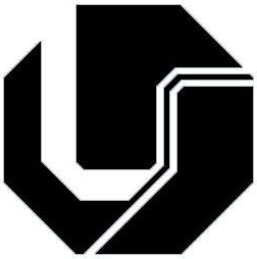
FEDERAL UNIVERSITY OF UBERLÂNDIA

TECHNOLOGICAL UNIVERSITY OF THE SHANNON



DÉBORA PEREIRA SALGADO

A FRAMEWORK FOR EVALUATION OF WHEELCHAIR SIMULATOR-BASED MOBILITY TECHNOLOGIES

THESIS IS SUBMITTED FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

SEPTEMBER 2025

DÉBORA PEREIRA SALGADO

A FRAMEWORK FOR EVALUATION OF WHEELCHAIR SIMULATOR-BASED MOBILITY TECHNOLOGIES

Doctoral thesis submitted to the Post-Graduate Program at the Federal University of Uberlândia and Technological University of the Shannon as part of the requirements for the degree of Doctor of Philosophy.

Supervisors:

Prof. Dr. Eduardo Lázaro Martins Naves

Dr. Sheila Fallon

Dr. Yuansong Qiao

UBERLÂNDIA, MG

SEPTEMBER 2025

Ficha Catalográfica Online do Sistema de Bibliotecas da UFU
com dados informados pelo(a) próprio(a) autor(a).

S164 Salgado, Débora Pereira, 1992-
2025 A Framework for Evaluation of Wheelchair Simulator-based Mobility Technologies [recurso eletrônico] : N/A / Débora Pereira Salgado. - 2025.

Orientadora: Eduardo Lázaro Martins Naves.
Coorientadora: Sheila Fallon .
Coorientadora: Yuansong Qiao.
Tese (Doutorado) - Universidade Federal de Uberlândia, Doutorado em Engenharia Biomédica.
Modo de acesso: Internet.
DOI http://doi.org/10.14393/ufu.te.2025.512
Inclui bibliografia.
Inclui ilustrações.

1. Engenharia biomédica. I. Naves, Eduardo Lázaro Martins, 1970-, (Orient.). II. , Sheila Fallon,1974-, (Coorient.). III. Qiao, Yuansong ,1972-, (Coorient.). IV. Universidade Federal de Uberlândia. Doutorado em Engenharia Biomédica. V. Título.

CDU: 62:61

Bibliotecários responsáveis pela estrutura de acordo com o AACR2:
Gizele Cristine Nunes do Couto - CRB6/2091
Nelson Marcos Ferreira - CRB6/3074

UNIVERSIDADE FEDERAL DE UBERLÂNDIA
Coordenação do Programa de Pós-Graduação em Engenharia Biomédica
Av. João Naves de Ávila, 2121, Bloco 3N, Sala 115 - Bairro Santa Mônica, Uberlândia-MG, CEP 38400-902
Telefone: (34) 3239-4761 - www.ppgeb.feelt.ufu.br - ppegb@feelt.ufu.br

ATA DE DEFESA - PÓS-GRADUAÇÃO

Programa de Pós-Graduação em:	Engenharia Biomédica				
Defesa de:	Tese, 011, PPGEB				
Data:	vinte de agosto de dois mil e vinte e cinco	Hora de início:	07:00	Hora de encerramento:	08:45
Matrícula do Discente:	12223EBI003				
Nome do Discente:	Débora Pereira Salgado				
Título do Trabalho:	A Framework for Evaluation of Wheelchair Simulator-Based Mobility Technologies				
Área de concentração:	Engenharia Biomédica				
Linha de pesquisa:	Engenharia de Reabilitação e Biomecânica				
Projeto de Pesquisa de vinculação:	Desenvolvimento e teste de estratégias telerreabilitadoras para usuários de cadeiras de rodas com comprometimento de ombro para a plataforma Reabnet				

Reuniu-se via plataforma Videoconferência Teams, a Banca Examinadora, designada pelo Colegiado do Programa de Pós-graduação em Engenharia Biomédica, assim composta: Professores Doutores: Sridhar Krishnan - TMU/Canada; Ludymila Ribeiro Borges - University of Calgary/Canada; Rhaíra Helena Caetano e Souza - IFB/Brasília e PPGEB/UFU; Camille Marques Alves - PPGEB/UFU; Eduardo Lázaro Martins Naves - PPGEB/UFU orientador da candidata.

Iniciando os trabalhos o presidente da mesa, Dr. Eduardo Lázaro Martins Naves, apresentou a Comissão Examinadora e a candidata, agradeceu a presença do público, e concedeu a Discente a palavra para a exposição do seu trabalho. A duração da apresentação da Discente e o tempo de arguição e resposta foram conforme as normas do Programa.

A seguir o senhor presidente concedeu a palavra, pela ordem sucessivamente, aos examinadores, que passaram a arguir o candidato. Ultimada a arguição, que se desenvolveu dentro dos termos regimentais, a Banca, em sessão secreta, atribuiu o resultado final, considerando o candidato:

Aprovada.

Esta defesa faz parte dos requisitos necessários à obtenção do título de Doutora.

O competente diploma será expedido após cumprimento dos demais requisitos, conforme as normas do Programa, a legislação pertinente e a regulamentação interna da UFU.

Nada mais havendo a tratar foram encerrados os trabalhos. Foi lavrada a presente ata que após lida e achada conforme foi assinada pela Banca Examinadora.

Documento assinado eletronicamente por **Ludymila Ribeiro Borges**, **Usuário Externo**, em 20/08/2025, às 10:06, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do Decreto nº 8.539, de 8 de outubro de 2015.

Documento assinado eletronicamente por **Camille Marques Alves**, **Usuário Externo**, em 20/08/2025, às 10:12, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do Decreto nº 8.539, de 8 de outubro de 2015.

Documento assinado eletronicamente por **Eduardo Lazaro Martins Naves**, **Professor(a) do Magistério Superior**, em 20/08/2025, às 10:25, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do Decreto nº 8.539, de 8 de outubro de 2015.

Documento assinado eletronicamente por **Sridhar Krishnan**, **Usuário Externo**, em 20/08/2025, às 10:29, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do Decreto nº 8.539, de 8 de outubro de 2015.

Documento assinado eletronicamente por **Rhaíra Helena Caetano e Souza**, **Usuário Externo**, em 22/08/2025, às 09:59, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do Decreto nº 8.539, de 8 de outubro de 2015.

A autenticidade deste documento pode ser conferida no site https://www.sei.ufu.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0, informando o código verificador **6606478** e o código CRC **4445EC5A**.

Referência: Processo nº 23117.057330/2025-21 SEI nº 6606478

To God and my family

“Whatever you do, work at it with all your heart as working for the Lord, not for human masters.”

– Colossians 3:23

“Tudo o que fizerem, façam de todo o coração, como para o Senhor, e não para os homens.”

– Colossenses 3:23

ABSTRACT

Wheelchair simulators offer significant potential to support assessment and training in the power mobility provision. However, their adoption in clinical settings remains limited, largely due to the absence of standardised protocols for meaningfully evaluating user-simulator interaction in a safe and adaptable manner. Given the wide variability in users' motor and cognitive abilities, there is a clear need for a structured framework that defines how simulator-based sessions are delivered and how user responses are assessed.

This thesis presents a multidimensional evaluation framework, grounded in Quality of Experience (QoE) and Cognitive Load Theory (CLT), to assess user interaction with a virtual wheelchair simulator. The framework integrates subjective feedback (e.g., usability, emotion, cognitive workload, sense of presence), physiological signals (e.g., cardiac and electrodermal activity), and behavioural data (e.g., head and wrist movement) to capture users' quality of experiences and tolerances (e.g., cybersickness) during simulator use. Two lab-based studies and one field-based pilot study were conducted.

The first lab study compared immersive (Headset-1, N = 17) and non-immersive (Desktop, N = 24) conditions. Results revealed significant differences in workload (NASA-TLX, $p < .01$), presence (IPQ, $p < .01$), and heart rate variability change from baseline to first collision ($p < .01$, $r = 0.43$), with immersive use also eliciting higher cybersickness symptoms. These findings illustrate the trade-off between realism and tolerance.

The second lab study expanded on the first (Headset-2 group, N=16) tested a smoother motion profile (low jerk). Symptoms decreased (SSQ-oculomotor, $p < .05$), but not enough to justify continued headset use in clinical settings. The most notable finding was the combined effect of display type and motion dynamics ($p < .05$), influencing usability (SUS) and sense of presence (IPQ).

The final field-based pilot study (N = 20; 10 wheelchair users, 10 controls) examined the framework's feasibility and clinical relevance in Irish Wheelchair Association (IWA) centres. Simulator metrics distinguished groups, with collisions ($p < .01$, $r = 0.74$) and trajectory deviation (RMSE, $p < .01$, $r = 0.85$) higher among users. Cognitive functioning (MoCA) correlated negatively with collisions ($\rho = –.73$, $p < .05$) and RMSE ($\rho = –.74$, $p = .05$). Heart rate analysis showed a significant group difference in HR change ($p = .003$, $r = 0.67$), with users maintaining elevated HR while controls showed reductions. User feedback further guided refinement of the evaluation framework.

Together, these studies demonstrate the feasibility and potential clinical value of wheelchair simulators for delivering tailored, repeatable, and safe assessments. The resulting QoE-based

framework, EMPOWER-SIM, is presented as a set of preliminary, feasibility-informed guidelines for clinical integration, offering a foundation for future pilot and validation studies in power mobility assessment and training.

RESUMO

Simuladores de cadeira de rodas oferecem um potencial significativo para apoiar a avaliação e o treinamento na provisão de mobilidade motorizada. No entanto, sua adoção em contextos clínicos ainda é limitada, em grande parte devido à ausência de protocolos padronizados para avaliar de forma significativa a interação usuário–simulador de maneira segura e adaptável. Dada a ampla variabilidade nas habilidades motoras e cognitivas dos usuários, há uma clara necessidade de um framework estruturado que defina como as sessões baseadas em simuladores devem ser conduzidas e como as respostas dos usuários devem ser avaliadas.

Esta tese apresenta um framework de avaliação multidimensional, fundamentado em Qualidade da Experiência (QoE) e na Teoria da Carga Cognitiva (CLT), para analisar a interação do usuário com um simulador virtual de cadeira de rodas. O framework integra feedback subjetivo (usabilidade, emoção, carga cognitiva, senso de presença), sinais fisiológicos (atividade cardíaca e eletrodérmica) e dados comportamentais (movimentos da cabeça e do punho) para capturar a qualidade da experiência e a tolerância dos usuários (por exemplo, ciberdoença) durante o uso do simulador. Foram conduzidos dois estudos laboratoriais e um estudo piloto de campo.

O primeiro estudo comparou condições imersivas (Headset-1, N = 17) e não imersivas (Desktop, N = 24). Os resultados revelaram diferenças significativas na carga de trabalho (NASA-TLX, $p < .01$), presença (IPQ, $p < .01$) e variação da variabilidade da frequência cardíaca entre o baseline e a primeira colisão ($p < .01$, $r = 0.43$), sendo que o uso imersivo também provocou sintomas mais elevados de ciberdoença. Esses achados ilustram o equilíbrio necessário entre realismo e tolerância.

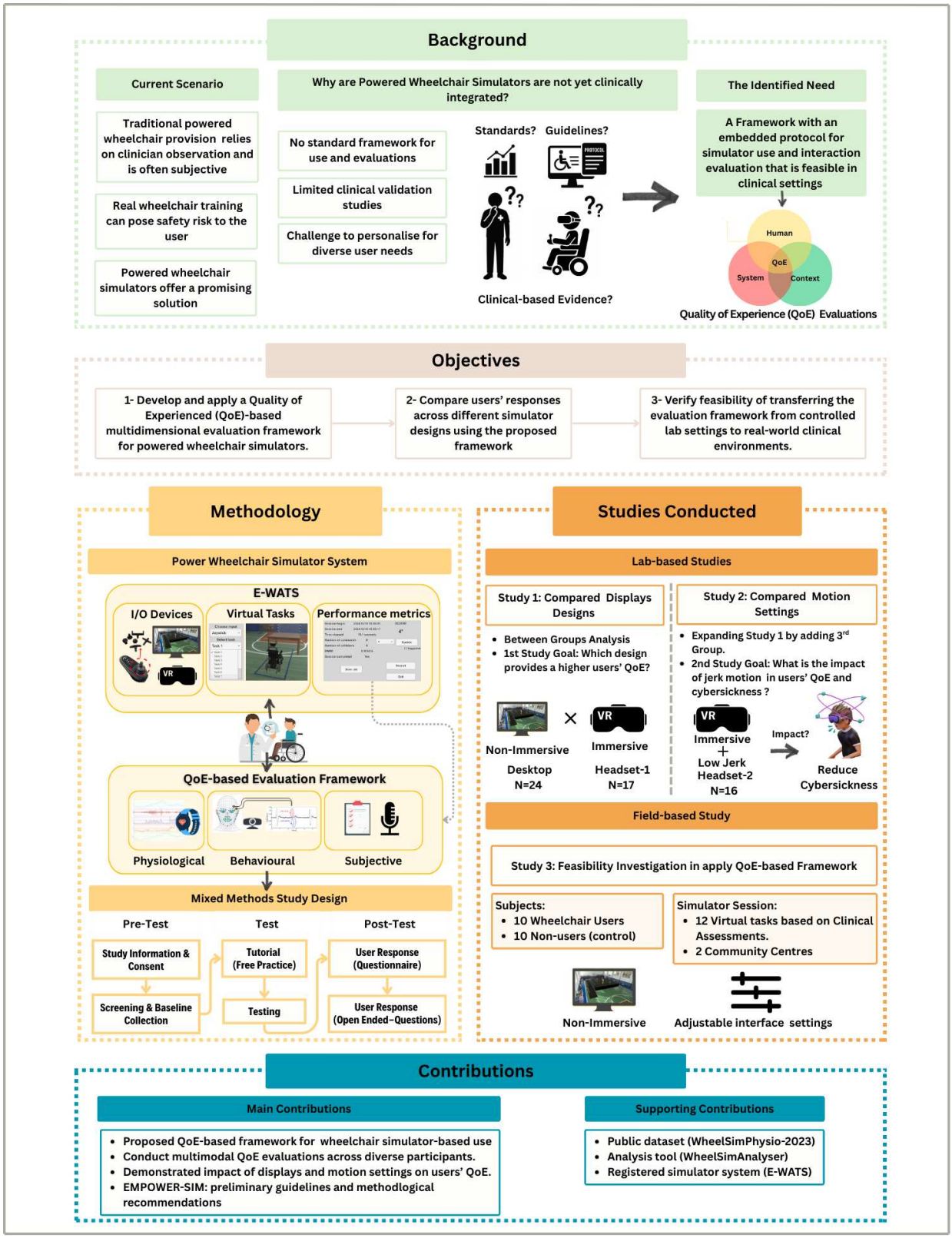
O segundo estudo expandiu o primeiro (Headset-2, N = 16), testando um perfil de movimento mais suave (baixo jerk). Os sintomas diminuíram (SSQ–oculomotor, $p < .05$), mas não o suficiente para justificar o uso contínuo de headsets em contextos clínicos. O achado mais relevante foi o efeito combinado de tipo de display × dinâmica de movimento ($p < .05$), influenciando usabilidade (SUS) e senso de presença (IPQ).

O estudo piloto de campo (N = 20; 10 usuários de cadeira de rodas, 10 controles) examinou a viabilidade e relevância clínica do framework em centros da Irish Wheelchair Association (IWA). As métricas do simulador distinguiram os grupos, com colisões ($p < .01$, $r = 0.74$) e desvio de trajetória (RMSE, $p < .01$, $r = 0.85$) mais elevados entre usuários. O funcionamento cognitivo (MoCA) correlacionou-se negativamente com colisões ($\rho = –.73$, $p < .05$) e RMSE ($\rho = –.74$, $p = .05$). A análise da frequência cardíaca (FC) mostrou uma diferença significativa entre grupos em mudança de FC ($p =$

.003, r = 0.67), com usuários mantendo FC elevada enquanto controles apresentaram reduções. O feedback dos usuários também orientou o refinamento do framework de avaliação.

Em conjunto, esses estudos demonstram a viabilidade e o valor clínico preliminar dos simuladores de cadeira de rodas para oferecer avaliações seguras, repetíveis e adaptadas. O framework baseado em QoE, EMPOWER-SIM, é apresentado como um conjunto de diretrizes preliminares, informadas pela viabilidade, para integração clínica, oferecendo uma base para futuros estudos piloto e de validação no contexto de avaliação e treinamento em mobilidade motorizada.

GRAPHICAL ABSTRACT



ACKNOWLEDGEMENTS

I extend my sincere gratitude to my supervisors, Dr. Sheila Fallon, Dr. Yuansong Qiao, and Prof. Eduardo Lázaro Martins Naves, for their invaluable guidance, encouragement, and constructive feedback throughout the development of this work.

I would also like to express my heartfelt thanks to Dr. Enda Fallon for his support and mentorship during key stages of my academic journey.

This research was made possible through the collaboration between Technological University of the Shannon (TUS) and the Federal University of Uberlândia (UFU). I am deeply grateful to both institutions for providing the environment, resources, and opportunities that enabled this work to take shape.

I am especially thankful to the Irish Wheelchair Association, in particular Paul Ryan, for their outstanding support and warm welcome during my fieldwork. I am truly grateful to all the staff and community members who engaged with the study, your openness, time, and insights have been invaluable to this research.

Finally, I wish to thank all friends, and family for their unwavering support, both within the scope of the project and beyond. Their belief in me, encouragement during difficult moments, and continuous presence were essential in helping me reach this milestone. I could not have completed this journey without them.

CONTENTS

List of Tables

List of Figures

List of Abbreviations and Acronyms

ABI – Acquired Brain Injury

ACC - Accelerometer

ACR – Absolute Category Rating

AI – Artificial Intelligence

ALS – Amyotrophic Lateral Sclerosis

ANS – Autonomic Nervous System

AT – Assistive Technology

AUC – Area Under the Curve

BVP – Blood Volume Pressure

CASP – Cognitive Assessment for Stroke Patients

CCR – Comparison Category Rating

CNS – Central Nervous System

CSO – Central Statistics Office

CIF – Context Influence Factors

CX – Customer Experience

DCR – Degradation Category Rating

EDA – Electrodermal Activity

EEG – Electroencephalogram

EOG – Electrooculogram

E-WATS – Electric Wheelchair Assessment and Training Simulator

FAIR – Findable, Accessible, Interoperable, Reusable

FERS – Functional Evaluation Rating Scale

fNIRS – functional Near-Infrared Spectroscopy

HCI – Human Computer Interaction

HIF – Human Influence Factors

HMD – Head Mounted Display

HR – Heart Rate

HRV – Heart Rate Variability

IBGE – Brazilian Institute of Geography and Statistics

IBI – Inter-beat Intervals

ICF – International Classification of Functioning, Disability and Health

IF – Influence Factors

IPQ – Igroup Presence Questionnaire

IQR – Interquartile Range

ITU-T – International Telecommunication Union Telecommunication

IWA – Irish Wheelchair Association

LSL – Lab Streaming Layer

MAE – Mean Absolute Error

MoCA – Montreal Cognitive Assessment

MOS – Mean Opinion Score

MS – Multiple Sclerosis

NASA-TLX – NASA Task Load Index Assessment

OSF – Open Science Framework

PCDA – Power Community Driving Assessment

PIDA – Power Mobility Indoor Driving Assessment

PMD – Power Mobility Device

PMRT – Power Mobility Road Test

PMFET – Power Mobility Functional Evaluation Tasks

PoMoDATT – Power Mobility Device Assessment Training Tool

PMTT – Power Mobility Training

PNS – Parasympathetic Nervous System

PWC – Power Wheelchair

QoE – Quality of Experience

QoS – Quality of Service

RA – Rheumatoid Arthritis

ROC – Receiver Operating Characteristics

RBANS – Repeatable Battery for the Assessment of Neuropsychological Status

REC – Research Ethics Committee

RMSE – Root Mean Square Error

RTC – Randomized Controlled Trials

SAM – Self-Assessment Manikin

SIF – System Influence Factors

SNS – Sympathetic Nervous System

SSQ – Simulator Sickness Questionnaire

STARD – Standards for Reporting Diagnostic Accuracy Studies

STROBE – Strengthening the Reporting of Observational studies in Epidemiology

SUS – System Usability Scale

TBI – Traumatic Brain Injury

TEMP - Temperature

TMT – Trail Making Test

TUS – Technological University of the Shannon

UFU – Universidade Federal de Uberlândia

UX – User Experience

VR – Virtual Reality

WHO – World Health Organization

WST – Wheelchair Skill Test

WSTP – Wheelchair Skills Training Program

Part I INTRODUCTION

Chapter 1 Introduction

1.1 Research Motivation

According to the World Health Organization (WHO), over 80 million people require a wheelchair for mobility (WHO, 2023). In UK, over 1.2 million individuals are wheelchair users (NHS, 2021) while in Brazil, approximately 9 million people live with mobility impairments (CENSO, 2010). In Ireland, 22% of the population reported a long-term disability (CSO, 2022). The last detailed national estimate of wheelchair users in Ireland was provided by the 2006 National Disability Survey, which reported that 31,342 individuals used manual or powered wheelchairs, with a further 8,923 indicating an unmet need for one (NDA, 2006). As the populations age and the incidence of chronic conditions such as stroke multiple sclerosis and spine cord injury increases, the demand for assistive mobility solutions, including manual and electric-powered wheelchairs (EPWs), is expected to rise (Kirk-Wade et al., 2024).

