=play

Group B (Task 1 & Task 2 – Salesforce; Task 3 – ServiceNow):
https://sway.cloud.microsoft/82mHEGEnSUIZZsoO?ref=Link&loc=play

Information required for Task 1 for both Microsoft Dynamics & Salesforce

TÂCHE 1: CRÉER UN CONTACT

Veuillez utiliser les informations ci-dessous pour créer un dossier de contact client.

Formule d'appel (le cas échéant) : **Mr**

Prénom : **John**

Nom de famille : **Smith**

Nom du compte : **Café Rouge**

Numéro de Téléphone mobile: **902-330-3388**

Courrier électronique : **johnsmith@gmail.com**

Adresse : **3280 Rue Goyer, Montréal, QC H3S 1J1**

Lorsque vous avez terminé la tâche, veuillez en informer le modérateur.

Information required for Task 2 for Microsoft Dynamics

TÂCHE 2: CRÉER UN INCIDENTS

Veuillez utiliser les informations ci-dessous pour créer un dossier incidents client.

- Nom du client : **Marlène Dumoulin**
- Titre de l'affaire : **Product Malfunction**
- Type de incidents : **Problème**
- Origine**: E-mail**
- Sujet : **Général**
- Produit : **Café Duo**
- Description : **La machine à café ne s'allume pas**

Lorsque vous avez terminé la tâche, veuillez en informer le modérateur.

Information required for Task 2 for Salesforce

TÂCHE 2: CRÉER UNE REQUÊTE

Veuillez utiliser les informations ci-dessous pour créer un dossier requête client.

- Type d'enregistrement : **RFI**
- Nom du client : **Jane Doe**
- Compte : **Coffee Lab**
- Statut : **New**
- Origine de la requête : **Email**
- Type : **Problem**
- Motif de la requête: **New problem**
- Priorité : **Medium**
- Objet : **Alimentation de la machine à café**
- Description : **La machine à café n'arrive pas à se mettre en marche**

Lorsque vous avez terminé la tâche, veuillez en informer le modérateur.

Information required for Task 3 for ServiceNow

TÂCHE 3: CRÉER UN CAS (SERVICENOW)

Veuillez utiliser les informations ci-dessous pour créer un dossier cas client.

- Type d'affaire : **Commander**
- Canal : **Email**
- Entreprise : **Golddex**
- Contact : **Hector Currie**
- Priorité : **Moderate**
- Affecté à : **Jamie Erwin**
- Courte description : **Demande d'un nouveau logiciel ou d'une mise à jour**

Appendix D – Expert Evaluation Protocol

	PROTOCOL for INTER-RATER RELIABILITY
Step 1	Define the observation time frame. 10s before the PP marker time 10s after the PP marker time *This will be the time frame where you begin your observation and end the observation
Step 2	Watch the video and assess if the user is experiencing a pain point or not. Scale assessment from 0-3. **Pain Point Assessment** **0: Absolutely not a Pain Point** **1: Somewhat a Pain Point** **2: Most likely a Pain Point** **3: Absolutely a Pain Point**
Step 3	**Type of Pain Point** Identify what type of pain point the user is experiencing in the video. If it is a Pain Point, what type of Pain Point do you think it was? **0: Not a Pain Point** **1: PP1: Actively searching, trying to understand** **2: PP2: Frustrated, not making effort** **3: PP1 & PP2: mix of actively searching and frustration** **4: PP1.a: Actively searching only** **5: PP1.b: Trying to understand** **6: PP2.a: Frustrated** **7: PP2.b: not making an effort**
Step 4	**Saccade/Eye Jumps** Rewatch the video, but this time you will be observing the saccade/eye jumps exhibited by the user within the observation time frame. Does the participant show erratic eye jumps that deviates from the group of saccades within the observation time? **0: Smooth:** No noticeable eye jumps beyond typical saccades (small, rapid eye movements between fixation points). **1: Occasional:** Participant shows occasional, brief eye jumps that deviate slightly from the group's saccade patterns. (1 to 3 long eye jumps) **2: Frequent:** Participant exhibits frequent and/or sustained eye jumps that significantly deviate from the group's saccade patterns. (4 or more) Note: Only count eye jumps (long) that deviates from the group of saccades. Please note that the participant may not show eye data because they were looking at the iPad. Also, disregard eye jumps when interface switches from one page to another.
Step 5	**Scan Path Pattern** Rewatch the video, but this time you will be observing the scan path pattern exhibited by the user. within the observation time frame. Does the participant's scan path pattern shows logical progression as they explore the screen within the observation time? **0: Highly Systematic:** Participant's scan path follows a clear and logical progression, efficiently moving between relevant areas of the screen in a way that suggests a strategic exploration. **1: Somewhat Orderly:** Participant shows some attempt at logical progression, but there might be backtracks, revisits, or inefficient jumps between areas of interest **2: Random:** Participant's eye movements show no clear order or progression as they explore the screen.