EPWs offer essential mobility for individuals with limited motor function. They are frequently prescribed for users with progressive neurological conditions (e.g., multiple sclerosis or muscular dystrophy), spinal cord injuries, or age-related physical functional decline. However, prescribing an EPW is a complex, multi-step process that requires evaluation of physical, cognitive and environmental factors (Fishleigh et al., 2024; WHO, 2008, 2023). Despite the availability of tools such as the Wheelchair Skills Test (WST) (Kirby et al., 2002) and the Power Mobility Road Test (PMRT) (Massengale et al., 2005), these assessments rely heavily on professional observation, which introduces subjectivity and potential variability in scoring (Bigras et al., 2020; Jenkins et al., 2015).

User diversity further complicates prescription, as individuals present with varying cognitive and physical profiles, influenced by personal and environmental contexts (Hoenig et al., 2002; WHO, 2011). This highlights the need for assessment approaches that balance personalization (tailoring to individual needs) with standardization (ensuring comparability across clinical settings), so that evaluations both reflect real-world usage and provide reliable benchmarks (Mannion & Exworthy, 2017; Sinsky et al., 2021).

A further challenge in this process is the lack of safe, structured environments for skill development prior to real-word wheelchair use (Fishleigh et al., 2024; Jenkins et al., 2015; Mathis & Joan Gowran, 2021). In clinical practice, wheelchair training content, duration and frequency varies depending on

the user's needs, abilities, and available support (Dalhousie University, 2023; WHO, 2023). The World Vision guidelines for holistic wheelchair services provision estimate that the training time typically lasts from two weeks to two months (World Vision, 2017). In structured intervention studies, such as those based on the Wheelchair Skills Training Program (WSTP) (Dalhousie University, 2023), users may receive up to five 30-minute individual sessions over one to two weeks, with caregiver participation and encouraged practice between sessions (Kirby et al., 2015). Despite the availability of these training models, many users, particularly those in non-urban or resource-limited areas, have restricted access to consistent training. As consequence, new users may face a steep learning curve and increased risk of falls or collisions, especially among older adults and individuals with spinal cord injuries (Mikolajewska, 2013; Sung et al., 2019; Sydor et al., 2017; Xiang et al., 2006).

Moreover, wheelchair service delivery is highly variable across settings. According to the WHO Wheelchair Provision Guidelines(WHO, 2023) , services may be delivered through primary, secondary, or tertiary healthcare, or in educational, social, and community-based environments. These services include assessment, training, maintenance, follow-up, and referral.

However, factors such as geographic location, workforce capacity, referral systems, and the complexity of user needs influence their consistency and quality. Outreach, task-sharing, and telehealth are often required to reach underserved areas. WHO categorises user needs into basic, intermediate, and advanced levels, each requiring different levels of expertise and resources (WHO, 2023). Equitable access depends on coordinated, multisectoral referral pathways and trained personnel across all levels of care. This variability underscores the need for flexible yet standardised assessment and training approaches that can adapt to diverse delivery models and population needs.

To address these limitations, wheelchair simulators have emerged as a promising digital health solution. They enable repeatable, safe and data-driven assessment and training, supporting skill development before on-road wheelchair use (A. R. de Sá et al., 2022; Zorzi et al., 2024). Evidence shows improvements in user confidence and driving performance following simulator training (Faure et al., 2023; Morère et al., 2018) and show potential for standardised assessment through objective metrics and inter-rater reliability (Mahajan et al., 2013).

Despite these advances, the adoption of wheelchair simulators in clinical practice remains limited. Recent literature reviews (A. R. de Sá et al., 2022; Arlati et al., 2020; Zorzi et al., 2024) highlights key gaps hindering their wider integration:

I. Weak alignment with clinical settings, tools and guidelines – many simulator-based studies does not reflect established clinical assessment protocols (e.g., WST, PMRT and cognitive abilities evaluations), limiting their clinical relevance.
II. Underuse of wearable data – objective physiological measures such as heart rate, heart rate variability, or electrodermal activity are rarely incorporated, even though they can provide insights into user workload, stress, or engagement.
III. Lack of methodological standardisation – simulator studies often use different protocols, tasks, and outcome measures, making results difficult to compare across projects.
IV. Limited inclusion of diverse end-user populations – most studies involve small samples of able-bodied participants, with limited validation among wheelchair users of different ages, diagnoses, or functional profiles.
V. Scarcity of longitudinal and controlled studies – existing research is mainly cross-sectional and short-term, leaving little evidence on long-term learning effects, retention, or real-world transfer.

In addition, aspects that look at how users perceive the technology, such as Quality of Experience (QoE) assessment methods have not yet been systematically applied to the design and evaluation of wheelchair simulators (Vlahovic et al., 2022). QoE, as defined by the International Telecommunication Union (ITU), refers to the user's overall perception of system utility and enjoyment (ITU, 2017). QoE can encompass multiple dimensions relevant to assistive and immersive technologies, including usability, immersion, emotional response and cognitive workload. These multi-dimensions are recognised in QoE and immersive systems research and reflect the experiential demands of virtual reality-based interaction (Vlahovic et al., 2022). Therefore, researchers have explored how human-, system-, and context-related factors influence QoE through both implicit and explicit assessment methods (Callet et al., 2013; Möller & Raake, 2013) (see Figure 1.1). These methods have been applied in various multimedia systems and immersive applications and are critical for understanding how users interact with such systems in assistive and clinical contexts.

In this PhD study, human-related factors include participants' motor and cognitive abilities and simulator and real-wheelchair experience level; system-related factors refer to display modality (e.g., desktop vs. VR headset), virtual motion feedback settings (e.g., low jerk vs. high jerk), and joystick interface design; context-related factors encompass the setting in which the simulator is used, such as controlled-laboratory or rehabilitation centres, and the alignment with clinical assessment tools (e.g., WST(Kirby et al., 2002), PMRT (Massengale et al., 2005), MoCA(Nasreddine et al., 2005)).

Together, these factors shape the users' QoE and are assessed using both explicit and implicit measures.

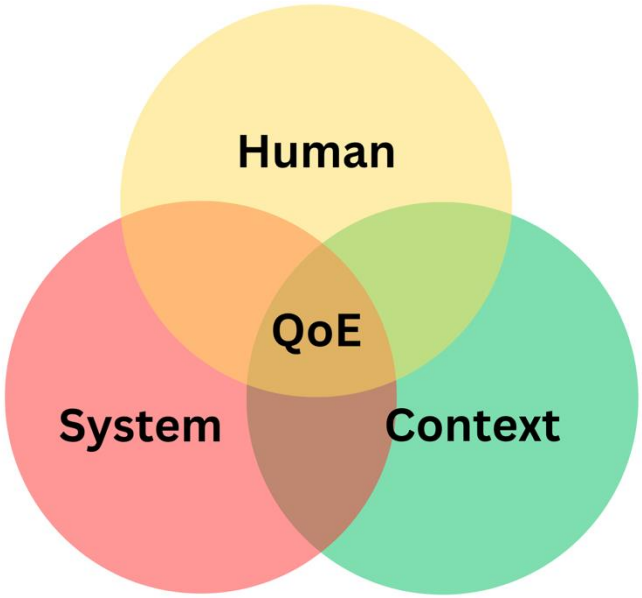


Figure 1.1: Factors that influence QoE in Multimedia Systems. Adapted from (Callet et al., 2013; Möller & Raake, 2013).

In summary, the absence of a structured, QoE-based framework for evaluating user interaction with wheelchair simulator systems has limited their effective use in assessment and training. This thesis specifically addresses two of these gaps: (i) enhancing clinical alignment by designing simulator tasks that reflect established assessments and cognitive evaluation, and (ii) integrating wearable physiological data as implicit indicators of user response. Together, these contributions advance clinical validation, improve usability across diverse user groups, and generate meaningful data to support powered mobility prescription and training. To guide this investigation, a main research question was formulated, supported by two sub-research questions and specific objectives, as detailed in the following sections.

1.2 Research Questions and Objectives

The primary aim of this PhD research is to investigate how a Quality of Experience (QoE)-based evaluation framework can support the structured and clinically applicable use of wheelchair simulators for assessment and training. The framework focuses on evaluating the user's experience during simulator-based tasks, capturing how individuals interact with and respond to the system across experiential, physiological and behavioural-performance dimensions.

The project integrates multiple layers of investigation, including simulator-based performance metrics (e.g., task completion time and driving errors), users' QoE capture through subjective feedback and cognitive workload), physiological responses (e.g., heart rate and electrodermal activity), system design variations (immersive vs. non-immersive conditions), and clinical relevance through comparison with conventional assessment tools and feedback from wheelchair users.

Building on these components, this research explores how immersive and adaptive simulator features can inform the design of preliminary guidelines and protocols that consider individual user needs while maintaining clinical standardisation. The investigation includes lab-based studies examining the impact of immersive settings and capturing physiological responses, as well as a field study conducted with wheelchair users in real-world clinical settings.

1.2.1 Overarching Research Question

The overarching research question (RQ) guiding this thesis is:

> ***RQ: How can a virtual wheelchair simulator be integrated into clinical settings for power mobility training and assessment, by defining protocols and metrics that support structured, safe, and clinically applicable use across a diverse population of power wheelchair users?***

This overarching research question reflects the need to bridge the gap between technological development and clinical implementation in power mobility provision. A key consideration in is the variability in training and assessment practices across countries and service models (WHO, 2023) highlights the importance of grounding the simulator design within real-world clinical settings while ensuring its broader applicability. The Irish Wheelchair Association (IWA) centres serve as use cases to contextualise the investigation within a functioning rehabilitation environment.

This research question comprises two interrelated components. The first concerns the design and evaluation of a simulator-based system through lab-controlled studies, with an emphasis on Quality of Experience (QoE), physiological signals, and user interaction metrics. The second focuses on transferring and evaluating the feasibility of the simulator in clinical environments, focusing on integration with routine practice, alignment with established assessment tools, and the development of pilot guidelines and preliminary protocols for clinical use.

Together, these two components define the scope of this research, which encompasses the development and evaluation of a technically robust and user-responsive simulator, as well as the

demonstration of its clinical feasibility and relevance for structured, safe, and adaptable use within power mobility provision services.

1.2.2 Sub-Research Questions

To support this investigation, the following sub-research questions (SRQs) are addressed, each associated with a set of specific objectives that frame the methodological and design considerations of the work:

1. System Design Investigation via Quality of Experience Evaluations in Controlled Laboratory environment

   SRQ1: How can a virtual wheelchair simulator be designed and tested in a controlled environment to establish a clinically relevant proof of concept that supports multidimensional assessment, incorporating immersive technologies, physiological signals, subjective feedback, and Quality of Experience (QoE) evaluation?

This question was addressed through controlled lab studies focused on system design evaluation. The following objectives supported this investigation:

1.1. Evaluate QoE in wheelchair simulator use by combining subjective ratings and biomarkers.
1.2. Assess the influence of immersive technology design on usability, performance, cognitive workload, and simulator-induced discomfort (cybersickness).
1.3. Examine how virtual motion settings affect user experience, including usability, performance, cognitive workload and cybersickness.
1.4. Design an initial evaluation protocol for clinical pilot studies, informed by findings from the above objectives.

These objectives were developed to follow a logical progression aligned with the technical and experiential priorities of the simulator. The first objective focused on establishing a multidimensional Quality of Experience (QoE) assessment approach. QoE was examined through four key dimensions: usability (efficiency and ease of use), emotional response (such as frustration or satisfaction), immersion (the extent to which users felt present in the virtual environment), and cognitive workload (mental effort required to perform tasks). These experiential factors are increasingly being used in immersive systems research to understand how users engage with virtual environments. In this study, they were applied to evaluate user interaction with the simulator beyond performance outcomes, capturing perceptual and affective responses that are critical to clinical relevance and user acceptability (Vlahovic et al., 2022).

The second objective compared immersive and non-immersive display conditions to determine their impact on both user experience and functional performance. Immersive technologies refer to interactive systems, such as head-mounted displays (HMDs), that present users with visual, auditory, and spatial information to simulate real-world environments. In the scope of this study, immersive Virtual Reality (VR) was used, delivered through HMDs to simulate powered mobility tasks. While Augmented Reality (AR) has been explored in other training contexts, its reliance on overlaying physical spaces limits standardisation for assessment (Phadke et al., 2024). VR was therefore prioritised, as it enables controlled and repeatable replication of complex mobility scenarios, providing the ecological control and consistency essential for clinical validation (Caruso et al., 2025). Immersion is understood as the sense of presence (SoP), or the extent to which users perceive themselves to be situated within the virtual environment (Arlati et al., 2020). While increased immersion may enhance engagement and task realism, it may also contribute to cybersickness, a known challenge in virtual reality applications that remains difficult to manage effectively. Cybersickness typically manifests as a set of symptoms including nausea, dizziness, oculomotor discomfort, and disorientation, resulting from sensory conflicts experienced in virtual environments (Arlati et al., 2020; Vlahovic et al., 2022). To mitigate these risks, the simulator utilized tuned motion profiles and time-limited exposure, strategies consistent with recommendations for mitigating cybersickness in immersive systems (Vlahovic et al., 2022).

The third objective investigated how virtual motion settings, specifically the acceleration and deceleration profiles of the simulated wheelchair, affect user experience. These parameters were treated as planned design variables aimed at balancing ecological realism with user comfort. As cybersickness had been identified as a relevant usability concern, this objective focused on exploring how motion characteristics could be adjusted to minimise discomfort while maintaining system usability and task realism. Studies on wheelchair simulators confirm that motion feedback and tuning strategies can substantially influence both simulator effectiveness and user tolerance (Arlati et al., 2020).

The fourth objective consolidated findings on QoE, immersive design, and motion settings into a preliminary protocol and simulator configuration, intended as a basis for pilot implementation and discussion with healthcare professionals at the IWA. Rather than replacing established assessments such as the WST, PMRT and MoCA, this protocol demonstrates how a simulator can complement them by providing safe, repeatable, and objective metrics that are difficult to capture in real-world testing environments. In this way, the simulator addresses current gaps by offering structured practice

opportunities and quantifiable performance data, while remaining aligned with validated clinical assessment tools.

Together, these objectives addressed key challenges, including defining a QoE model suitable for immersive assistive technologies, balancing realism and usability in immersive design, and mitigating cybersickness through motion configuration. The outcomes provided practical insights to guide the simulator's design and its preparation for pilot use in clinical settings.

2. User Evaluation in Real-World Environment (Community Centre)

> ***SRQ2: How can the proof-of-concept simulator be transferred into clinical settings, using Irish Wheelchair Association (IWA) centres as a use case, and how can protocols and evaluation methods be developed to test the feasibility its components, reflect the perspectives of wheelchair users, and support standardised implementation?***

Following system development, the second phase of the investigation focused on evaluating how the simulator could be embedded in clinical practice and assessed for feasibility and preliminary alignment with standard tools. This question was addressed through a field study involving user evaluations, focusing on assessing the feasibility, acceptability and practical applicability of the simulator for real-world use. The following objectives supported this investigation:

2.1. Conduct a field pilot feasibility study to assess the feasibility and acceptability of the simulator within Irish Wheelchair Association (IWA) centres.
2.2. Defined and analyse the simulator-based metrics that are aligned with standard clinical assessments for power mobility skills and cognitive abilities (WST and MoCA) to evaluate alignment and potential complementarity.
2.3. Develop preliminary framework (guidelines and protocol recommendations) for simulator use in future pilot and validation studies, informed by feasibility findings, comparative analysis, and feedback from users.

These objectives supported the transition from controlled laboratory testing to pilot implementation in clinical environments by focusing on process feasibility, exploratory metric comparison, and practical recommendations.

The first objective examined the feasibility of implementing the simulator in routine practice across two IWA facilities, located in Athlone and Cork areas (Irish Wheelchair Association (IWA), n.d.). Simulator sessions were embedded within the centres' regular activity schedules and introduced to wheelchair users who are frequent visitors to the centres for activities such as physiotherapy, leisure programmes, mobility training, and other supports aimed at maintaining or improving quality of life.

By situating the sessions in familiar environments and among existing support services, the study aimed to ensure ecological validity and evaluate acceptability among both users without disrupting existing routines.

The second objective focused on analysing simulator-based metrics by comparing them with established clinical assessments. The WST and MoCA were selected due to their extensive validation and widespread use in mobility and cognitive assessment, and their inclusion in evidence-based resources such as the Rehabilitation Measures Database, a well-established tool curated by Shirley Ryan Abilitylab. However, within this database, validated tools for wheelchair users are primarily centred on manual wheelchair skills. The WST is included, but its validation and application are mainly focused on manual wheelchair users, with only limited adaptations for powered mobility proposed by its developers (Dalhousie University, 2023; Kirby et al., 2002; Kirby, 2017). The PMRT, while one of the few structured protocols created specifically for powered wheelchair users, is not included in the database and still depends on professional observation and four-point rating scales (Massengale et al., 2005). This highlights a broader gap, powered wheelchair assessment remains difficult due to the diversity of user needs, especially where motor and cognitive profiles vary widely (Pellichero, Best, et al., 2021; Torkia et al., 2015), Current approaches are therefore largely subjective and lack technology-supported, objective measures that could provide consistent, repeatable, and clinically transferable insights (A. R. de Sá et al., 2022). The simulator is positioned not as a replacement, but as a complementary tool to enhance existing assessments with data-driven indicators and to support healthcare professionals in training and evaluation.

To enable this comparison, the simulator recorded a range of performance-based metrics, including task completion time, number of collisions, and deviation from optimal paths. These metrics reflect competencies commonly assessed in simulator-based mobility evaluations (A. R. de Sá et al., 2022) and were intended to provide objective, repeatable data that could complement conventional assessment approaches.

The third objective brought together feasibility findings, comparative analysis with clinical tools, and feedback from users and staff into a set of preliminary recommendations. These guidelines provide a practical basis for pilot use of wheelchair simulators in clinical contexts and highlight considerations for future large-scale validation studies.

Together, these objectives addressed practical and methodological challenges. Integrating the simulator into clinical routines required adapting study procedures (pre-assessment, in-simulator tasks, and post-assessment) without disrupting users' engagement or the structure of scheduled activities. Comparing simulator outputs with clinical tools posed difficulties due to differences in

measurement approaches and subjectivity of traditional assessments. The outcomes offer early insights to guide clinical pilot use and inform subsequent validation research.

1.3 Contributions and Publications

The following contributions were achieved in alignment with the overarching research question, two sub-research questions, and the research objectives. As illustrated in Figure 1.2, the thesis followed a structured progression from lab-based to field-based studies, each designed to explore specific aspects of simulator design and user response. These studies were grounded in Quality of Experience (QoE) and Cognitive Load Theory (CLT), combining user- and human-centred design approaches to ensure usability, empathy, and performance. Each study directly informed a key contribution, culminating in the formulation of practical guidelines and methodological recommendations, derived from pilot evidence, to inform and support the future development of an adaptable evaluation framework for the clinical integration of wheelchair simulators.

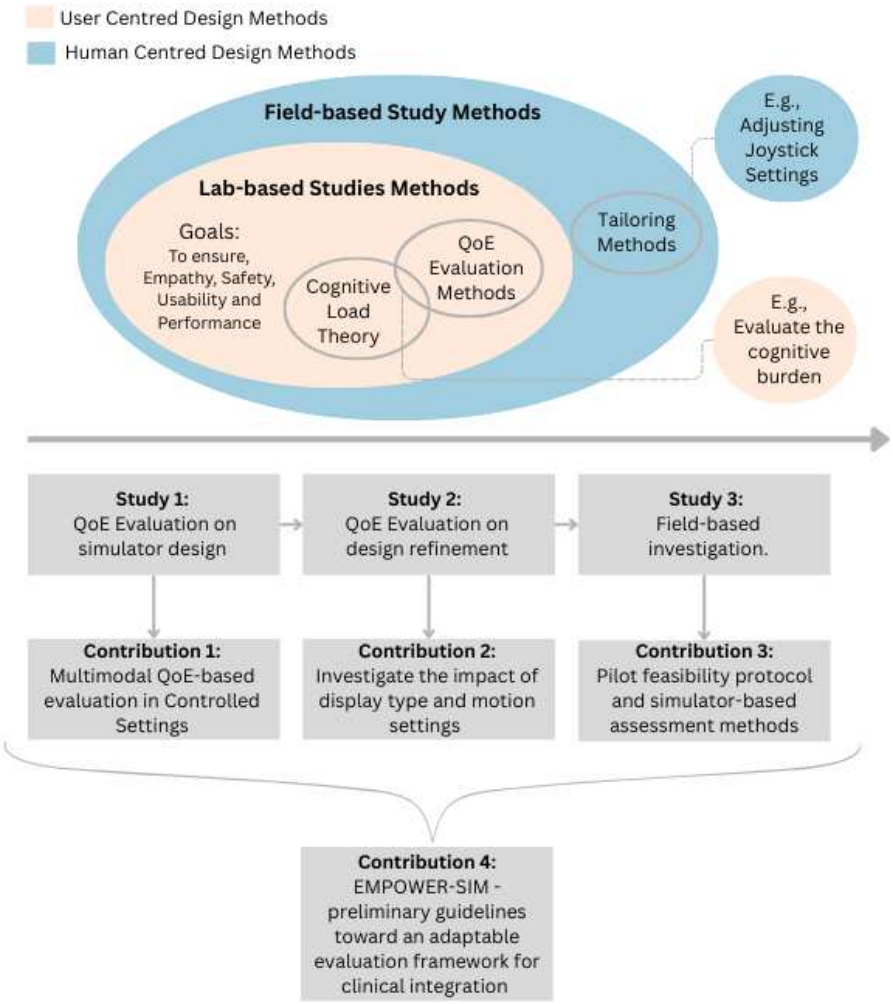


Figure 1.2: Overview of study design and contributions.

1.3.1 Main Contributions

Contribution 1: Multimodal Quality of Experience Evaluation in Controlled Settings

Applied and refined a multidimensional approach to assess user experience in immersive wheelchair simulators through controlled studies combining subjective and physiological data. This contribution addresses SRQ1 and supports objectives 1.1 and 1.2.

Publications:

- Débora Pereira Salgado, Ronan Flynn, Naves, Eduardo Lázaro Martins Naves, Niall Murray. A questionnaire-based and physiology-inspired quality of experience evaluation of an immersive multisensory wheelchair simulator. In Proceedings of the 13th ACM Multimedia Systems Conference (MMSys'22) (2022). DOI: https://doi.org/10.1145/3524273.3528175
- Débora Pereira Salgado, Thiago Braga Rodrigues, Conor Keighrey, Ronan Flynn, Felipe Roque Martins, Eduardo Lázaro Martins Naves, Niall Murray. *A QoE assessment method based on EDA, heart rate and EEG of a virtual reality assistive technology system*. In Proceedings of the 9th ACM Multimedia Systems Conference (MMSys'18) (2018). DOI: https://doi.org/10.1145/3204949.3208118

Contribution 2: Impact of Display Type and Motion Settings on Users Experience and Cybersickness

Investigated how immersive display types (VR headset vs. desktop) and motion feedback (low vs. high jerk) influence usability, sense of presence/immersion, cognitive load, and simulator-induced discomfort. Findings provided evidence to guide design decisions, including the adoption of desktop display for clinical settings (field-base study). This contribution addresses SRQ1 and supports objective 1.3.

Publications:

- Débora Pereira Salgado, Ronan Flynn, Eduardo Lázaro Martins Naves, Niall Murray. *The impact of jerk on quality of experience and cybersickness in an immersive wheelchair application*. In proceedings of the 12th International Conference on Quality of Multimedia Experience (QoMEX) (2020). DOI: https://doi.org/10.1109/QoMEX48832.2020.9123086
- Débora Pereira Salgado, Thiago Braga Rodrigues, Felipe Roque Martins, Eduardo Lázaro Martins Naves, Ronan Flynn, Niall Murray. *The Effect of Cybersickness of an Immersive Wheelchair Simulator*. Procedia Computer Science (160), (2019), 665-670, ISSN 1877-0509, DOI: https://doi.org/10.1016/j.procs.2019.11.030

Contribution 3: Pilot Feasibility Protocol and Simulator-Based Assessment Methods

Designed and piloted a mixed-methods feasibility protocol to explore the integration of the wheelchair simulator into real-world clinical workflows for assessing power mobility and cognitive abilities. The approach combines simulator-derived metrics with powered wheelchair users and clinician feedback, enabling feasibility testing and exploratory alignment in end-users. This contribution addresses SRQ2 and supports objectives 1.4, 2.2 and 2.3.

Publication:

- Débora Pereira Salgado, Caroline Valentini de Queiroz, Eduardo Lázaro Martins Naves, Yuansong Qiao, Sheila Fallon. Protocol for evaluation of a virtual wheelchair simulator in assessing mobility skills and cognitive abilities in diverse populations: A multicentric mixed-methods pilot study. PLOS ONE 20(6): e0325186, (2025). DOI: https://chatgpt.com/c/68750a52-b6dc-800a-9225-6fa6ceee1717https://doi.org/10.1371/journal.pone.0325186

Contribution 4: EMPOWER-SIM – Preliminary Guidelines Toward an Adaptable Evaluation Framework for Clinical Integration

Formulated EMPOWER-SIM, as a set of preliminary, feasibility-informed guidelines and methodological recommendations, informed by findings across Contributions 1–3. Contribution 1 established multidimensional QoE assessment methods; Contribution 2 examined display and motion design factors affecting usability and tolerance; Contribution 3 applied these insights in a pilot feasibility protocol. These stages provided the empirical basis for EMPOWER-SIM, which integrates user experience, performance metrics, and physiological responses to align simulator outcomes with clinical tools (e.g., WST, PMRT, MoCA). It represents an initial step toward a future validated framework for clinical integration. This contribution addresses RQ (Main), SRQ1, SRQ2 and supports objectives 1.1, 1.2, 1.4, 2.1 and 2.3.

Publication (planned):

- Débora Pereira Salgado, Caroline Valentini de Queiroz, Eduardo Lázaro Martins Naves, Yuansong Qiao, Sheila Fallon. An Adaptable Quality of Experience-Based Evaluation Framework for Virtual Wheelchair Simulators in Clinical and Training Contexts. PLOS ONE. Status: to be submitted.

1.3.2 Supporting Contributions

In addition to the core research outcomes, a set of supporting contributions emerged throughout the project. These contributions were developed in parallel with the primary investigations and provide practical resources and outputs that enhance the accessibility, reproducibility, and potential application of the research. They include the release of a public dataset, the development of an analysis tool, and the co-development and registration of the simulator system.

Supporting Contribution 1: WheelSimPhysio-2023 Dataset

Published a dataset comprising physiological, performance, and QoE data from 58 participants, supporting benchmarking and reproducibility in multimodal simulator research.

Publication:

- Débora Pereira Salgado, Sheila Fallon, Yuansong Qiao, Eduardo Lázaro Martins Naves. WheelSimPhysio-2023 dataset: Physiological and questionnaire-based dataset of immersive multisensory wheelchair simulator from 58 participants. Elsevier Data in Brief, 54 (2024). DOI: https://doi.org/10.1016/J.DIB.2024.110535

Supporting Contribution 2: WheelSimAnalyser – Multimodal Data Analysis Tool

Developed a MATLAB tool for pre-processing, synchronisation, feature extraction, and visualisation of simulator data, enabling reproducible multimodal analysis.

Publication:

- Débora Pereira Salgado, Niall Murray, Ronan Flynn, Eduardo Lázaro Martins Naves, Yuansong Qiao, Sheila Fallon. *WheelSimAnalyser: A MATLAB tool for multimodal data analysis of WheelSimPhysio-2023 dataset*. Elsevier *Software Impacts*, 23 (2024). DOI: https://doi.org/10.1016/j.simpa.2024.100731

Supporting Contribution 3: Simulator Development and Software Registration (E-WATS)

Co-developed the simulator system and registered it under E-WATS for intellectual property protection.

Registration:

- Eduardo Lázaro Martins Naves, Felipe Roque Martins., Débora Pereira Salgado, Angela Rosa de Sá. *E-WATS - Electric Wheelchair Assessment and Training Simulator*. (2019). INPI. https://busca.inpi.gov.br/pePI/jsp/programas/ProgramaSearchBasico.jsp.

1.4 Thesis Outline

The structure of thesis reflects the progression from identifying clinical and technological gaps in power wheelchair provision to proposing, piloting, and refining a feasibility-focused evaluation framework for simulator-based assessment. It is organised into five parts comprising six chapters, each contributing to different aspects of the framework's design, implementation, and empirical investigation. Figure 1.3 illustrates the overall structure of the thesis and the relationships between its chapters.

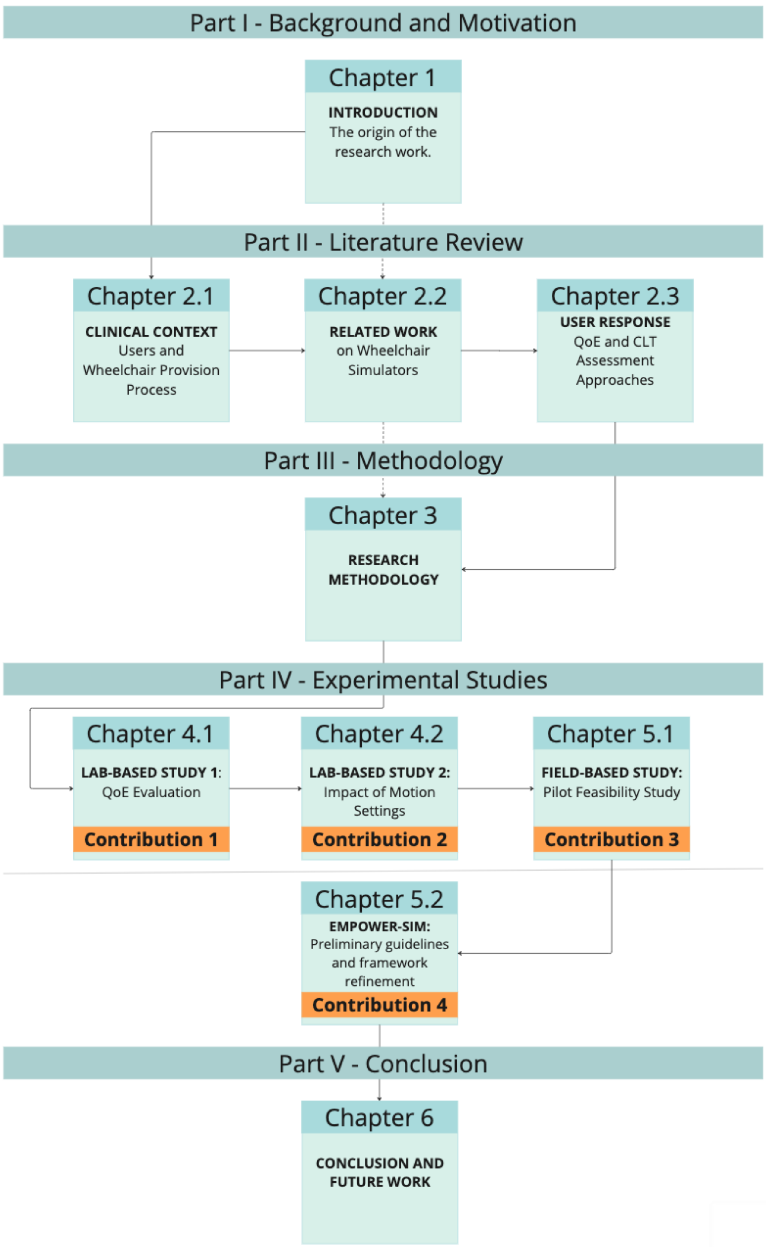


Figure 1.3: Thesis Outline.

Part I – Introduction

Chapter 1 – Background and Motivation: Introduces the clinical and technological challenges of power wheelchair provision, outlines the potential of simulators as digital health tools, and presents the research problem, questions, objectives, and contributions.

Part II – Literature Review

Chapter 2 – Clinical and Research Background on Power Mobility and Wheelchair Simulators: Combines three areas of focus. The first is the **clinical background of powered mobility**, including user diversity, assessment practices, and the WHO wheelchair provision guidelines. The second is **wheelchair simulator research**, which surveys existing systems, hardware and software components, metrics, study designs, and application domains. The third is **user response assessment approaches**, which explores explicit measures such as subjective ratings and behavioural indicators alongside implicit measures such as physiological signals, framed by Quality of Experience (QoE) and Cognitive Load Theory (CLT).

Part III – Methodology

Chapter 3– Research Methodology: Describes the mixed-methods research design, detailing study procedures, participant groups, and the development of an adaptable evaluation framework grounded in QoE.

Part IV – Experimental Studies

Chapter 4 – Lab-based Studies: Lab-based Studies presents controlled experiments on simulator design and immersive technologies. These address Sub-Research Question 1 (SRQ1) and contribute to Contributions 1 and 2 by developing a multidimensional QoE approach, testing display and motion settings, and informing the initial protocol.

Chapter 5 – Field-based Study: Reports a pilot feasibility study addressing Sub-Research Question 2 (SRQ2) and Contributions 3 and 4, focusing on process feasibility, exploratory alignment with WST and MoCA, and preliminary refinement of the evaluation framework.

Part V – Conclusion

Chapter 6 – Conclusion and Future Work: Summarises the findings across all studies, highlights both main and supporting contributions, discusses study limitations, and outlines directions for future validation studies and clinical adoption.

Part II LITERATURE REVIEW

Chapter 2 Clinical and Research Background on Power Mobility and Wheelchair Simulators

This chapter provides the foundation for the simulator-based framework proposed in this thesis. It reviews three interconnected areas: section2.1 the clinical background of powered mobility provision, section 2.2 research on wheelchair simulators, and section 2.3 about methods for evaluating user responses in QoE and CLT contexts. Together, these strands establish both the current challenges and the rationale for integrating simulator-based approaches into clinical practice.

2.1 Clinical Background of Powered Mobility Provision

Power mobility devices (PMDs), including powered wheelchairs and scooters, are critical enablers of independence and quality of life for individuals with severe physical disabilities. Prescribing a powered wheelchair goes beyond device selection; it requires evaluating user capabilities, safety considerations, and the long-term fit of the device within the user's daily environment (WHO, 2022, 2023). Distinct challenges arise from (i) highly diverse physical, cognitive, and psychosocial profiles that shape outcomes (Hoenig et al., 2002; WHO, 2011), (ii) reliance on professional observation that can introduce subjectivity and variability (Bigras et al., 2020; Jenkins et al., 2015), (iii) international variability in service pathways and training opportunities (WHO, 2008, 2023), and (iv) limited access to structured pre–on-road practice environments, which raises risk for novice users (Fishleigh et al., 2024; Mathis & Joan Gowran, 2021).

2.1.1 Power Wheelchair Users

Power wheelchair users represent a highly diverse population, encompassing individuals with varying degrees of motor, cognitive, and sensory impairments (Hoenig et al., 2002; WHO, 2011). These users may differ in age, condition onset (congenital or acquired), disease progression, and levels of functional independence. Some rely on powered mobility due to significant motor impairments, while others may also experience cognitive or perceptual challenges that affect navigation, safety, and decision-making (WHO, 2023). Recognising and accounting for this diversity is essential in the development of adaptive technologies and inclusive assessment strategies.

Disability is dynamic and multifactorial, shaped by ageing, recovery from injury, progression of chronic conditions, and environmental factors. The WHO's International Classification of Functioning, Disability and Health (ICF) provides a framework that moves beyond medical diagnoses to consider body structures and functions, activity limitations, participation restrictions, and contextual factor (WHO, 2011).

Powered mobility devices (PMDs) are prescribed to support independence and participation for those with moderate to severe mobility limitations. Disabilities may be permanent (e.g., congenital musculoskeletal anomalies), progressive (e.g., multiple sclerosis), or situational (e.g., post-injury/surgery) (WHO, 2022). Technologies must therefore accommodate both stable and evolving needs (WHO, 2022).

Disabilities commonly associated with PMD use can be grouped as neurological/neuro-physical, musculoskeletal, and other conditions, with frequent overlap. **Error! Reference source not found.** summarises conditions frequently linked to powered wheelchair prescription, including neuro-physical disabilities such as spina bifida, hydrocephalus, cerebral palsy, muscular dystrophy, osteogenesis imperfecta, arthrogryposis, spinal muscular atrophy, spinal cord injury, stroke, and multiple sclerosis.

Table 2.1: Common Impairments Affecting Mobility and Power Mobility Device

Category	Conditions
Neurological & Neuro-physical	Spinal Cord Injury (SCI), Stroke / Cerebrovascular Accident (CVA), Multiple Sclerosis (MS), Parkinson's Disease, Cerebral Palsy (CP), Traumatic Brain Injury (TBI), Amyotrophic Lateral Sclerosis (ALS), Spina Bifida, Hydrocephalus, Spinal Muscular Atrophy (SMA)
Musculoskeletal	Muscular Dystrophy (MD), Rheumatoid Arthritis (RA), Osteogenesis Imperfecta (OI), Arthrogryposis, Limb Amputation, Bone Fractures, Osteoporosis, Congenital Musculoskeletal Anomalies.
Others	Cognitive Impairments (e.g., intellectual disability, dementia), Respiratory Disorders (e.g., chronic obstructive pulmonary disease), Diabetes Mellitus, Cardiovascular Disease (e.g., heart failure).

The range of conditions listed in Table 2.1 illustrates the heterogeneity of the powered wheelchair user population. The empirical work in this thesis does not attempt to encompass this full spectrum; rather, it focuses on feasibility-oriented studies approach with a limited participant sample, consistent with methodological recommendations for feasibility research (Bowen et al., 2009).

To address variability within this scope, participants were functionally profiled using validated clinical assessments. The Wheelchair Skills Test (WST) captured self-reported and observed motor competence, while the Montreal Cognitive Assessment (MoCA) provided structured insight into

cognitive functioning. In addition, observations by healthcare professionals, who supported the recruitment process, contributed contextual understanding of participants' functional abilities and mobility needs. These measures offered a pragmatic means of reflecting user heterogeneity during simulator development and evaluation, ensuring that both motor and cognitive aspects were considered systematically.

Future research will need to extend this approach to broader user groups to fully address the complexity outlined in Table 2.1. The following section explores the motor and cognitive abilities that underpin power mobility use, highlighting their relevance to the design of simulator-based assessments and training protocols, and to the alignment of evaluation methods with clinical practice.

2.1.2 Essential Motor and Cognitive Abilities

Motor and cognitive abilities are critical to the safe and effective use of PMDs. Motor impairments can limit the capacity to initiate, control, and sustain movements required for device operation, affecting muscle strength, joint coordination, or postural control. In such cases, powered wheelchairs become essential tools for mobility and participation. These limitations also influence how users engage with adaptive technologies, including VR-based training or simulator tools, underscoring the need for flexible and personalised design.

Effective PMD operation requires coordination between gross and fine motor skills. Gross motor skills encompass movements of large muscle groups necessary for maintaining posture, balance and trunk control during movement. Fine motor skills are associated with more precise actions, such as manipulating joysticks or alternative input controls. Additionally, hand-eye coordination is important for navigation (Navarro et al., 2020), and trunk stability is necessary to prevent falls (Garner & Ricard, 2022; Patel et al., 2017; Sprigle et al., 2003). Head and neck control also contribute to safe operation by preventing dizziness and disorientation during mobility tasks (Cooper et al., 2006; Geers et al., 2023; Sprigle et al., 2007).

Cognitive functions are equally essential, enabling users to process environmental information, make decisions, and adapt responses. In this thesis, cognitive functions refer to specific domains such as attention, memory, and executive control, while cognitive abilities describe measurable performance within these domains. These processes are particularly important in dynamic or unsupervised environments, where judgment and adaptability determine safe navigation (Amudha & B. William Dharma, 2016). Individual variation in cognitive abilities is shaped by factors such as age, health status, and personal characteristics(S. R. Diamond & Royce, 1980).

Cognitive functioning, or cognition, encompasses the effective application of these processes in task contexts, manifesting as behaviour (Fisher et al., 2017). This includes executive functions, attention, and memory (Deligkaris et al., 2014). For PMD users, relevant functions include problem-solving to overcome obstacles, spatial awareness for navigation, and memory to recall operational steps (Massengale et al., 2005).

Core executive functions (CEFs) are a higher-order subset of cognition responsible for coordinating information to support goal-directed behaviour (Friedman & Miyake, 2017). These include working memory, inhibitory control, and cognitive flexibility. They are described as "distinct latent variables" because, while correlated, each can be modelled as a separate construct (e.g., through factor analysis of test performance)(A. Diamond, 2013). CEFs can improve through training or intervention, but are also sensitive to stress, isolation, and physical inactivity, all of which can impair performance (A. Diamond, 2013). Such interdependencies highlight the relevance of both physical and psychosocial factors to PMD safety and training.

Recent evidence from a scoping review underscores the complexity of the relationship between cognitive functioning and PMD use (Pellichero, Kenyon, et al., 2021a). Key domains—including visuospatial perception, attention, memory, judgment, and executive functions—were found to influence users' capacity to operate PMDs, with both predictive and correlational associations identified across age groups(Cullen et al., 2008; Furumasu et al., 2004; Massengale et al., 2005; Tefft et al., 1999). Despite this, clinical assessments often rely on subjective judgment, which can limit access for individuals with cognitive impairments and reduce consistency in decision-making (Maywald & Stanley, 2015; Mortenson et al., 2013; Pellichero, Kenyon, et al., 2021a). Several tailored training interventions have demonstrated improvements in both mobility and cognitive outcomes, particularly when adapted to individual needs (Benford, 2017; Jones et al., 2003; Kenyon et al., 2015; Kenyon, Jones, et al., 2018; Pellichero, Kenyon, et al., 2021a). This adaptability is central to simulator-based methods, which allow scenarios and metrics to be adjusted to user profiles rather than assuming a uniform baseline.