Step 6	**Regressions** Rewatch the video, but this time you will be observing the regressions exhibited by the user within the observation time frame. Does the participant look back at the previously visited area, button, or field? **0: Minimal Regressions:** Participant shows very few instances of revisiting previously fixated areas. Their scan path progresses smoothly with minimal backtracking (0 to 1 times). **1: Moderate Regressions:** Participant exhibits some regressions, revisiting previously fixated areas occasionally. (2 to 3 times) **2: Excessive Regressions:** Participant frequently revisits previously fixated areas. Their scan path shows a lot of backtracking and may appear scattered. (4 or more)
Step 7	**Fixations** Rewatch the video, but this time you will be observing the fixations exhibited by the user within the observation time frame. Did the participant showed longer fixation time on different elements during the obeservation timeline? **0: Uniform Fixation:** Participant showed similar fixation times on all elements (short fixations), or there were no clear differences in fixation durations between elements. (0-1 long fixation) **1: Varied Fixation:** Participant exhibited some variation in fixation times between elements. There might be a few elements that stood out with slightly longer or shorter fixations compared to the rest. (2-3 long fixations) **2: Distinct Fixation:** Participant showed clear and significant differences in fixation times between elements. Some elements held the participant's attention for much longer than others. (4 or more)
Step 8	**Hesitations** Rewatch the video, but this time you will be observing the hesitation exhibited by the user within the observation time frame. Pay attention to the cursor movement. Does it hover over an element for an unusually long time before clicking? **0: No Unusual Hovers:** The cursor hovers over elements for a typical amount of time before clicking. There's no indication of hesitation or prolonged focus on specific elements. **1: Occasional Long Hovers:** The cursor occasionally hovers over an element for a slightly longer duration than usual before clicking. This might happen once or twice during the observation period. (1 or 2 times) **2: Frequent Long Hovers:** The cursor frequently hovers over elements for an unusually long time before clicking. (3 or more) Note: Disregard cursor hovers if the participant is typing, reading, or scrolling. Also, if there is no eye-tracking data on the screen, the participant might be looking at the iPad, do not count this as "hovers". Make your own judgement if hovering is hesitating to click.

Step 9	**Cursor Movement** Rewatch the video, but this time you will be observing the cursor movement exhibited by the user within the observation time frame. Does the cursor move back and forth repeatedly between sections of the screen or elements with no click event? **0: No Back-and-Forth:** The cursor moves purposefully between sections of the screen or elements without repeated back-and-forth movements. Clicks likely happen after reaching the intended section. **1: Occasional Back-and-Forth:** The cursor occasionally makes brief back-and-forth movements between sections or elements before settling or clicking. (1 or 2 times) **2: Frequent Back-and-Forth:** The cursor frequently moves back and forth repeatedly between sections of the screen or elements without clicking. (3 or more)

Step 10	**Missed Clicks** Rewatch the video, but this time you will be observing the missed clicks exhibited by the user within the observation time frame. Did the participant need to reclick an item because it is not what they intented to click? **0: No Reclicks:** The participant clicked on items only once and did not need to reclick anything to achieve their intended action. **1: Single Reclick:** The participant reclicked on an item once, possibly due to a minor misclick or needing to confirm their selection. (1 reclick) **2: Multiple Reclicks:** The participant reclicked on items multiple times, suggesting significant difficulty selecting the intended target. (2 or more reclicks) NOTE: This click action refers to the participant going back and correcting their previous click to achieve their intended action. If it is not the case, please see next column "Re-evaluation Click".
Step 11	**Re-evaluation Clicks** Rewatch the video, but this time you will be observing the re-evaluation clicks exhibited by the user within the observation time frame. Did the participant clicked an element (button, field, etc), and clicked a different element, then went back and reclicked the initial element clicked during the observation time period? **0: No Re-evaluation Clicks: T**he participant clicks an element and completes the action without revisiting the same element. **1: Single Re-evaluation Click:** The participant clicks an element, then clicks a different element, but returns to the initial element with a single re-click. **2: Multiple Re-evaluation Clicks:** The participant clicks an element, then clicks a different element, and then reclicks the initial element multiple times (more than once) within the observation period. NOTE: For this category, element refers to buttons, text box field, drop down menus, etc)
Step 12	**Audio Reactions** Rewatch the video, but this time you will be observing the audio reactions exhibited by the user within the observation time frame. Did the participant produce audio reactions while performing the task? **0: No Audible Reactions:** The participant remains silent throughout the task, producing no audible sounds like sighs, grunts, or verbalizations. **1: Occasional Vocalizations:** The participant occasionally produces brief, quiet vocalizations like sighs, soft murmurs, or single words. (1 or 2 audio) **2: Frequent Vocalizations:** The participant frequently produces vocalizations that might be louder, more sustained, or involve clear words or phrases. (3 or more)

Appendix E – Ethics Certificate

HEC MONTRÉAL

Comité d'éthique de la recherche

CERTIFICAT D'APPROBATION ÉTHIQUE

La présente atteste que le projet de recherche décrit ci-dessous a fait l'objet d'une évaluation en matière d'éthique de la recherche avec des êtres humains et qu'il satisfait aux exigences de notre politique en cette matière.

Projet # : 2024-5933

Titre du projet de recherche : Étude sur la découvrabilité et la complexité des applications SaaS d'entreprise

Chercheur principal : Pierre-Majorique Léger, Professeur titulaire, Technologies de l'information, HEC Montréal
Cochercheurs : Sylvain Sénécal; Constantinos K. Coursaris; Marc Fredette; Frédérique Bouvier; Luis Carlos Castiblanco; Juan Fernandez Shaw; David Brieugne; Salima Tazi; Xavier Côté; Théophile Demazure; Élise Imbeault; Andrada Toma; Kelvin Jacinto

Date d'approbation du projet : 08 mai 2024
Date d'entrée en vigueur du certificat : 08 mai 2024
Date d'échéance du certificat : 01 mai 2025

Maurice Lemelin
Président
CER de HEC Montréal

Signé le 2024-05-08 à 16:30