In summary, motor and cognitive abilities contribute to safe and effective PMD use. **Error! Reference source not found.** outlines key abilities required for operation, alongside their specific relevance to the simulator, indicating whether each domain is directly measurable, indirectly observable, or beyond the current scope. This mapping clarifies how different skill areas are addressed within the simulator design and evaluation.

One study showed that cognitive functioning, assessed via Montreal Cognitive Assessment (MoCA), was associated with power wheelchair performance (Power mobility Indoor Driving Assessment –

PIDA), user confidence (Wheelchair Skills Test Questionnaire – WST-Q) and life-space mobility (Life-Space Assessment – LSA) (Pellichero, Best, et al., 2021). Users with moderate impairment scored significantly lower than those with mild or no impairment, underscoring the role of both motor and cognitive domains in functional capacity. These findings highlight why simulator-based assessment must incorporate both dimensions when evaluating performance and training effectiveness. The next section examines how these abilities translate into task-based mobility skills.

Table 2.2: Overview of motor skills and cognitive abilities for PMD use and Simulator Assessment

Types	Specific Skills	Characteristics	Relevance to Simulator
1. Motor Skills	**1.1 Gross Motor Skills**	Large muscle groups for posture, balance, trunk control.	Not in scope
	Balance	Ability to stay upright and stable while manoeuvring the PMD.	Not in scope
	Posture	Maintaining appropriate seated position.	Not in scope
	1.2 Fine Motor Skills	Precise use of small muscles in hand/fingers.	Assessed through joystick and wrist accelerometer input data.
	Hand-eye Coordination	Synchronising visual input with joystick control.	Reflected in accurate navigation, minimal path deviation, and reduced collisions.
	Dexterity	Performing smooth and precise movements.	Measured through jerkiness/smoothness of joystick signals and driving path.
	Grip Strength	Sufficient force to manipulate controls.	Not in scope
2.Cognitive Skills	**2.1 Lower-Order Cognitive Skills**	Fundamental processes supporting basic PMD use.	Observed indirectly through performance outcomes.
	Perception	Interpreting obstacles and environment.	Reflected in head movements (scanning), obstacle avoidance, and collision frequency.
	Attention	Sustaining focus on navigation tasks.	Seen in missed cues, task completion time, and error rate.
	Memory	Recalling operational steps or navigation rules.	Reflected in improvement across repeated tasks.
	2.2 Higher-Order Cognitive Skills	Advanced processes for decision-making and adaptation.	Performance in complex driving scenarios.
	Problem-Solving	Overcoming unexpected obstacles.	Measured through alternative route choices and recovery from errors.
	Decision-Making	Choosing safe/efficient paths.	Reflected in junction choices, hazard negotiation, and path selection.
	Critical Thinking	Anticipating and correcting errors.	Observed through avoidance of repeated mistakes and strategic manoeuvres.

2.1.3 Power Mobility Skills and Clinical Performance

While the previous section outlined the foundational motor and cognitive domains necessary for powered mobility use, this section focuses on how these abilities translate into observable, task-based skills. Power mobility skills refer to the integrated application of physical control, cognitive processing, and perceptual awareness required for operating a powered wheelchair or similar PMD. These skills are typically assessed through tasks-based evaluations and represent a key focus in both clinical rehabilitation and simulator-based training programmes (Smith et al., 2022).

Task analysis frameworks have mapped power mobility skills to International Classification of Functioning (ICF) domains, with a substantial proportion attributed to mental functions (Smith et al., 2022). Skills such as spatial reasoning, problem-solving, and visual perception are critical to navigation and safety, particularly in dynamic or unfamiliar environments (Massengale et al., 2005; Tefft et al., 1999). Importantly, individuals with cognitive and perceptual impairments can learn to use power wheelchairs effectively through structured and tailored training approaches (Mountain et al., 2010).

The simulator developed in this thesis was designed with adaptability as a core principle. At the task level, the level of challenge can be adjusted, for example by adding more turns to a route, including narrow doorways, or introducing moving obstacles. At the interface level, adaptability refers to the types of controls available. Conventional joystick input is supported, but unlike most real powered wheelchairs where only speed can usually be changed without engineering support for acceleration and deceleration, the simulator allows these parameters to be modified directly. In addition, alternative input methods such as eye-gaze control and facial EMG were implemented to accommodate users who cannot rely on standard joystick operation. This adaptability allows the system to reflect a wide range of user needs without assuming a uniform baseline.

In clinical settings, the evaluation of power mobility performance often focuses on task execution, environmental adaptation, and confidence in navigation. These elements are influenced by both fine motor control, such as joystick manipulation and cognitive functions including planning, judgment, and attention (Pellichero, Kenyon, et al., 2021a). Research has demonstrated that targeted training programmes can lead to improvements not only in mobility capacities but also in cognitive functioning among individuals with heterogeneous impairments (Pellichero, Kenyon, et al., 2021a). Variation across settings is to be expected given the diversity of PMD user requirements, but this variability complicates comparability of outcomes across studies and clinical contexts (Kenyon et al., 2020).

In summary, the development of power mobility skills requires consideration of both cognitive and motor abilities in task-oriented contexts. However, ongoing challenges persist in evaluating and

training powered wheelchair users due to safety risks, heterogeneous user profiles, diverse training approaches, and inconsistent assessment tools (Kirby, 2017; Pellichero, Best, et al., 2021; Tu et al., 2017). As highlighted in a recent scoping review (Pellichero, Kenyon, et al., 2021a), power mobility device (PMD) training programmes range from standardized to highly individualized interventions, leading to inconsistent outcomes and limited comparability across studies. Simulator-based approaches have been proposed as one means of addressing these gaps: they generate objective, repeatable performance data to complement traditional clinical tools, while still allowing training to be tailored to specific needs. Adaptation may involve focusing on particular skills—such as joystick control, path following, or obstacle avoidance—based on a user's baseline performance. In this way, simulators can support both comparability and personalisation, aligning with evidence-based practice without promoting a prescriptive "one-size-fits-all" model.

These findings underscore the need for training and evaluation protocols that strike a balance between structured and adaptable approaches while remaining grounded in evidence. Such protocols must reflect the complex interplay between cognitive abilities, motor demands, environmental contexts, and technology use. The following section explores how powered wheelchair design and user interface features further influence these dynamics and impact safe and effective mobility outcomes.

2.1.4 Technology, Interfaces and Safety

Wheelchairs have undergone significant technological advancements over time, evolving from heavier steel models used in the 1930s to modern versions made of lightweight materials such as aluminium and titanium (Cooper et al., 2006). As technology has progressed, power mobility devices (PMDs), particularly powered wheelchairs (PWCs), have become more advanced, offering greater adaptability, programmability, and integration with users' physical and cognitive capabilities.

PWCs typically rely on electronic control systems, with joystick interfaces being the most common. Directional input is often divided into four quadrants: forward-right, forward-left, reverse-left and reverse-right. The VR2 joystick system, for example, translates directional input into movement commands using hall-effect sensors that detect joystick displacement along horizontal and vertical axes. Wheelchair controllers determine motion using vector velocity across two or more motors, with forward and reverse speeds programmable via motor speed limits (Cooper et al., 2006). Additional parameters such as acceleration, deceleration, turning radius, and braking can also be adjusted to match user capabilities, enhancing personalisation and safety. This approach is particularly useful in virtual simulations, where core motor control strategies can be assessed under controlled conditions.

Despite their benefits, PWC use is associated with safety challenges. Users report high rates of incidents, with collisions being a leading cause of injury in powered mobility compared to falls in manual use (Edwards & McCluskey, 2010; Leblong et al., 2021). Single-user accidents have also increased, often linked to cognitive or perceptual limitations and environmental barriers (Carlsson & Lundälv, 2019). Advanced features such as obstacle detection and programmable speed limits can reduce risks, but their effectiveness is partial; studies show that these features help prevent certain incidents yet do not fully mitigate risks arising from user variability, environmental complexity, and cognitive demands (Dicianno et al., 2019).

Findings by (Dolan & Henderson, 2017) emphasise the importance of tailoring powered wheelchair control interfaces to individual user needs. In a large postal survey of over 250 users, the vast majority (94.6%) relied on hand joysticks, yet a significant proportion reported issues such as fatigue, discomfort, and accidental activation. These challenges were particularly prevalent among users with limited upper limb function or progressive conditions. While most participants expressed high levels of satisfaction, the study highlighted that a subset of users might benefit from adjustments, alternative control types (e.g., chin joysticks or switches), re-evaluation of device settings or a simulator-based trial to determine optimal settings.

The simulator developed in this thesis incorporates this principle of flexibility. It supports configurable joystick parameters such as sensitivity, acceleration and deceleration, and includes alternative input methods such as eye-gaze and facial EMG. These options allow for testing of different configurations within a controlled environment, offering a safer pathway to explore suitability before real-world implementation. However, for the empirical studies reported in this thesis only the joystick interface was applied, to maintain comparability across participants.

In summary, the effectiveness of powered wheelchair control interfaces varies across individuals, influenced by factors such as upper limb strength, dexterity, fatigue, and sensory or cognitive impairments, all of which affect control accuracy, comfort, and safety during use (Dolan & Henderson, 2017; Massengale et al., 2005). These demands shape how users interact with the device and respond to environmental challenges. Ensuring safe and effective power wheelchair use depends not only on device design but also on how services assess, train, and support users.

The configurability factor is particularly relevant for clinical simulations, where user responses to different control schemes can inform assessment protocols and guide individualised training approaches. The following section introduces international guidelines that structure these provision processes, with a focus on those developed by the World Health Organization.

2.1.5 Provision Processes and Guidelines

Wheelchair provision involves more than the distribution of a mobility device; it includes a coordinated process of assessment, fitting, training and follow-up that should align with everyone's needs and context. Over time, global guidance has shaped how these services are structured, particularly through documents produced by WHO. In 2008, the WHO introduced the Guidelines on the Provision of Manual Wheelchairs in Less Resourced Settings, outlining a structured 8-step service model. This framework influenced the design rationale of the simulator presented in this thesis, particularly in relation to the assessment and user training stages, where controlled environments can provide opportunities for skill development and evaluation prior to on-road use.

WHO 8-step model remains as a reference in many countries, especially in rehabilitation programmes and community-based services (Dicianno et al., 2019; Quiñones-Uriostegui et al., 2023). It emphasises the role of trained personnel, individual assessment, and continuity of care across the user journey. **Error! Reference source not found.** summarises these steps. Each stage was designed to support a person-centred and context-aware approach to service delivery, especially in settings where resources and infrastructure may be limited.

Table 2.3: WHO 8-Step Model for Wheelchair Provision (2008)

Step	Description
1. Referral and appointment	Initiating the process and scheduling a service appointment.
2. Assessment	Evaluating the user's physical, environmental, and functional needs.
3. Prescription	Selecting the appropriate wheelchair type and features based on assessment.
4. Funding and ordering	Securing resources and ordering the selected wheelchair.
5. Product preparation	Assembling or configuring the wheelchair prior to fitting.
6. Fitting	Adjusting the wheelchair to the user's body and environment.
7. User training	Teaching the user (and caregiver, if needed) to use and maintain the wheelchair.
8. Follow-up, maintenance and repair	Ongoing support to ensure the wheelchair remains safe, functional, and appropriate.

Although this model remains relevant, growing diversity in wheelchair users, services, and policy environments has led to the development of updated guidance. In 2023, the WHO released the Wheelchair Provision Guidelines, offering a more streamlined framework intended to support implementation across a broader range of systems and populations. These guidelines acknowledge that wheelchair service delivery is highly variable across settings. Services may be delivered through primary, secondary, or tertiary healthcare, or in educational, social, and community-based environments.

This version introduced a simplified 4-step model (Select, Fit, Use, Follow-Up) intended to offer a more adaptable structure while retaining core service elements. **Error! Reference source not found.** outlines these steps. The updated framework retains a person-centred approach. The steps are grounded in principles that frame wheelchair provision as a human right, promote equitable access, and recognise its role within universal health coverage. They emphasise that services should be responsive to individual needs and supported by trained personnel across all stages.

Table 2.4: WHO 4-Step Model for Wheelchair Provision (2023)

Step	Description
1. Select	Assessment and selection of the most appropriate wheelchair for the user.
2. Fit	Configuring and fitting the wheelchair to the user's posture, mobility needs, and environment.
3. Use	Providing training and support to ensure the user can operate and maintain the wheelchair safely and effectively.
4. Follow-up	Monitoring wheelchair use, performing maintenance, and reassessing user needs over time.

While the updated model offers a simplified structure, the practical application of wheelchair provision continues to depend on context-specific factors, including available expertise, infrastructure, and user diversity. Persistent challenges remain in assessing user needs, providing adequate training, and selecting appropriate wheelchair configurations, particularly for individuals with cognitive or perceptual difficulties (Fishleigh et al., 2024; Gowran et al., 2021; Mathis & Joan Gowran, 2021). These challenges underscore the potential added value of simulation by offering safe and repeatable environments for novice users, generating objective performance data to complement professional judgement, and enabling controlled trials of different control settings prior to adjustments on a physical wheelchair. In this way, simulation supports provision processes rather than replacing them, addressing areas where conventional clinical pathways often lack capacity or consistency.

Gowran et al. (2021) further highlight systemic gaps in provision services, including limited training opportunities, policy inconsistency, and fragmented service delivery, which hinder equitable access to appropriate wheelchairs. The authors emphasise the need for coordinated strategies that extend beyond structured models to include skilled personnel, user-centred practices, and sustained investment. Within this context, the simulator developed in this thesis is not positioned as an alternative to existing provision frameworks but as a supporting tool that can be embedded particularly within the “Use” and “Follow-up” stages, strengthening both training opportunities and clinical decision-making.

The following section examines the clinical assessment and training tools currently used to support these stages, highlighting opportunities for integration with simulation-based approaches.

2.1.6 Clinical Assessment and Training Tools

2.1.6.1 Power Mobility skills assessments

Structured observation tools have long been used to evaluate power mobility capacity and training needs. The Wheelchair Skills Test (WST) remains one of the most widely adopted instruments, offering a reliable, standardised battery of indoor and outdoor tasks scored by clinicians (Kirby et al., 2002). Its broad adoption is attributed to its structured design, clinical feasibility, and ability to track progress over time. However, like many observational tools, its effectiveness depends on controlled testing conditions and clinician supervision, and it emphasises physical execution more than the cognitive or perceptual strategies used by the individual. This limitation has been noted in the literature (Kirby, 2017; Mortenson et al., 2018).

Self-report measures, such as the Wheelchair Skills Test Questionnaire (WST-Q), provide a scalable way to capture user-reported ability and confidence. They are practical in community and remote settings and have strong test–retest reliability (Mortenson et al., 2018; Rushton et al., 2016). Yet, like all post-activity questionnaires, their accuracy is affected by recall, insight, and safety awareness. Therefore, they are most informative when combined with observational or performance-based methods.

Other instruments expand on specific contexts. The Power Mobility Indoor Driving Assessment (PIDA) evaluates competency in 30 indoor tasks and has shown inter-rater reliability, though intra-rater variation and observer subjectivity remain concerns (Dawson et al., 1994). The Power Community Driving Assessment (PCDA), developed through consensus methods, focuses on real-world community tasks, giving it high ecological validity. However, despite its structured tasks, its scoring still depends heavily on clinician interpretation (Letts et al., 2007), raising the same concerns about variability.

The Power Mobility Road Test (PMRT) incorporates both structured and semi-structured driving tasks, assessing hesitation, confidence, and safety across 16 scenarios (Massengale et al., 2005). Although validated for inter-rater reliability, its four-point scale relies on verbal instructions and subjective clinician judgment, which can result in inconsistent application, particularly among users with cognitive or perceptual differences. In this thesis, the simulator builds on the structured approach of the PMRT by replicating its task design and extending it with objective, repeatable metrics. For each PMRT-based task, the simulator generates performance indicators such as task completion time, path

deviation, and number of collisions, thereby complementing clinician scoring with quantitative data and reducing reliance on subjective observation.

The Functional Evaluation Rating Scale (FERS) was originally developed for simulator-based assessment and has been applied in training studies(Hasdai et al., 1998). However, recent literature does not report widespread clinical adoption of FERS, and its use appears largely confined to research contexts (A. R. de Sá et al., 2022; Arlati et al., 2020). The Power Mobility Functional Evaluation Tasks (PMFET) similarly targets simulated tasks, particularly for paediatric users, but its use has remained confined to research contexts (Deitz et al., 1991). These examples suggest that simulator-based tools have encountered barriers to broader clinical uptake, with their use remaining primarily within research contexts. This highlights the need for future systems to demonstrate alignment with established assessments, integration into clinical workflows, and added value for both clinicians and users.

Finally, more recent tools such as the Powered Mobility Device Assessment Training Tool (PoMoDATT) and the Power Mobility Training Tool (PMTT) broaden assessment to include cognitive, physical, and psychosocial domains (Kenyon, Farris, et al., 2018; Townsend & Unsworth, 2019). While these instruments illustrate the field's movement toward more holistic and flexible evaluation, their adoption to date appears limited, with reported use largely in specialised or research-focused contexts rather than widespread clinical practice. Table 2.5 provides a condensed summary of these tools.

In summary, power mobility assessment tools vary widely in structure and application, with most relying on clinician observation and subjective scoring. This provides flexibility for adapting evaluations to diverse user needs but also introduces variability and potential bias. Self-report instruments such as the WST-Q offer scalable alternatives, though they remain sensitive to response bias and require clinical interpretation. Despite these advances, gaps remain in the integration of cognitive, perceptual, and environmental factors into assessments, particularly with respect to balancing comparability and individualisation. Simulator-based methods may offer one avenue to address this by generating objective performance metrics while permitting tasks and scenarios to be adjusted to user profiles. The following section examines assessment instruments focused on cognitive functioning and decision-making capacity in powered wheelchair use.

Table 2.5: Power Mobility Skills Assessment Tools

Assessment Tool	Purpose	Key Features	Considerations	Implementation in this Thesis
Wheelchair Skills Test (WST) (Kirby et al., 2002)	Evaluate wheelchair skills in controlled settings	Standardized tasks, obstacle-laden environments, expert presence for adjustments.	Requires expert supervision; limited insight into cognitive strategies	Partially implemented: simulator tasks were mapped to WST.
Wheelchair Skills Test Questionnaire (WST-Q) (Mortenson et al., 2018; Rushton et al., 2016)	Self-reported wheelchair skill ability and confidence	Practical for remote/community use; strong test–retest reliability	Subject to self-report bias; should be interpreted with clinical input	Implemented.
Power Indoor Driving (PIDA) (Dawson et al., 1994)	Assess indoor driving performance	30 structured tasks; strong inter-rater reliability	Intra-rater variation; observer subjectivity influences results	Not implemented
Power Community Driving Assessment (PCDA) (Letts et al., 2007)	Assess real-world outdoor wheelchair use	Person–environment interaction in community settings	High ecological validity but scoring depends on clinician judgment	Not implemented
Power Mobility Road Test (PMRT) (Massengale et al., 2005)	Assess structured and unstructured driving ability	Combines real and simulated challenges; 16 tasks	Relies on subjective 4-point scoring; varies across clinicians	Implemented the structured tasks form PMRT into simulator
Functional Evaluation Tasks (FERS) (Hasdai et al., 1998)	Simulator-based skill training	Tracks progression; repeatable virtual scenarios	Mainly research-based; limited real-world clinical uptake	Not implemented
Power Mobility Functional Evaluation Tasks (PMFET) (Deitz et al., 1991)	Assess wheelchair use in simulated indoor environments	Focus on children, functional positioning, and navigation	Used in research; not yet integrated into clinical pathways	Not implemented
Powered Mobility Device Assessment Training Tool (PoMoDATT) (Townsend & Unsworth, 2019)	Comprehensive assessment for users with complex needs.	Integrates cognitive, physical, and psychosocial domains.	Time-intensive; requires interdisciplinary input	Not implemented
Power Mobility Training Tool (PMTT) (Kenyon, Farris, et al., 2018)	Paediatric-focused early power mobility tool.	Supports joystick/switch access; developmental progression.	Limited generalisability; for specific paediatric contexts	Not implemented

2.1.6.2 Cognitive assessments

Cognitive assessment tools play a vital role in determining a person's cognitive abilities and their suitability for operating a PMD. In addition to observational assessments conducted during therapy sessions, standardized cognitive assessments can provide valuable insights into the client's cognitive functions. These tools are typically used to identify impairments that may interfere with mobility planning, attention to the environment, or safe navigation. For instance, impairments in executive function may affect a user's ability to plan routes or respond to unexpected changes in the environment (Ashley et al., 2017). Similarly, attention deficits can reduce situational awareness and increase the risk of collisions or missed hazards (Chang et al., 2014).

Table 2.6 summarizes commonly used cognitive assessments tools that, although not developed specifically for powered mobility, are frequently applied to support clinical decision-making during the wheelchair prescription process. While each tool listed has demonstrated clinical utility, they vary in their scope and applicability.

Table 2.6: Cognitive Abilities Assessment Tools

Cognitive Assessment Tool	Description	Key Cognitive Areas Assessed	Target Populations	Implementation in this Thesis
Montreal Cognitive Assessment (MoCA) (Nasreddine et al., 2005)	A brief screening tool assessing multiple cognitive domains, requiring training for administration.	Attention, memory, visuospatial abilities, executive function, language.	Individuals with neurological injuries	Implemented: Used to profile participants' cognitive performance and linked to simulator metrics.
Trail Making Test (TMT) (Tombaugh, 2004)	Evaluates visual perceptual skills and cognitive flexibility through a two-part task.	Task-switching, sequencing, working memory.	Individuals with neurological impairments	Not Implemented
Repeatable Battery for the Assessment of Neuropsychological Status (RBANS) (Randolph et al., 1998)	Measures various cognitive functions, validated for individuals with neurological injuries.	Memory, attention, language, visuospatial abilities.	Individuals with stroke, TBI, or dementia	Not Implemented
Cognitive Assessment for Stroke Patients (CASP) (Benaim et al., 2022)	A standardized tool specifically designed for individuals' post-stroke (CVA).	Attention, memory, language, visuospatial abilities.	Stroke survivors	Not Implemented

The MoCA (Nasreddine et al., 2005) is one of the most used screening tools in rehabilitation due to its broad coverage and sensitivity to mild cognitive impairment. The TMT (Tombaugh, 2004) offers insights into processing speed and flexibility, useful for evaluating users' ability to shift attention or manage sequential tasks. The RBANS (Randolph et al., 1998) provides a more comprehensive profile of cognitive performance, while the CASP (Benaim et al., 2022) addresses need specific to individuals recovering from stroke.

Recent findings by (Pellichero, Best, et al., 2021) provide strong support for the clinical relevance of cognitive assessments in powered wheelchair provision. Their exploratory study involving experienced PWC users demonstrated that cognitive functioning, as measured by the MoCA, significantly influenced PWC performance (PIDA), user-reported confidence (WST-Q), and life-space mobility (LSA). Principal component analysis showed that cognitive function, visual perception, and physical performance explained a large portion of the variability in mobility outcomes. Importantly, the study cautioned against using MoCA scores in isolation, recommending a combined approach using both cognitive screening and functional assessments to inform more accurate and inclusive clinical decisions.

A complementary scoping review by the same authors (Pellichero, Kenyon, et al., 2021)further highlighted the importance of cognitive evaluation in power wheelchair provision. The review identified core cognitive domains related to PMD use, such as attention, executive function, memory, and visuospatial processing, and found that individuals with cognitive impairment are often excluded from PMD access due to clinician-perceived safety risks. However, the authors emphasised that specific cognitive thresholds for safe PMD use remain unclear. Instead, they advocate for the integration of cognitive tools like the MoCA within mobility skills assessments to support both ethical and effective clinical practice.

Moreover, the results from cognitive assessments can guide the selection of alternative control strategies. Users with sufficient cognitive capacity but significant motor impairment may benefit from non-standard control interface methods, such as head arrays, sip-and-puff systems, or eye-gaze interfaces (Fager, 2018). While these clinical assessment tools were not developed for powered mobility contexts, their inclusion alongside observational and simulator-based assessments supports more holistic and equitable decision-making in wheelchair provision.

In this thesis, the MoCA was incorporated alongside simulator-based metrics to profile participants' cognitive performance and explore its relationship with driving outcomes. Other tools listed in Table 2.6, while clinically relevant, were not implemented here but are included to illustrate the wider

landscape. The following section explores current strategies used to train users in operating PMDs, with a focus on structured interventions for individuals with cognitive and physical impairments.

2.1.6.3 Training programs

Training in power mobility use is a vital component of wheelchair provision, aiming to enhance user competence, confidence, and safety. Clinical training programs are designed not only to support initial skill acquisition but also to facilitate long-term adaptation and autonomy. These programs often incorporate structured assessment, task progression, and motor learning principles. Among them, the Wheelchair Skills Training Program (WSTP) (Dalhousie University, 2023; Kirby, 2017)is one of the most widely adopted evidence-based approaches. It forms part of the Wheelchair Skills Program (WSP), which also includes the Wheelchair Skills Test (WST) and the Wheelchair Skills Test Questionnaire (WST-Q). The WSTP integrates goal setting, task breakdown, demonstration, feedback, and progressive practice to guide both users and clinicians through a structured training pathway.

A systematic review and meta-analysis by (Tu et al., 2017) confirmed the short-term effectiveness of the WSTP in improving wheelchair skills capacity. Data from 10 randomised controlled trials (RCTs) and 7 non-randomised studies showed a 13.26% improvement in WST scores in RCTs and a 23.44% improvement in non-randomised studies immediately post-training. The review also reported that the WSTP was generally safe, with few adverse events occurring during training. However, evidence for long-term (3–12 month) effectiveness and the impact of powered wheelchair skills training remains limited, warranting further investigation.

Despite its demonstrated benefits, structured training programs like the WSTP are still underutilised in clinical practice. For example, only 7.9% of surveyed assistive technology professionals reported using standardised assessment tools or training frameworks in PMD provision (Jenkins et al., 2015), with lack awareness being a significant factor in their adoption. Another recent survey by (Kirby et al., 2020) found that only 43.5% of occupational therapists in Nova Scotia reported routinely providing wheelchair skills training to clients, and even fewer (34%) felt adequately prepared for the trainer role. Most reported providing only brief training sessions, with a median of two sessions lasting 30 minutes each.

Comparable issues have been identified in the Irish context, where a cross-sectional survey (Mathis & Joan Gowran, 2021) found that while basic training is commonly delivered, health professionals lack confidence in teaching advanced wheelchair skills and have limited access to standardised training programs. Many rely on self-directed learning and express a strong desire for continued professional education in this area. These studies highlighted significant barriers to implementation, including time

constraints, insufficient support, and lack of formal training. Yet, the majority of therapists recognised wheelchair skills training as very important for both clients and caregivers, underscoring a disconnect between clinical priorities and practical realities.

The research presented on (Nilsson & Durkin, 2014; Nilsson & Kenyon, 2022) created a comprehensive assessment tool for learning powered mobility use, emphasizing a user-led approach and facilitating strategies across the learning continuum. Likewise, the study in (Livingstone, 2010) highlighted the importance of focusing on a continuum of learning, the reciprocal relationship between trainer and trainee, and environmental factors in skill development.

Collectively, these studies reinforce the value of standardised, evidence-based training programs while highlighting persistent gaps in implementation. Training approaches grounded in established frameworks such as the WSTP and informed by user-centred focus models offer strong potential for improving power mobility outcomes. These insights are especially relevant to the current study, which explores the use of wheelchair simulators as an extension of traditional training Simulators enable safe, repeatable, and contextualised learning environments that can replicate real-world challenges and, within current clinical constraints, often support only limited, iterative skill development. By generating objective performance data and allowing tasks to be adapted to user needs, simulators may help bridge the gap between standardisation and personalisation.

In summary, clinical training programs such as the WSTP provide a structured and evidence-based foundation for supporting power mobility users in developing essential skills and confidence. These programs have demonstrated short-term effectiveness and are generally considered safe, largely due to therapist support and additional supervision that helps prevent injuries during training (Tu et al., 2017). However, evidence regarding their long-term effectiveness remains limited, and their implementation in clinical settings continues to face practical challenges. Common barriers include time constraints, insufficient training for clinicians, and variability in service delivery (Kirby et al., 2020). In the Irish context, many clinicians report limited access to formal training and a lack of confidence in teaching more advanced wheelchair skills (Mathis & Joan Gowran, 2021). These limitations underscore the need for complementary strategies that can enhance accessibility, individualisation, and safety in training delivery. The following chapter explores how wheelchair simulators may address some of these challenges by offering controlled, repeatable, and adaptable environments for assessment and training in powered mobility.

2.1.7 Summary

Powered mobility provision is complex: diverse user presentations, evolving needs, and variable service contexts intersect with safety, access, and equity. Robust assessment and training exist (e.g., WST/PMRT, WST-Q, MoCA, WSTP) but are often observer-dependent and inconsistently implemented, with limited integration of cognition and constrained training capacity in routine care (Kirby et al., 2020; Mathis & Joan Gowran, 2021).

These realities motivate practical, clinically aligned protocols and guidelines, including the use of controlled, repeatable environments, to support safe skill development and more consistent decision-making. In the following work, we pursue such protocols, aligning simulator-based assessment and training with established clinical tools and user needs, while carefully situating findings as preliminary and feasibility-focused.

2.2 Research on Wheelchair Simulator

Simulation technologies are being explored in wide range applications in healthcare, such as education (Curran et al., 2023), rehabilitation and medical skills training (Abbas et al., 2023; Mao et al., 2021), offering controlled environments for safe practice and performance assessment. Developments in virtual reality (VR), augmented reality (AR), and extended reality (XR) have enabled the creation of immersive, interactive and customizable simulated environments (Vincenzi et al., 2023), typically accessed through head-mounted displays (HMDs), projection systems, or mixed-reality platforms.

In the context of powered mobility, simulators have gained attention as a promising approach to complement traditional real-world assessment (A. R. de Sá et al., 2022) and training methods (Zorzi et al., 2024). Originally adopted in aviation, defence and automotive testing (Vincenzi et al., 2023), simulators are increasingly explored in clinical research and pilot programs involving wheelchair users (A. R. de Sá et al., 2022; Arlati et al., 2020; Zorzi et al., 2024). These systems provide the opportunity to evaluate user performance and deliver structured training in repeatable, low-risk scenarios, particularly when real-world testing poses safety or accessibility concerns (Gefen et al., 2019; Morère et al., 2018).

Despite these promising developments, the integration of wheelchair simulators into clinical workflows remains limited. Key challenges include the absence of standardised outcome measures (A. R. de Sá et al., 2022) and validation protocols across studies, which hinders comparisons and limits generalisability of findings (Zorzi et al., 2024). Considerable heterogeneity in simulator configurations (Arlati et al., 2020; Zorzi et al., 2024), spanning hardware and software features, immersion levels, and user tasks, further complicates the establishment of clinical guidelines (Arlati et al., 2020). Additional concerns include usability barriers such as cognitive overload (Vailland et al., 2021)interface complexity, and the occurrence of cybersickness, especially with immersive displays, can negatively impact user experience and tolerability (Zorzi et al., 2024). Furthermore, many existing studies are proof-of-concept or prototype-focused (Arlati et al., 2020), with limited clinical validation or deployment in real-world settings (Zorzi et al., 2024).

Unlike many existing prototypes, the simulator developed in this thesis is explicitly designed to bridge the gap between proof-of-concept systems and clinical applicability. Its key differentiating features are: (i) alignment with established clinical tools such as the WST and PMRT, enabling comparison with real-world assessments; (ii) incorporation of QoE methods to evaluate usability, immersion, and cognitive demand, moving beyond performance metrics alone; and (iii) integration of multimodal data, combining task performance, subjective feedback, and physiological signals, to capture both

explicit and implicit user responses. Additionally, this work adopts an iterative, user-centred design process involving end-users and clinicians, with the goal of improving ecological validity and supporting eventual integration into healthcare workflows.

As immersive simulation technologies continue to evolve, regulatory bodies are placing emphasis on clinical safety, effectiveness, and data governance. For example, the European Union for medical device regulation 2017/746 and the U.S. Food and Drug Administration (FDA) guidance on *Augmented Reality and Virtual Reality in Medical Devices* and *Software as a Medical Device (SaMD)*outline expectations for validation, risk management and transparency. These evolving requirements highlight the importance of rigorous, standardised evaluation and alignment with clinical objectives if wheelchair simulators are to be successfully integrated into healthcare workflows.

This section presents a literature review on the current state of wheelchair simulator research. It begins with examining core components and design features relevant to wheelchair simulator systems. It continues by analysing findings from key studies, particularly those focused on virtual reality powered wheelchair simulator-based training and assessment and concludes with an evaluation of the methodological and practical challenges that hinder clinical translation. It builds a foundation for proposing a structured, simulator-based framework for assessing and training powered mobility users. The exploratory questions guiding this literature review are:

- In what ways are wheelchair simulators currently being designed and studied for training and assessment in powered mobility provision?
- What evaluation methods and study design strategies are employed to ensure clinical relevance, usability, and inclusivity?
- How can the approach to wheelchair simulators be standardised so that adoption becomes more widespread?
- How does the simulator proposed in this thesis differ from existing systems in terms of clinical alignment, multidimensional assessment, and readiness for integration into healthcare workflows?

2.2.1 Wheelchair Simulator Main Components

Wheelchair Simulators (WSs) are assistive technology systems designed to replicate aspects of real-world wheelchair operation within controlled virtual or semi-virtual environments. They offer structured and repeatable settings for both training and assessment, especially valuable for users with motor, cognitive, or perceptual impairment (Roberts et al., 2012; Vailland et al., 2019). Their design typically integrates both hardware and software components to create an interactive system capable of delivering realistic movement, feedback, and task-based scenarios. While the specific configurations vary widely, most simulators share a set of core components. At the same time, prior reviews note that publications vary in how explicitly they link component choices to clinical outcomes, user tolerability, and real-world transfer, which can make translation beyond laboratory feasibility less straightforward (A. R. de Sá et al., 2022; Arlati et al., 2020; Lam et al., 2018; Zorzi et al., 2024).

Figure 2.1 presents a high-level framework for a general wheelchair simulator system and its integration into an assessment-and-training context. This model is informed by a synthesis of design characteristics reported across simulator studies published between 2000 and 2024 and is used here as an organising scaffold for the subsequent analysis of systems and findings (A. R. de Sá et al., 2022; Arlati et al., 2020; Lam et al., 2018; Pithon et al., 2009; Zorzi et al., 2024).

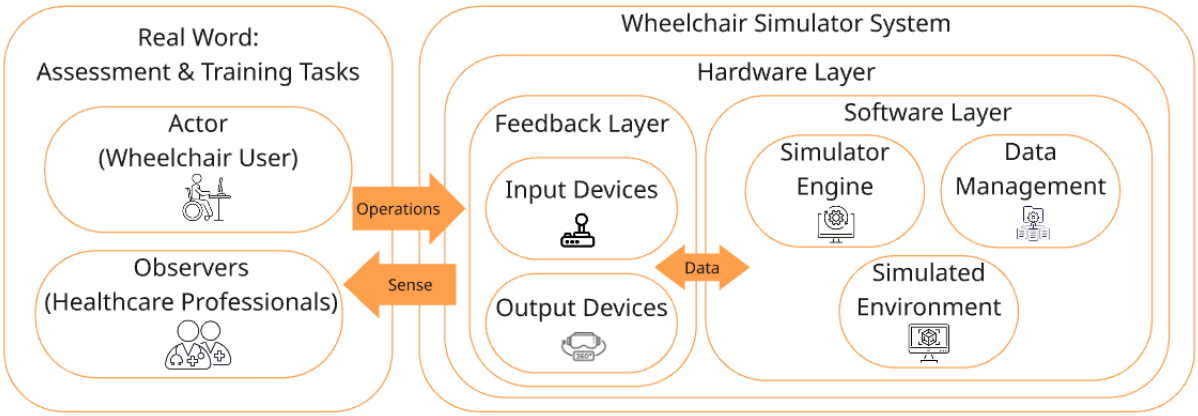


Figure 2.1: System architecture of general wheelchair simulator framework.

On the left side of the diagram, the real-world layer depicts the actor (wheelchair user) engaged in simulator-based tasks, while observers (typically healthcare professionals) evaluate the user's performance. These interactions produce both operational and observational data. On the right, the wheelchair simulator system is decomposed into hardware and software layers. The hardware layer includes input devices (e.g., joystick controllers) and output devices (e.g., head-mounted displays, haptic feedback units, or auditory systems) that facilitate bidirectional communication with the user. These components form the feedback layer, enabling the user to interact with and respond to the

simulated environment. The software layer comprises the simulator engine, which governs real-time control, physics modelling, and communication protocols, and the simulated environment where tasks are rendered. Additionally, a data management module is responsible for logging user activity, storing system settings, and supporting performance evaluation.

This architecture enables immersive user interaction while allowing observers to monitor progress and extract multimodal performance metrics. Together, these components support the structured delivery of wheelchair training and assessment in clinically meaningful ways.

2.2.1.1 Hardware Features: I/O Devices and sensory feedback

The hardware layer of wheelchair simulators shapes how users interact with the system and what feedback they receive. In turn, these choices influence sense of presence, immersion, user engagement, and overall quality of experience (QoE). Across systems reported between 2000 and 2024, configurations vary in complexity, cost, and purpose—from desktop displays to high-fidelity multisensory platforms (A. R. de Sá et al., 2022; Arlati et al., 2020; Pithon et al., 2009; Zorzi et al., 2024). While higher fidelity can support greater realism, reviews also note recurring issues such as cybersickness and motion–visual mismatch (particularly with motion platforms), and practical barriers to deployment related to cost and complexity (A. R. de Sá et al., 2022; Arlati et al., 2020; Zorzi et al., 2024). These considerations highlight a design trade-off between immersive feedback and clinical feasibility that subsequent sections examine in more detail.

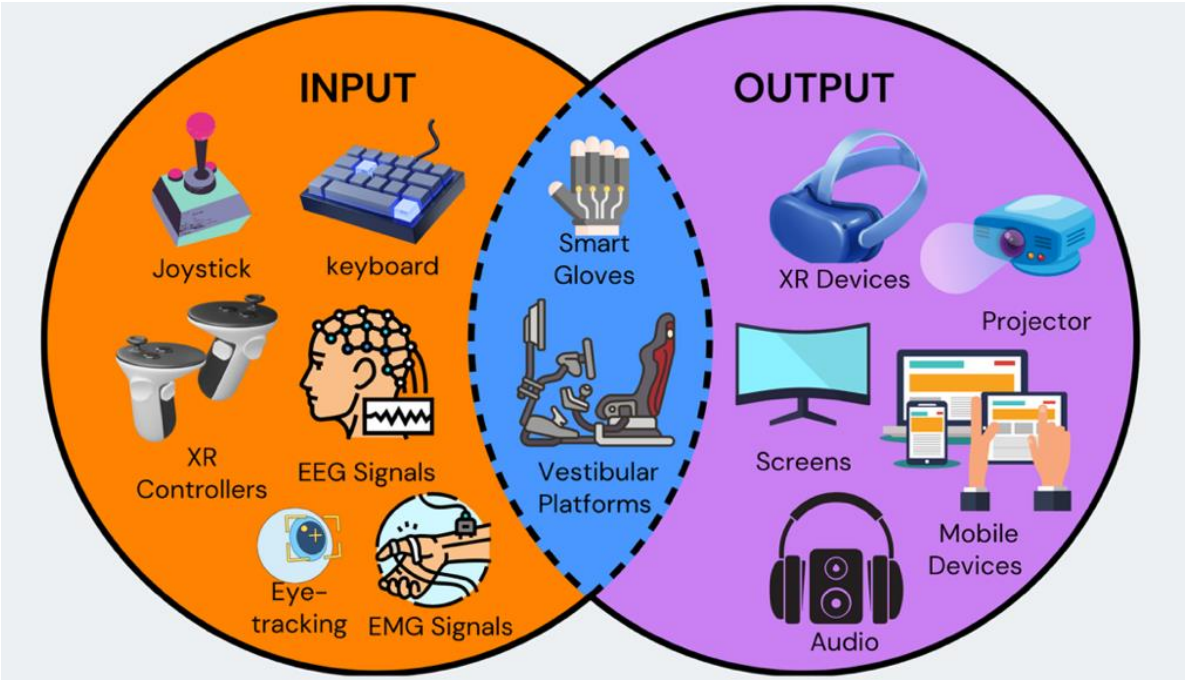


Figure 2.2: Venn diagram of input and output components in VR Power Wheelchair Simulators.

As illustrated in Figure 2.2, input devices may include physical controllers (e.g., joystick, keyboard, XR hand controllers), as well as physiological signals as alternative controls (e.g., EEG, EMG, eye-tracking). On the output side, systems typically incorporate XR displays, screens, auditory feedback, and mobile projection.

Input Devices

Input devices are essential for capturing user commands and translating them into simulated actions. The most common input is a standard or adapted wheelchair joystick (Arlati et al., 2020; Zorzi et al., 2024), typically modelled after standard power wheelchair controllers to ensure transferability of skills. Some simulators also incorporate alternative control interfaces to accommodate users with limited upper-limb functions (A. R. de Sá et al., 2022; Zorzi et al., 2024). These include Eye-tracking (e.g., SimCadRom (H. Montenegro-Couto et al., 2018)), EMG (e.g. E-WATS (Martins, 2022), Brain-computer interfaces (BCI) using EEG (Younis et al., 2019), Switch-based controls (Rodriguez, 2018), Head arrays and adaptable controllers (e.g., miWE (Tao & Archambault, 2016)). This diversity reflects an increasing focus on accessibility and customization, particularly in clinical and paediatric contexts. Accessibility-driven inputs increase inclusivity, but many reports are lab-only with small samples and limited clinical anchoring, which weakens claims of real-world benefit.

Output Devices and Sensory Feedback

Output devices are responsible for delivering sensory feedback to the user. These include visual displays, such as monitors, projectors, or head-mounted displays (HMDs), which render the virtual environment. Auditory systems that replicate environmental sounds or provide navigation cues. Haptic feedback units, such as vibration motors on the joystick or seat, which provide tactile reinforcement for collisions or surface changes. Motion platforms (in high-end systems), which simulate real-world wheelchair movement by physically shifting the user's base in response to virtual terrain. The combination of these devices influences the level of sense of presence /SoP (involvement) and immersion (the feeling of being part of the virtual world, interacting directly with the environment), user tolerability, and the realism of the simulated experience. Systems differ in the extent to which they prioritize realism, cost-efficiency, and accessibility, especially when designed for clinical use (Arlati et al., 2020).

Simulator Design Implications

Visual and auditory feedback are commonly employed to simulate navigation cues and environmental interaction, haptic feedback remains less frequent and often limited to vibrational cues. More immersive configurations incorporate vestibular feedback through motion platforms, simulating

acceleration, incline, or uneven terrain. However, as noted by (A. R. de Sá et al., 2022) these setups remain rare due to their cost and complexity. Their clinical utility and impact on performance and learning outcomes also require further validation (Arlati et al., 2020; Pithon et al., 2009). These configurations illustrate a wide spectrum of design choices, from experimental research platforms to clinically oriented and training-focused tools.

In summary, the Configurations vary according to study aims, participant characteristics, and available resources. In this thesis, hardware trade-offs are examined specifically in relation to visual immersion: a desktop/monitor condition is compared with a VR HMD condition, with effects evaluated on usability, presence/immersion, cognitive workload/ emotion affect, tolerability/cybersickness, and implicit user state (physiological arousal). These analyses provide evidence regarding tolerability, safety-relevant proxies, and performance with direct relevance to clinical adoption. Motion platforms and dedicated haptic feedback (seat) are outside the present scope and are identified as priorities for future work. Table 2.7 situates these design choices within recent literature and underscores the rationale for prioritising clinical feasibility and measurement breadth over high-cost fidelity. The following section examines the software components that complement these hardware features, focusing on virtual-environment design and simulator-derived metrics.

Table 2.7: Study-level comparison of Wheelchair Simulators

Simulator / Authors	Population	Env.	Clinical tools	Immersion	Outcome measures	Physiological response	Study
VRSIM-2 (Kamaraj, Dicianno, et al., 2016a)	Experienced PWC users (N=21). Clinicians (N=5)	Lab/ Clinic	PMRT	3 plasma screens	PMRT	No	Feasibility/ pilot
VRSIM-2 (Kamaraj, Dicianno, et al., 2016b)	Experienced PWC users (N=21)	Lab/ Clinic	PMRT	3 plasma screens	PMRT and NASA-TLX	No	Feasibility/ cross-sectional
miWe (Tao & Archambault, 2016)	Experienced PWC users (N=12)	Lab/ Clinic	MoCA used in screening	Desktop	Task and reaching time; Movements; IPQ; Usability;	No	Pilot/cross-sectional
ViEW (Morère et al., 2018)	Young CP PWC users (N=12)	Clinic	WST, screening (GMFCS, MACS)	Desktop	Path distance, jerk and amplitude of movements, WST	No	Pilot/RCT elements
PWS (L. Devigne et al., 2017)	Able-bodied (N=9)	Lab	Not reported	Four screens + motion	Commands, collisions, time, QoE (IPQ, NASA-TLX)	No	Feasibility/ pilot
SIMADAPT (Vailland et al., 2020)	Able-bodied (N=29)	Lab	Not reported	VR +3 DoF motion	QoE (NASA-TLX, IPQ and IVEQ)	No	Feasibility
SIMADAPT (Vailland et al., 2021)	Regular PWC users (N=29)	Lab	Not reported	VR + 4/6-DoF motion	Time, Collisions, Head movements, QoE (IPQ, SSQ)	No	Feasibility
SIMADAPT (Fraudet et al., 2024)	Regular PWC users (N=31)	Lab	WST	VR + 4/6-DoF motion	Time, Collision, WST, QoE (NASA-TLX, IPQ, USE, SSQ, Graybiel)	No	Feasibility
Wheelchair-VR (John et al., 2018)	Able-bodied (N=33)	Lab	Not reported	Desktop/ VR (HMD)	Time and SSQ	No	Feasibility/ cross-sectional
Wheelchair-VR(Day & John, 2019)	Able-bodied (N=35)	Lab	Not reported	MR. VR (HMD)	Time, collision and SSQ	No	Feasibility/ cross-sectional
(Zorzi et al., 2023a, 2023b)	Able-bodied (N=14)	Lab	Tasks based on WSTP	VR (HMD)	Time, amplitude of movement, SSQ, IPQ	HR (polar)	Feasibility/ RCT
(Zorzi et al., 2023b)	Able-bodied (N=22)	Lab	WSTP/WST	VR (HMD)	Time, amplitude of movement, SSQ, IPQ	HR (polar)	Feasibility
IndieTrainer (Kenyon et al., 2024)	Children PWC users (N=25)	Clinic	WSC, COPM, ALP	Screen-based	WSC, ALP, COPM and CSQ-8	No	Small scale clinical trial
WheelUp! (M. Chen et al., 2023)	Experienced PWC users (N=3)	Lab	Not reported	desktop/ VR (HMD)	Collision, movements, map locations, usability	No	Feasibility/ pilot
(Younis et al., 2019)	Able-bodied (N=16)	Lab	Not reported	BCI-/Desktop/ VR (HMD)	Time, number of tasks completed, Usability	No	Feasibility
(Rodriguez, 2018)	Children w/ disabilities (N= not reported)	Lab	Not reported	Desktop	Not reported	No	Exploratory
E-WATS (F. R. Martins et al., 2022)	PWC users (N=4)	Lab/ Clinic	Tasks based on PMRT	Desktop	Time, commands, collisions and usability	No	Feasibility/ pilot
E-WATS (Valentini et al., 2024)	PWC users (N=6)	Lab/ Clinic	PMRT	Desktop	Time, commands, collisions and PMRT	No	Feasibility/ pilot
E-WATS This thesis	Able-bodied (N=67), PWC users (N=10)	Lab/ Clinic	Task (WST + PMRT), WST-Q, and MoCA	Desktop/ VR (HMD);	Perf. + Subj. (SUS, IPQ, SAM, SSQ) + Physio. (EDA, HR/HRV)	EDA, HR, HRV and ACC (E4)	Pilot Feasibility/ cross-sectional

2.2.1.2 Software Features: Development Environments and Simulator-based outcome Measures

The software layer integrates inputs, physics, task logic, rendering, and data logging. Game engines (Unity/Unreal) have improved extensibility and fidelity however, many studies still employ study-specific tasks and non-standardised outcome sets, which complicates cross-study comparison and clinical interpretation (A. R. de Sá et al., 2022; Arlati et al., 2020). Reported systems vary widely in immersion, realism, and data resolution, ranging from 2D bird's-eye interfaces to fully immersive 3D simulations with real-time logging and feedback.

An important consideration is the software development environment used to build the simulator. Recent systems often leverage general-purpose game engines (Unity3D, Unreal Engine) that offer high graphical fidelity, configurable physics, and modular sensor integration. Earlier systems used proprietary or bespoke platforms (e.g., VRSIM-2 in Vega Prime; ViEW in 3DVIA Virtools). The selected environment influences extensibility, interface design, and the ability to support real-time data capture and feedback.

A second consideration is the type and resolution of simulator-derived outcome measures. Across the powered wheelchair simulators, these metrics include system-level data such as joystick input, trajectory tracking, time to complete tasks and number of collisions. The availability and relevance of these data points are closely tied to the simulator's design intent, task complexity, and the fidelity of its simulation engine. Early systems such as VAHM2 (Hafid Niniss & A Nadif, 2000) and VRSIM (Spaeth et al., 2008) captured basic joystick commands, angular velocities, or RMSE for assessing motor behaviour. As simulator capabilities expanded, systems like VRSIM-2 (Mahajan, 2012), SimCadRom (Hernandez-Ossa et al., 2017), and AccessSim (Goncalves et al., 2014) incorporated more detailed kinematic data including linear/angular velocities, control command frequency, and trajectory boundary violations.

Notably, VRSIM-2 (Mahajan, 2012) contributed to the Quantitative Driving Metrics (QDM) framework, which formalised performance indicators (time, velocity, RMSE). Limitations of a performance-only paradigm remain, taken in isolation, these indicators can be difficult for clinicians to map to functional goals or safety and are infrequently paired with validated clinical tools, tolerability measures, or implicit indices of user state.

More recent simulators, including E-WATS (F. R. Martins, 2017), (Zorzi et al., 2023a, 2023b) and WheelUp! (M. Chen et al., 2023), also integrate quantitative metrics and explore additional sensors metrics. For example, Zorzi et al. (Zorzi et al., 2023a) log IMU-based joystick motion and explores the

use of polar-based heart rate metrics. WheelUp! (M. Chen et al., 2023) saves movement direction, XY axis location, rotation, and collisions, and offers replay features for session-level analysis. In this thesis, performance logs are combined with QoE instruments (SUS, IPQ, SAM, SSQ) and wearable physiology (EDA, HR/HRV) under both monitor and HMD conditions, to characterise explicit (performance/QoE) and implicit (physiological) responses that support clinical interpretation of feasibility, tolerability, and progression.

Not all simulators emphasize quantitative outputs. Several feasibility or co-design-focused systems, such as WST (Nunnerley et al., 2017), miWe-CC (P. S. Archambault et al., 2013), and IndieTrainer (Kenyon et al., 2024), rely primarily on qualitative evaluation, clinician observation, or external assessments (e.g., Wheelchair Skills Checklist (WSC) (Butler et al., 1984; Gefen et al., 2022; Kenyon et al., 2024), Assessment of Learning Powered Mobility (ALP) (Kenyon et al., 2024; Nilsson & Durkin, 2014), Power Mobility Program (PMP) (Furumasu et al., 1996; Gefen et al., 2022; Kenyon et al., 2024), Canadian Occupational Performance Measure (COPM) and Client Satisfaction Questionnaire-8 (CSQ-8) with little or no reporting of objective variables recorded automatically by the simulator during tasks (e.g., task duration, collision events, trajectory deviation, speed profiles, joystick command traces). These studies have advanced participatory design; however, the absence of embedded quantitative tracking linked to clinical anchors limits automated feedback, within- and between-session monitoring, and cross-study comparability.

Overall, simulator software components vary substantially depending on the application context, whether exploratory, clinical, or research oriented. This diversity reinforces the need for standardized approaches to simulator-based data collection and transparent reporting of simulator architecture and evaluation tools.

In light of the heterogeneity across systems and outcome sets, this thesis develops and documents a reproducible evaluation protocol that uses WST/PMRT-aligned tasks, quality-of-experience instruments (SUS, IPQ, NASA-TLX, SAM, SSQ), and wearable physiology (EDA, HR, HRV) under both desktop and VR head-mounted display conditions. It also presents Table 2.7 as a concise literature summary that situates the present work by reporting, for each included study, the population, setting, clinical tools, immersion (display type), outcome measures, use of physiological signals, and study design. The aim is to improve comparability and interpretability and to support replication in clinical and research settings. The protocol is offered as an adaptable example rather than a standard and is limited here to desktop and VR HMD conditions; evaluation of motion platforms and dedicated haptics lies outside the present scope. Progress toward any formal standard would require broader community consensus and multi-site validation.

2.2.2 Overview of Main Developments

Early systems, such as VAHM2 (Hafid Niniss & A Nadif, 2000), established foundational features still found in more recent simulators. These included configurable driving modes (manual, semi-autonomous, and autonomous), virtual obstacle detection, and the use of head-mounted displays (HMDs). Though limited by early hardware, VAHM2 demonstrated how virtual tasks could support safer, more accessible training for users with disabilities.

Subsequent simulators like VRSIM (Spaeth et al., 2008) and its successor VRSIM-2 (Mahajan, 2012) transitioned from 2D top-down views to immersive 3D environments. These platforms implemented structured assessments based on the Power Mobility Road Test (PMRT), capturing quantitative driving metrics such as RMSE, collision count, and joystick velocities. This work led to the development of the Quantitative Driving Metrics (QDM) framework by (Kamaraj, 2020)., which formalized performance indicators for evaluating virtual driving proficiency. However, most cohorts were modest and lab-only, and the metrics were rarely paired with validated clinical tools or user tolerability measures, limiting transfer to practice.

The ViEW simulator (Morère et al., 2011, 2018) focused on evaluating and training driving skills in users with cerebral palsy. Developed with 3DVIA Virtools and 3D Studio Max, it offered a joystick-controlled experience across seven levels of increasing complexity. Key contributions included quantitative metrics like jerk and joystick amplitude. Later studies (Zatla et al., 2015, 2018) applied the Optimal Preview Control Model (OPCM) to analyse gaze behaviour and trajectory planning. Despite promising results, challenges remain regarding ecological validity and broader accessibility. Strengths include a clinical cohort and gaze/trajectory analysis; however, bespoke tasks and condition-specific sampling constrain generalisability and guideline development.

PWCsim (Alshaer et al., 2013, 2015) investigated the impact of field of view (FOV), stereoscopic vision, and visual joystick representation on user performance and perception. Conducted with healthy participants, the studies revealed that wider FOVs improved performance and presence. Findings also showed that visual realism of the joystick interface influenced performance more than physical configuration, raising important questions about perceptual fidelity in simulator design. These perception–performance links are informative, but clinical benefit or transfer in wheelchair users remains under-evidenced.

Simulators like miWe (P. Archambault et al., 2016; P. S. Archambault et al., 2008, 2011, 2012, 2017; Bigras et al., 2019; Faure et al., 2023; Tao & Archambault, 2015, 2016) and its variants (miWe-C (Gefen et al., 2019, 2022), miWe-CC (P. S. Archambault et al., 2013; Torkia et al., 2019)) pioneered the use of

motion platforms and explored subjective fidelity through metrics such as presence and task realism. Developed using Unreal Engine and Unity3D, these simulators were validated through comparisons with real-world driving tasks in both adult and paediatric populations. The use of full-motion feedback and participatory evaluations positioned miWe as a benchmark in human factors and wheelchair simulation research. However, cybersickness and cost/complexity remain barriers; multi-site validation and standard task sets are still scarce.

A subset of simulators explored broader accessibility or alternative applications. AccessSim (Goncalves et al., 2014), for instance, aimed to evaluate environmental accessibility using WSTP-based scenarios rather than user performance. PhyMel_WS (Panadero et al., 2014a) presented a unique case, designed for awareness training rather than skill development. It used a motion platform and emotional learning framework to expose university students to urban accessibility barriers. Though not validated for clinical use, it underscored the role of simulator design in supporting empathy and policy education. Wheelchair-Rift (Headleand et al., 2015, 2016) incorporated serious game elements, HMDs, and gesture tracking with Leap Motion to enhance immersion. Simulators such as SimCadRom (Hernandez-Ossa et al., 2017, 2020) focused on adaptive interfaces for users with severe physical impairments, including eye-tracking and simplified command inputs. These broaden scope (accessibility, empathy, alternative inputs), yet calibration burden, small-N designs, and limited clinical anchoring restrict adoption.

From a user-centred design perspective, simulators like WTS (Nunnerley et al., 2017) and the PWS and SIMADAPT systems (Fraudet et al., 2024; L. Devigne et al., 2017; Vailland et al., 2019, 2020, 2021) were developed through participatory methods involving clinicians and wheelchair users. The PWS system evolved into a flexible, multisensory platform integrating visual, auditory, haptic, and vestibular feedback via a 4-DoF/6-DoF motion base. Developed in Unity3D with ROS middleware, the system supports diverse input devices (e.g., joystick, head array) and immersive configurations (HMDs, CAVEs). Although technically advanced, the system raised important challenges related to cybersickness, motion realism, and the balance between fidelity and accessibility. Follow-up work refined motion cueing and kinematics; nevertheless, higher fidelity did not consistently translate into better tolerability or clearly actionable clinical outcomes..

Simulators such as (Rodriguez, 2018) emphasized simplicity and customization for children with multiple disabilities. Their Unity-based system included real-world inspired virtual environments and alternative control inputs (switches, eye-tracking). Though exploratory and not formally validated, the project illustrated how simulator design can prioritize cognitive accessibility and clinician-guided task design.

Younis et al.(2019) introduced a Unity-based simulator that allowed powered wheelchair control using a low-cost EEG headset (Emotiv Epoc+). The system supported BCI and joystick modes and evaluated performance across navigation tasks. Despite calibration challenges, results showed that BCI users improved over sessions, demonstrating the feasibility of brain-controlled training in accessible VR environments. These inclusive directions are promising but remain feasibility-focused with unclear pathways to routine clinical use.

More recent systems illustrate a growing interest in gamification and modularity. IndieTrainer (Kenyon et al., 2024) offers a real-world joystick training system for children using the IndieGo® platform and screen-based games aligned with the Assessment of Learning Powered mobility use (ALP) model. Its staged framework supports real-world skill acquisition without requiring immersive hardware, emphasizing cause-effect learning and clinical accessibility. However, reliance on external scales with limited embedded performance logging, screen-only immersion, and small single-site design leave questions about long-term retention and transfer to real-world power wheelchair use.

WheelUp! (M. Chen et al., 2023) is an open-source simulator developed in Unreal Engine 5 with both monitor and VR modes. It includes photorealistic environments, checkpoint guidance, and joystick input support. Early pilot testing with experienced wheelchair users found high acceptability and embodiment. However, further validation with novice users and refinement of motion mismatch effects is needed.

Zorzi et al. (2023a) proposed a standardized and cost-effective simulator architecture for powered wheelchair training. Built in Unity3D and integrated with ROS, the system supports different hardware configurations and task types. Their study highlighted the need for harmonized development practices, focusing on simulation fidelity, scenario realism, and user adaptability as part of a standardized training toolset. Engagement and tooling have improved, but reliance on able-bodied samples, short-term lab protocols, and limited clinical anchoring constrain claims about generalisability and standardisation.

Finally, E-WATS (F. R. Martins, 2017; F. R. Martins et al., 2022; Valentini et al., 2024) is a modular simulator originally developed for desktop use with joystick-based daily activity scenarios. It purposes to be integrated into a clinical assessment and training protocols for individuals with spinal cord injury, E-WATS provides metrics such as joystick commands, time, collisions during structured scenarios (ramp, obstacle course, elevator) and deviation from the optimal path. While initial studies demonstrated feasibility, limitations were identified in spatial awareness and user comfort. This thesis builds upon E-WATS by extending it to immersive HMD environments, integrating QoE-based

assessments, and incorporating physiological monitoring, thus addressing fidelity, emotional engagement, and individualized user evaluation.

Table 2.7 presents representative powered wheelchair simulator systems (2016–2024) and associated studies, positioning the current work within a spectrum that spans training, assessment, clinical planning, interface prototyping, and inclusive design.

In summary, wheelchair simulators have progressed from basic feasibility tools toward platforms that can support personalised training, clinically aligned assessment, and design exploration. Beyond task-load evaluation, this thesis contributes by (i) mapping simulator tasks to established clinical frameworks (WST/PMRT), (ii) combining performance, QoE, and physiology to characterise both explicit and implicit user responses, (iii) comparing immersion modalities (desktop vs HMD) with a focus on tolerability and feasibility, and (iv) documenting a reproducible, clinically interpretable protocol that can be adopted or adapted in future work.

2.2.2.1 Wheelchair Simulator Studies and Outcomes Measures

This section includes studies published between 2014 and 2024, reflecting a decade of technological growth in immersive VR, HMDs, and sensor-based interaction. The focus is on studies that employed quantitative or mixed-method evaluation methods, clearly described simulator tasks, and reported structured outcome measures. These studies go beyond proof-of-concept or user impressions by incorporating protocols and metrics relevant to performance assessment and training and clinical utility.

2.2.2.1.1 Simulator Tasks

While Powered wheelchair simulators are commonly framed as tools for training and/or assessment. However, only a subset of studies specify task parameters with sufficient granularity (e.g., path geometry, doorway width, ramp grade, time limits) and link them to defined outcomes. Typical tasks model everyday navigation challenges such as obstacle avoidance, doorway traversal, ramp negotiation, and tight-space manoeuvring. A smaller subset explicitly derives or maps tasks from established clinical frameworks, notably WST/WSTP (Dalhousie University, 2023; Kirby, 2017) or PMRT (Massengale et al., 2005), but even in these cases, implementations often use subsets, adaptations, or approximations rather than the full protocol, and reporting can omit the exact task constraints.

2.2.2.1.2 Outcome Measurement Tools

The most common quantitative metrics include task completion time, number of collisions, joystick input data, root-mean-square error (RMSE), and counts of directional commands. Several studies complemented these with subjective assessments using tools such as: NASA-TLX for perceived cognitive workload (Hart, 2006; Hart & Staveland, 1988), Simulator Sickness Questionnaire (SSQ) for discomfort and disorientation (Balk et al., 2013; Kennedy et al., 1993) and Igroup Presence Questionnaire (IPQ) for evaluating immersion and spatial presence (Igroup Project Consortium, 2015).

Based on Table 2.7, IPQ is reported in (Fraudet et al., 2024; L. Devigne et al., 2017; Tao & Archambault, 2016; Vailland et al., 2020, 2021; Zorzi et al., 2023a, 2023b) used the IPQ to assess the sense of presence (SoP), validating the importance of immersion and realism in VR-based mobility training.

The SSQ was employed in studies by (Day & John, 2019; Fraudet et al., 2024; John et al., 2018; Vailland et al., 2020, 2021; Zorzi et al., 2023a, 2023b) demonstrating growing attention to cybersickness and user tolerance. In this thesis, SSQ is used not only for evaluation across desktop and HMD conditions but also to quantify the effect of a comfort-oriented display setting on cybersickness, with the aim of reducing symptoms while preserving task performance and presence.

NASA-TLX was adopted by (Fraudet et al., 2024; Kamaraj, Dicianno, et al., 2016b; L. Devigne et al., 2017; Vailland et al., 2020) to quantify mental demand during simulator-based tasks, reaffirming its suitability in assessing cognitive load in wheelchair simulator contexts. However, it is rarely analysed together with physiology, nor is it commonly incorporated into predefined progression rules (e.g., criteria that determine when to increase task complexity or move from desktop to HMD based on performance, tolerability, and safety signals).

Cognitive screening is rarely reported. Notably, (Tao & Archambault, 2016) was the only study to administer a validated cognitive instrument, the Montreal Cognitive Assessment (MoCA) (Zorzi et al., 2023a, 2023b) as part of participant screening to examine the influence of cognitive status on virtual navigation and reaching performance in experienced wheelchair users.

In terms of incorporating physiological data to assess user states, only two studies investigate, (Zorzi et al., 2023a, 2023b) included heart rate and the E-WATS studies presented in this thesis, systematically integrated wearable-based monitoring (e.g., HR, HRV, EDA) as a method for assessing user arousal, stress, and workload.

2.2.2.1.3 Use of Clinical Assessment Frameworks and Ecological Validity

Only seven studies integrated elements of structured clinical frameworks such as the PMRT (Massengale et al., 2005) or WST/WSTP (Dalhousie University, 2023; Kirby, 2017): (Fraudet et al., 2024; Kamaraj, Dicianno, et al., 2016a, 2016b; Kamaraj, Mahajan, et al., 2016; F. R. Martins et al., 2022; Morère et al., 2018; Tao & Archambault, 2016; Valentini et al., 2024; Zorzi et al., 2023a, 2023b). Their use reflects a deliberate effort to bridge simulation environments with validated mobility assessments

Ecological validity was reinforced in ten studies that involved powered wheelchair users: (M. Chen et al., 2023; Fraudet et al., 2024; Kamaraj, Dicianno, et al., 2016b, 2016a; Kenyon et al., 2024; F. R. Martins et al., 2022; Morère et al., 2018; Tao & Archambault, 2016; Vailland et al., 2021; Valentini et al., 2024). Among these, seven of these were clearly mentioned that were conducted within clinical or rehabilitation contexts (see Table 2.7). These studies demonstrate progress toward integrating simulator use in practice. However, heterogeneity across the study designs, small sample sizes, and lack of alignment with clinical standards persist.

2.2.3 Literature Reviews on Wheelchair Simulators

Research on WSs has expanded considerably over the past two decades, driven by advances in immersive technology and a growing recognition of the need for safe and standardised training environments. Three recent reviews (A. R. de Sá et al., 2022; Arlati et al., 2020; Zorzi et al., 2024) have synthesized the state of research and development on wheelchair simulators, each offering a unique perspective on the field's evolution, current challenges, and future directions.

2.2.3.1 Review Paper 1 – Scoping Review on Wheelchair Simulators Studies focusing on Sense of Presence factor (SoP)

A scoping review by (Arlati et al., 2020) identified 62 papers describing 29 unique wheelchair WSs, focusing particularly on VR-based simulators and the concept of sense of presence (SoP) as a potential factor influencing simulator effectiveness. SoP, defined as the degree to which users feel immersed in the virtual environment (Witmer & Singer, 1998), was investigated across different feedback modalities, including visual, auditory, haptic, and vestibular feedback (P. S. Archambault et al., 2012; Crichlow et al., 2012; Panadero et al., 2014b). While some systems used full-motion platforms and stereoscopic displays (Crichlow et al., 2012; Panadero et al., 2014b), others achieved adequate SoP using only desktop monitors (P. S. Archambault et al., 2012), suggesting that factors beyond multisensory input, such as task realism and user motivation, also contribute to presence (Baños et al., 2004; Nichols & Patel, 2002).

Arlati et al. review also highlighted a high heterogeneity in both the hardware and software components used, such as joystick controls, projection screens, HMDs, motion platforms, and various interaction devices, which made it difficult to determine which technologies were most effective in promoting SoP (Alshaer et al., 2013, 2016; Mahajan, 2012). Although, many individual studies reported improvements in performance and engagement, most were feasibility or pilot studies, with only a small subset employing randomized or controlled study designs (Arlati et al., 2020). As a result, this review emphasized the lack of standardized/structured study protocols and a limited body of high-level evidence from adequately powered, controlled trials, recommending further controlled clinical studies to assess simulator components and training strategies (Arlati et al., 2020; Sonar et al., 2005).

2.2.3.2 Review Paper 2 – Survey on Skills Assessment Metrics on Powered Wheelchair Simulators Studies

(A. R. de Sá et al., 2022) conducted a targeted survey that examined skills assessment metrics in studies using electric-powered wheelchair (EPW) simulators within virtual environments, reviewing 42 studies published from 2000 to 2020. The primary objective was to categorise and synthesise the metrics used to evaluate EPW driving in virtual environments. Among the 42 studies, the authors identified 29 quantitative and 3 qualitative performance parameters, at the field level; including task completion time, number of collisions, path deviation, and joystick movement patterns (A. R. de Sá et al., 2022; P. S. Archambault et al., 2008; Inman et al., 2011; Liu et al., 2014; Mahajan et al., 2013; Spaeth et al., 2008). Notably, 21 of the studies were published between 2016 and 2020, indicating a rising interest in PWSs field (A. R. de Sá et al., 2022).

Despite the diversity of metrics, the review emphasized a lack of standardisation in task design, metric selection, and reporting approaches, which limited comparability across studies, even for common metrics such as task completion time (A. R. de Sá et al., 2022). While several studies included repeated tasks, only a few reported user performances across multiple sessions, limiting the ability to assess learning effects or skill retention.

Some studies included qualitative measures such as expert observation evaluation or user feedback; These were always used in conjunction with quantitative metrics and were not applied in isolation (A. R. de Sá et al., 2022). Moreover, none of these reviewed studies systematically validated simulator-derived metrics against real-world performance or clinical assessment tools. Although four studies incorporating task elements inspired by PMRT (A. R. de Sá et al., 2022; P. S. Archambault et al., 2008; Kamaraj, Dicianno, et al., 2016a; Mahajan et al., 2013; Spaeth et al., 2008). This disconnection from

clinical benchmarks, alongside the lack of integration into clinical workflows, underscores barriers to the broader adoption of simulators in rehabilitation and prescription contexts (A. R. de Sá et al., 2022).

2.2.3.3 Review Paper 3 – Literature Review on Wheelchair Simulators Studies for Training Applications

A more recent review by (Zorzi et al., 2024) conducted a comprehensive review of 28 studies published between 2017 and 2024 focused on the use of virtual reality (VR) simulators for wheelchair skills training. The review aimed to examine how immersive VR could supplement conventional training practices by identifying limitations in current real-world approaches and by analysing the core features and methodologies of existing simulation systems.

Most of the included studies targeted powered wheelchair users and employed joystick-based input within either simplified or ecologically valid virtual environments. Performance-related outcome measures were prevalent, particularly task completion time, number of collisions, and joystick trajectory metrics. These were frequently paired with subjective assessments of user experience, such as usability (e.g., Ease-of-Use Questionnaire (USE), Short Feedback Questionnaire (SFQ) and User Experience Questionnaire (UEQ)), presence and immersion (e.g., IPQ), cognitive workload (e.g., NASA-TLX), and simulator-induced discomfort (e.g., SSQ). Some studies also utilized qualitative approaches like interviews or open-ended feedback to complement quantitative data.

Most of studies employed performance-based metrics, such as task completion time, number of collisions, and joystick movement patterns. These were frequently paired with subjective assessments of presence, usability, satisfaction, or simulator-induced discomfort. Common instruments included instruments to assess immersion/sense of presence (e.g., IPQ) and cybersickness (e.g., SSQ). A smaller number of studies assessed cognitive workload using NASA-TLX, and usability or satisfaction was measured using tools such as the Ease-of-Use Questionnaire (USE), Short Feedback Questionnaire (SFQ) and User Experience Questionnaire (UEQ). In several cases, custom interview guides or open-ended feedback forms were also employed.

However, emphasized key limitations: the frequent occurrence of cybersickness, small and often able-bodied samples, inter-study heterogeneity in outcome batteries, and limited longitudinal follow-up for skill retention. The review also noted that very few studies incorporated clinical assessment frameworks at the protocol level (e.g., clinician-led tailoring, structured progression); studies outside this scope (earlier or non-immersive) that mapped tasks to WST/PMRT were therefore not included in the review analysis. Although two studies explored physiological as implicit user metrics, the review highlighted this area as underdeveloped and inconsistent, an issue already outlined in this chapter. As

such, Zorzi et al. called for more robust and multidimensional evaluation methodologies that align simulator design with clinical objectives and real-world performance indicators.

2.2.4 Synthesis of Review Findings on Wheelchair Simulator Studies

Together, the three reviews offer a complementary, but not exhaustive, account of the field. Arlati et al. (2020) emphasized the importance of immersion/Sense of Presence and system architecture; A. R. de Sá et al. (2022) examined how power mobility skills and performance are quantified; and (Zorzi et al., 2024) analysed the feasibility of using wheelchair simulators as training tools.

Despite their different emphases, all three reviews pointed to the same overarching challenges: a lack of methodological standardisation, weak alignment with clinical assessment tools/guidelines, limited use of diverse wheelchair users' groups, and a general absence of longitudinal or high-quality controlled studies. Each review calls for greater standardization in simulator design and evaluation, and for stronger validation of outcome metrics against real-world practices. **Error! Reference source not found.** provides a comparative summary of the focus, findings, and recommendations of these three review papers.

The next section builds on these insights by summarising the core findings of this chapter and linking them to the central aim of this thesis: the development of a multidimensional, clinically relevant framework for evaluating powered wheelchair simulator use. This transition underscores the importance of bridging methodological gaps with structured, scalable, and user-centred solutions.

Table 2.8: Summary of key literature reviews on wheelchair simulators

Review	Time frame/ Scope	Primary Focus	Key Findings	Identified Gaps	Recommendations
Arlati et al. (2020)	62 studies on 29 WS systems (2000 to 2018).	Sense of Presence (SoP); influence of multisensory feedback (VR system design).	SoP influenced by visual, haptic, auditory, and vestibular cues; high hardware/software heterogeneity; performance gains often linked to immersive setups.	High heterogeneity in hardware/software; few controlled trials; lack of standardised protocols for assessing SoP and training impact.	Conduct controlled trials; systematically test components (e.g., display type, feedback modalities).
Angela de Sá et al. (2022)	42 studies (2000–2020) on EPW simulators.	Performance metrics in EPW simulators.	Identified 29 quantitative and 3 qualitative metrics; frequent use of task time, collisions, and joystick data; rising publication trend from 2016 onward.	No standardisation in metric design or task protocol; limited use of repeated tasks; no validation against clinical tools like PMRT.	Link VR metrics to clinical assessments (e.g., PMRT); improve reporting standards; include multi-session tracking; exploration on physiological signals applications in PWSs.
Zorzi et al. (2024)	28 studies (2017–2024) on VR for wheelchair training.	VR as complement to WSTP and real-world training.	Frequent use of IPQ, SSQ; subset included biomarkers(HR, HRV, EDA, skin temp) for presence and stress monitoring.	Noted lack of standardisation, frequent cybersickness, and underuse of physiological data; recommended structured evaluation approaches; limited inclusion of powered wheelchair users;	Adopt structured training protocols (task designs); mitigate cybersickness; explore VR for cognitive/emotional rehab applications, incorporate physiological monitoring;

Summary and Rationale for This Thesis Approach

In summary, simulator-based evaluation methods encompass a growing array of performance metrics, subjective responses, and more recently, physiological signals. Despite this evolution, methodological limitations remain, including:

I. Lack of standardised task protocols and measurements strategies
II. Limited inclusion of diverse user groups and real-world clinical contexts studies.
III. Adaptive or personalised features based on user needs are reported in a minority of studies and are inconsistently defined, implemented, and evaluated.

Addressing these gaps is critical to improving the reliability, clinical validity, and translational potential of simulator-based powered mobility assessments. These gaps form the foundation for the present thesis, which proposes a multidimensional, user-centred evaluation framework including a clinical protocol that integrates subjective experience, objective metrics, and physiological monitoring. The goal is to develop a framework that is not only adaptable and scalable but also grounded in clinical relevance and usability.

To complement this analysis of individual simulator studies and their metrics, the next section examines how the field has been synthesised in recent review articles. These reviews provide a broader perspective on methodological trends, gaps, and priorities in simulator-based assessment research.

2.2.5 Summary

This section critically examined the evolution of powered wheelchair simulator (PWS) research by integrating insights from both empirical studies and literature reviews. Section 2.2.1 highlighted how technological innovation, particularly in immersive displays, input devices, and virtual environments, has driven simulator development, while also revealing persistent inconsistencies in study design, evaluation practices, and clinical integration. Section 2.2.2 reviewed the variability in PWS studies, task structures, and outcome measures across recent WS implementations. Only a limited number of PWS tried to incorporate validated clinical assessments such as the PMRT, WST/WSTP, or employed physiological monitoring to assess user responses. Similarly, few systems offered adaptive input interfaces or personalized task adjustment to accommodate user diversity.

Section 2.2.3 synthesised three reviews. Arlati et al. (2020) emphasized heterogeneity in system design and the need for controlled studies on sense of presence. A. R. de Sá et al. (2022) identified a lack of standardization in performance metrics and limited alignment with clinical benchmarks such

as the PMRT. Zorzi et al. (2024) focused on how virtual reality can enhance training, but highlighted gaps related to tasks structure, underuse of physiological signals, and persistent cybersickness issue.

Together, these findings demonstrate a growing recognition of the importance of multidimensional evaluation in powered wheelchair simulator research, with several studies incorporating both performance-based metrics and user-reported outcomes. However, these components are often applied independently, without being integrated into a structured or clinically grounded framework.

To address the limitations identified in simulator design, task standardization, and evaluation practices, it is essential to better understand how users perceive, interact with and respond to virtual wheelchair simulators.

Key findings from this chapter include:

- Joystick input, HMDs, and Unity-based platforms are the dominant technical configurations.
- Most studies emphasize training and usability rather than clinical assessment or prescription workflows.
- Outcome metrics commonly include task performance and self-reported experience; few use physiological signals or behaviour-based indicators.
- There is a lack of standardised task protocols, validated clinical outcome measures, and alignment with real-world practice.
- Diversity in user inclusion, regulatory considerations, and clinician involvement remains limited.
- The field urgently requires structured, flexible frameworks that are both user-informed and clinically relevant.

The next chapter introduces the range of user response modalities employed in existing research, focusing on both explicit responses (e.g., questionnaires, interviews) and implicit indicators (e.g., physiological signals, behavioural data). Analysed through the lenses of Quality of Experience (QoE) and Cognitive Load Theory (CLT), these responses offer a more comprehensive understanding of how wheelchair simulators are experienced and how they might be tailored to support real-world training and clinical decision-making.

2.3 User Response Assessment in Simulator Studies

As discussed in the previous section, research on powered wheelchair simulators (PWSs) has expanded significantly. However, persistent limitations remain in how user experience and performance are evaluated, largely due to the challenges of implementing standardized assessments in a field marked by high user variability, differences in simulator design, and the inherently subjective nature of many evaluation methods. Most systems rely on quantitative performance metrics and/or qualitative feedback collected through interviews or usability questionnaires (A. R. de Sá et al., 2022; Arlati et al., 2020; Zorzi et al., 2024). While informative, these indicators alone offer limited insight into the cognitive and experiential demands placed on users, particularly those with diverse physical or cognitive abilities.

To address these limitations, this thesis proposes a multidimensional evaluation strategy, that captures both internal cognitive states and the overall quality of interaction with the simulator. De Sá et al. (2022) highlighted that future work should explore QoE assessment methods to better capture user experience in wheelchair simulators. Similarly, studies reviewed by Arlati et al. (2020) and Zorzi et al. (2024) noted the incorporation of cognitive load assessment into evaluation approaches. Building on these insights, this thesis integrates both Cognitive Load Theory (CLT) and Quality of Experience (QoE) to provide a structured framework capable of addressing these gaps.

This section introduces such two key frameworks: Cognitive Load Theory (CLT) (Sweller, 1994, 2011, 2018) and Quality of Experience (QoE) (Bañuelos-Lozoya et al., 2021; Callet et al., 2013). CLT focuses on the mental effort required to complete tasks, while QoE addresses usability, sense of presence/immersion, emotional response, and overall satisfaction. Together, these frameworks can help to guide wheelchair simulator design that is not only effective but also cognitively accessible and user-friendly.

The section is structured as follows:

- Section 2.3.1 outlines the theoretical foundation of CLT and QoE as complementary methodologies.
- Section2.3.2 details CLT, including its application to simulator tasks and related assessment principles.
- Section 0 focus on QoE, defining its key dimensions in virtual reality contexts, present both subjective and implicit methods for QoE assessment.

Together, these sections form a user-centred evaluation framework to support simulator design, personalization, and clinical integration.

2.3.1 Theoretical Foundations

Effective assessment of user experience in wheelchair simulators requires a clear theoretical grounding. This thesis adopts an approach informed by two well-established frameworks: CLT (Sweller, 1994) and QoE (Callet et al., 2013). These frameworks are not mutually exclusive; indeed, they often intersect, and their integration supports a more comprehensive understanding of how users interact with, and respond to, simulation environments (Bañuelos-Lozoya et al., 2021).

CLT, grounded in instructional design and cognitive psychology, provides insight into how limited working memory resources are distributed during task performance. It distinguishes between in intrinsic load (task complexity), extraneous load (design-related burden), and germane load (effort devoted to learning). CLT has been widely applied in simulation-based training to structure task demands and reduce cognitive overload (Bañuelos-Lozoya et al., 2021; Reedy, 2015).

QoE, by contrast, addresses the user's subjective evaluation of their interaction with a system. It incorporates both system-level factors (e.g. usability and immersion) and individual characteristics such as expectations, affective state, and context of use (Bañuelos-Lozoya et al., 2021; Callet et al., 2013; Perkis et al., 2020). QoE expands beyond usability by capturing emotional and contextual dimensions that shape the overall quality of interaction.

By integrating CLT and QoE, researchers and developers can achieve a more holistic perspective on simulator design evaluation. For example, CLT provides insights into the cognitive demands and task complexity, while QoE accounts for user satisfaction, emotional engagement, and perceived usability. This integrated view facilitates:

- **Personalised adaptation** by identifying user-specific cognitive and experiential thresholds, enabling real-time or post-session adjustments to simulator difficulty or feedback.
- **Improved accessibility** by revealing experiential or cognitive barriers that may not be evident through performance metrics alone, guiding inclusive design.
- **Clinical relevance** by ensuring that simulator-based assessments and training are not only effective but also usable, engaging, and acceptable in real-world rehabilitation or clinical settings.

The following sections present each framework in more detail, emphasizing their relevance to simulator evaluation.

2.3.2 Cognitive Load Theory in Simulator-Based Tasks

As mentioned in the previous section, Cognitive Load Theory (CLT) provides a psychological and instructional framework for understanding the mental effort exerted during simulator-based tasks. Developed by John Sweller in the context of instructional design and problem-solving research, CLT posits that the human working memory has a limited processing capacity (Sweller, 1994, 2011). When task demands exceed this capacity, learning and performance can be compromised. Effective system or instructional design should therefore aim to optimize how cognitive (mental) resources are allocated during task engagement. CLT distinguishes three load types (F. Paas et al., 2003; Sweller, 2011):

- Intrinsic load: mental effort from inherent task complexity, shaped by interacting elements and prior knowledge. Example: navigating a slalom course with joystick control requires coordination of spatial orientation, turning, and velocity.
- Germane load: cognitive effort dedicated to learning and schema construction. Example: analysing joystick sensitivity to refine strategy or aligning the wheelchair before a ramp.
- Extraneous load: unnecessary effort caused by poor design. Example: delayed feedback or ambiguous instructions divert attention from task objectives.

These loads correspond to how information moves from sensory memory to working memory, and eventually into long-term memory. Maintaining balance across them is essential for learning, skill retention, and schema development (Agency for Healthcare Research and Quality, 2024). Figure 2.3 illustrates how these types of cognitive load contribute to total working memory use. Effective simulator design should aim to reduce extraneous load (e.g., confusing visuals, unclear instructions), simplify intrinsic load (e.g., by adjusting task complexity), and maximize germane load (by promoting meaningful learning and engagement). Optimising these components free up cognitive capacity and supports more effective training.

In simulation-based environments, particularly those designed for mobility training and assessment, these types of load manifest in distinct ways. For instance, intrinsic load may be generated when users navigate tight spaces, align virtual wheelchairs to ramps, or interpret unfamiliar control dynamics. Extraneous load can arise from mismatches between user expectations and system responses-such as latency in control feedback, cluttered visual scenes, or inconsistent sensor mappings. Germane load is more prominent in repeated training sessions; it reflects how well the simulator supports learning by enabling users to develop transferable strategies and motor schemas over time (F. G. W. C. Paas & Van Merriënboer, 1994b, 1994a).

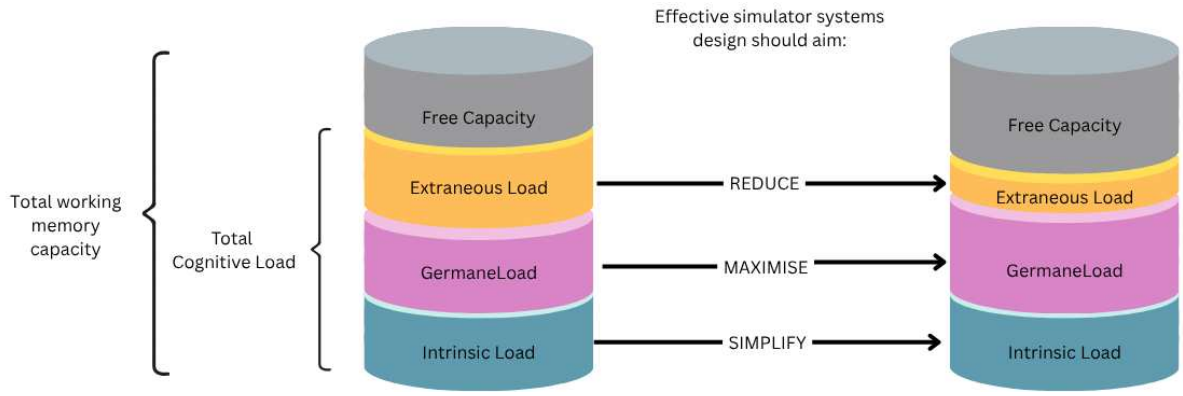


Figure 2.3: Cognitive Load Theory applied to simulator system design. Image adapted from (Krieglstein et al., 2022; Mancinetti et al., 2019; Young et al., 2018).

As simulators become more immersive, CLT grows increasingly relevant in HCI and VR design (N. Hollender, C. Hofmann, M. Deneke, 2010). Such environments place high perceptual and attentional demands, which can lead to overload, particularly for users with brain injury, cognitive impairments, or attentional challenges (Cinaz et al., 2010). For these users, excessive demand may result in disengagement or frustration. Moreover, the same task may impose different loads depending on prior experience, cognitive capacity, or age (F. Paas et al., 2001), underscoring the need for personalised design and careful interpretation of performance across diverse populations. To balance personalisation with methodological rigour, this thesis preserves standardisation at the level of task definitions, instruments, scoring, and reporting, while allowing limited, pre-specified parameter adjustments (e.g., speed scaling, visual complexity, assistance levels) governed by explicit progression rules, thereby maintaining comparability across participants and studies.

Understanding cognitive load within simulator use helps tailor task complexity, pacing, and interface design to user needs. It also informs the selection of appropriate assessment tools and provides a foundation for evaluating whether simulator-based training is educationally effective, cognitively appropriate, and clinically relevant.

2.3.2.1 Cognitive Load Assessment Methods

Cognitive load is a latent psychological construct[1] that reflects internal information processing during task execution (Moreno & Park, 2010). Because it is not directly observable, valid and reliable measurement methods are essential, particularly in wheelchair simulator studies where users may

[1] A latent psychological construct refers to a mental or psychological process that cannot be directly measured or observed but must be inferred through indirect indicators such as behavioural performance, self-reported effort, or physiological signals (Moreno & Park, 2010).

present motor or cognitive impairments. In such contexts, appropriate assessment is critical to ensure usability, accessibility, and task suitability. A widely accepted classification, proposed by (Brünken et al., 2003), organizes cognitive load assessment methods along two key dimensions: objectivity (subjective vs. objective) and causal relationship (direct vs. indirect). This leads to four methodological categories as shown at Table 2.9.

Table 2.9: Classification of Cognitive Load Assessment Methods by Objectivity and Causal Relationship adapted from (Brünken et al., 2003)

Objectivity	Indirect	Direct
Subjective	Self-reported mental effort questionnaires (e.g., NASA-TLX , Paas scale).	Perceived task difficulty ratings or mental stress reports.
Objective	Performance metrics (e.g., errors, completion time; physiological signals (e.g., HR, GSR, eye tracking).	Dual-task paradigm; neurophysiological imaging (e.g., EEG ,fMRI).

These methods vary in complexity, sensitivity, and applicability. Subjective indirect tools such as NASA-TLX are widely used due to their simplicity and multidimensionality, covering mental, physical, and temporal demands as well as effort, performance, and frustration. The Paas scale offers a concise, unidimensional alternative (F. Paas et al., 2003). However, subjective measures depend on introspection and verbal reporting, which may be limited in users with reduced communication or cognitive insight. To address these limitations, many studies employ physiological signals as objective and implicit indicators

- Heart Rate (HR) and Heart Rate Variability (HRV) – associated with autonomic nervous system activity and task complexity (Hughes et al., 2019; Solhjoo et al., 2019);
- Galvanic Skin Response (GSR) – reflects arousal and mental workload under stress or attentional demand (Widyanti et al., 2017; Zihisire Muke et al., 2022);
- Electroencephalography (EEG) – used for direct brain-based estimation of cognitive states across different frequency bands(Kyriaki et al., 2024) ;
- Oculometry parameters – metrics such as pupil dilation, fixation duration, and blink rate have shown sensitivity to mental load in multiple domains (Gambiraza et al., 2021).

In this thesis, instrument selection prioritises feasibility in clinical and training contexts. Wrist-worn photoplethysmography and electrodermal sensors are used to derive HR, HRV, and skin conductance, since they are non-intrusive and require minimal setup time for seated participants.

A distinction must also be drawn between explicit measures, which require conscious reflection (e.g., rating perceived effort), and implicit measures, which infer cognitive states from involuntary

physiological or behavioural responses. This is especially important in populations with neurodevelopmental conditions, stroke, or cognitive impairment, where verbal reports may be unreliable or fatiguing.

In summary, these methods support the use of multimodal indicators in this thesis. Clarifying their theoretical assumptions is essential for selecting appropriate tools, interpreting user responses, and maintaining ecological validity in simulator-based evaluations. The following section outlines the key hypotheses associated with commonly used cognitive load metrics, including subjective self-reports, behavioural outcomes, and physiological indicators.

2.3.2.2 Theoretical Assumptions and Hypotheses Behind Cognitive Load Metrics in Wheelchair Simulator Studies

Cognitive load assessment relies on the assumption that internal mental effort manifests through observable external indicators, whether self-reported, behavioural, or physiological (Ouwehand et al., 2021). These methods each involve trade-offs between validity, sensitivity, and feasibility, particularly in simulation contexts (N. Hollender, C. Hofmann, M. Deneke, 2010; Naismith et al., 2015; Ouwehand et al., 2021; F. Paas et al., 2003; Sweller, 2018). While general, the point is especially pertinent in powered wheelchair simulators given user heterogeneity, communication constraints, and safety considerations; hence the need to state measurement assumptions and triangulate measures.

Subjective self-report tools (Hart, 2006; Hart & Staveland, 1988) assume users can accurately reflect on and rate their cognitive effort, often across multiple dimensions such as mental demand, effort, or frustration. These tools are widely used in simulator studies. For example, (Kamaraj, Dicianno, et al., 2016b) reported higher NASA-TLX mental demand in VR compared with real-world driving, suggesting adaptation challenges; (L. Devigne et al., 2017) found moderate workload (mean = 27.2), indicating manageable demands; (Vailland et al., 2020) compared workload with and without vestibular feedback (3-DoF motion platform), using NASA-TLX alongside presence measures (IPQ), and found that a multisensory motion platform improved perceived VR quality and reduced cognitive demand relative to a visual-only setup. (Fraudet et al., 2024) observed higher scores in VR across circuit difficulties, indicating persistent cognitive burden. While versatile, NASA-TLX depends on user introspection and may be less reliable in populations with communication or cognitive limitations.

Physiological metrics are grounded in the assumption that cognitive load modulates autonomic and neural activity (Hebbar et al., 2021; Hughes et al., 2019). EEG studies link increased frontal theta and decrease parietal alpha to higher effort. HR and HRV are commonly used indicators of sympathetic arousal. GSR reflects arousal, stress, and attentional demand. These signals have been validated in

driving and teleoperation simulators: e.g., (Moya et al., 2017) found higher theta activity and performance degradation under network delay, while (Li et al., 2014) reported greater HR and NASA-TLX scores under dual-task driving. Oculometry (blink rate, pupil dilation) provides additional indicators of visual-cognitive engagement (S. Chen & Epps, 2014; R. Martins & Carvalho, 2015).

Table 2.10 compiles related work that illustrates how different approaches have been applied to capture workload variations under varying task demands and interface conditions. In contrast to earlier studies, which relied primarily on NASA-TLX, the present thesis broadens assessment by integrating session-level (NASA-TLX), task-level (PAAS), physiological measures (EDA, HR, HRV), and performance metrics.

In summary, these findings justify the adoption of a multimodal cognitive load assessment strategy in wheelchair simulator research. Cognitive load manifests not only through task performance and subjective perception but also through physiological and attentional behaviours, which are particularly valuable when working with users who have cognitive or expressive limitations. For users with cognitive or expressive limitations, attentional behaviour is operationalised using passive, non-verbal indicators that impose minimal burden, including: (i) responsiveness to salient events (e.g., time-to-first-move after a visual cue; stopping at command or hazards), (ii) control-stream intermittency (e.g., joystick command rate, micro-pause frequency, command entropy), (iii) spatial path variability (e.g., deviation from an optimal trajectory; unnecessary heading reversals), and (iv) viewpoint dynamics (e.g., head-orientation variability in viewport motion on desktop) as a coarse proxy of visual scanning when eye-tracking is unavailable. These logs are derived directly from the simulator and wearables and do not require verbal reporting. Nevertheless, understanding cognitive effort alone is not sufficient to fully evaluate user interaction. Equally important are the perceptual, emotional, and usability-related factors that shape how individuals experience and respond to simulator tasks. The next section therefore introduces the Quality of Experience (QoE) framework, which complements cognitive load assessment by addressing affective, perceptual, and experiential responses that influence user engagement and acceptance of simulator technologies.

Table 2.10: Cognitive Load Studies Including Wheelchair and Non-Wheelchair Simulators

Study	Simulator	Cognitive Load Metric	Interpretation
(Kamaraj, Dicianno, et al., 2016b)	VRSIM-2	NASA-TLX (Raw subscales)	Workload varied across HMI conditions; higher mental demand and frustration in VR vs real-world driving.
(L. Devigne et al., 2017))	PWS	NASA-TLX	Reported low-to-moderate workload (mean TLX: 27.2 ± 18.2); most participants rated workload below mid-scale.
(Vailland et al., 2020)	PWS	NASA-TLX	NASA-TLX score showed significantly higher workload in no vestibular scenario comparing with vestibular feedback scenario.
(Fraudet et al., 2024)	SIMADAPT	NASA-TLX	NASA-TLX used to compare VR and real driving; VR condition showed higher scores on average for mental workload.
[1](Moya et al., 2017)	Desktop Simulator	NASA-TLX, EEG and performance	Higher cognitive load linked to increased EEG theta activity and reduced performance under network delay.
[2](Li et al., 2014)	Driving simulator	NASA-TLX and HR	Greater NASA-TLX scores and increased HR observed during dual-task driving; younger drivers reported higher load.
This thesis	E-WATS	NASA-TLX (session-level),PAAS (task-level),physiological measures (EDA, HR, HRV) and performance	Reported workload variations across immersive conditions; correlations between NASA-TLX and cybersickness; demonstrated feasibility of multimodal assessment in clinical populations.

[1] Mobile robot teleoperation. [2] Car driving simulator for curve negotiation.

2.3.3 Quality of Experience (QoE): Definition and Features

While cognitive load offers insight into users' mental effort, it does not fully capture how users perceive or interact with simulation systems. To address this limitation, the Quality of Experience (QoE) framework is incorporated. QoE is a user-centred model that evaluates how well a system meets expectations under real-world conditions (C. W. Chen et al., 2015; Vlahovic et al., 2022). Unlike Quality of Service (QoS), which focuses on technical performance (e.g., latency, resolution), QoE encompasses both subjective and objective dimensions of user experience, including usability, emotional engagement, satisfaction, and contextual factors (Callet et al., 2013; Möller & Raake, 2013). The ITU-T defines QoE as:

> *"The degree of delight or annoyance of the user of an application or service, resulting from the fulfilment of his or her expectations with respect to the utility and/or enjoyment of the application or service in the light of the user's personality and current state" (ITU-T P.10/G.100(ITU, 2017).*

This definition underscores the subjective, dynamic, and context-sensitive nature of QoE. In simulator-based assistive technologies, such as wheelchair simulators, QoE is essential for understanding how factors such as motivation, immersion, and fatigue influence not only usability but also long-term engagement.

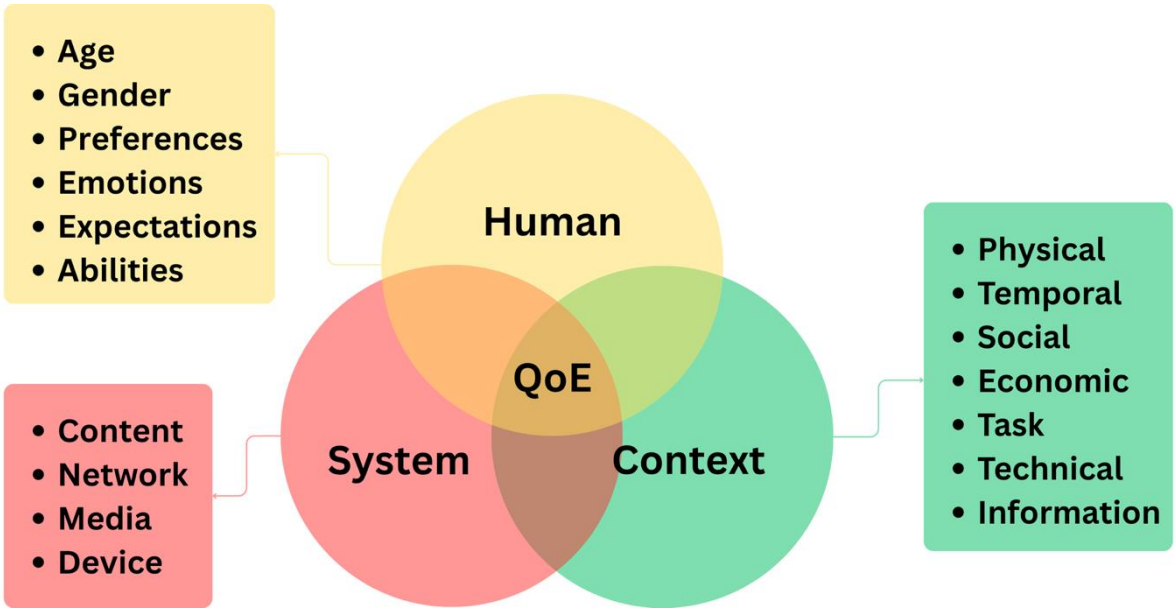


Figure 2.4: QoE influencing factors. Image adapted from (Callet et al., 2013; Möller & Raake, 2013).

To structure QoE, this study adopts the Influencing Factor (IF) model (Reiter et al., 2014), shown in Figure 2.4,. This framework acknowledges that early QoE research, rooted in QoS and network

engineering, focused mainly on system factors in controlled laboratory conditions (Callet et al., 2013; Möller & Raake, 2013). However, user-related and contextual factors also play a decisive role in shaping overall satisfaction. By accounting for the interplay between these factors, researchers and designers can develop systems that are more accessible, efficient, and engaging.

2.3.3.1 QoE Features in Interactive VR and Simulator System Context

In interactive and VR systems like powered wheelchair simulators, QoE is shaped by a complex interplay of perceptual, interaction and contextual characteristics. According to ITU definitions and Möller et al. (Ute Jekosch, 2005; Vlahovic et al., 2022), a QoE feature is:

> *"a perceivable, recognized and nameable characteristic of the individual's experience of a service which contributes to its quality" (Ute Jekosch, 2005)*

QoE features can be categorized in hierarchical levels based on how users experience and engage with the system (Callet et al., 2013; Möller et al., 2014). Perceptual level includes sensory qualities such as brightness, contrast, flicker, colour fidelity, loudness, and sound spatialization. Action level is related to immersion, spatial perception, and self-motion awareness within the virtual environment. Interaction level captures how responsive, intuitive, and natural the interaction mechanisms feel to the user. Usage instance level focusing on aspects such as ease of use, learnability, aesthetics, and task clarity. Service level encompasses overarching attributes like appeal, utility, acceptability, and perceived usefulness across repeated or long-term use.

These features provide a multidimensional lens through which to evaluate and design VR and multimedia simulator systems. In the domain of powered wheelchair simulators, many of these QoE features are directly applicable but often underexplored. While some taxonomies, such as those proposed in ITU-T P.809 (*ITU-T P. 809: Standardization Activities Targeting Gaming Quality of Experience – ACM SIGMM Records*, 2018) and adapted from gaming contexts, offer generalizable these dimensions, features like challenge, tension, or narrative engagement may be less relevant in clinical scenarios. Instead, greater emphasis must be placed on features more applicable to VR dynamic-based systems, as described at (Vlahovic et al., 2022) and analysed in the scoping review by (Arlati et al., 2020). These include:

- **Immersion, Sene of Presence (SoP)**: defined as the psychological state in which users perceive themselves as being inside a virtual environment. This feature significantly influences how realistically users engage with the simulation and perceive the interaction.

- **Discomfort, fatigue and cybersickness**: physical side effects that result from the immersive and often intrusive nature of VR platforms. These features are particularly relevant in clinical populations and can significantly degrade overall QoE.
- **System Realism (responsiveness and input fidelity)**: although the term "realism" may not be explicitly defined in some taxonomies, (Arlati et al., 2020) emphasize that without adequate system responsiveness and input fidelity, simulators lack the capability to support skill transfer and training reliability. These aspects directly impact both perceived and functional QoE in simulator-based applications.

Overall, QoE features serve not only as evaluation criteria but also as design goals to enhance user-centred performance and reduce barriers to adoption. As summarized in Table 2.11, key features such as immersion, discomfort, and system responsiveness are influenced by both system-level and human factors.

Table 2.11: QoE Features and Influencing Factors in VR-based Simulator System

QoE Feature	Main Influencing Factor(s)	Explanation
Sense of Presence (SoP) & Immersion	System, Human	Affected by system parameters like field of view and resolution; also depends on user's psychological state and attention.
Discomfort, Fatigue and Cybersickness	Human, System	Influenced by user physiology (e.g., vestibular sensitivity) and system issues like latency or optical misalignment.
Realism via System Responsiveness and Input Fidelity	System	Depends on technical system performance such as low-latency response, accurate control mapping, and feedback consistency.

However, the lack of a unified taxonomy of QoE features specifically tailored to VR-based clinical simulators remains a significant gap in the literature. A domain-specific framework that incorporates both immersive interaction factors and physical side effects is important for enabling evaluations in real-world clinical settings. To capture how users engage with and react to simulator-based systems, these features must be paired with appropriate assessment tools. The next section introduces the methods used to evaluate QoE, including both subjective and implicit measures, as applied in immersive and assistive simulation contexts.

2.3.3.2 QoE Assessment Methods

In QoE research, assessment and evaluation serve distinct yet complementary purposes. Assessment aims to inform improvement by providing feedback on user experience, while evaluation judges the

current quality of the system, often using scores or qualitative comments without prescribing changes (da Silveira et al., 2024; Perkis et al., 2020; Vlahovic et al., 2022). Both processes involve observation and data collection.

In the context of Assistive Technology, achieving high QoE is especially important. User satisfaction is shaped not only by system performance, but also by psychosocial and contextual factors such as motivation, fatigue, and personal goals (da Silveira et al., 2024; Domingues et al., 2019; Vlahovic et al., 2022). Evaluating QoE in immersive virtual reality (VR) systems, such as the wheelchair simulator used in this study, requires a mixed-methods approach. This involves integrating:

- Explicit subjective perceptions (e.g., sense of presence/immersion, usability, emotional response, cognitive workload)
- Implicit objective indicators (e.g., physiological or behavioural markers).

As emphasized in recent survey on QoE assessment in interactive VR environments (Vlahovic et al., 2022), meaningful insights depend on the ability to capture both what users report and how they respond implicitly during interaction. However, VR-based QoE evaluation presents specific challenges:

- High variability across individuals
- Complex interactions between sensory and cognitive load.
- Physical limitations or fatigue in users.

These challenges are further amplified in assistive technology contexts, where participants may present diverse abilities or limitations. To address this, the present study adopts dual-modality QoE assessment strategy, combining established subjective instruments, and implicit (physiological and simulator-based performance report) metrics.

2.3.3.2.1 Subjective QoE assessments

Subjective methods are the most used approach for assessing QoE in immersive systems (Vlahovic et al., 2022). These methods primarily rely on self-reported feedback, often collected via standardized questionnaires, rating scales, interviews, or diaries, and are typically administered either during or immediately after the VR experience. The prevalence of subjective methods stems from their ability to directly capture the user's perceived quality and affective response to a given system.

Despite their dominance, there is no standardized methodology for assessing QoE in VR applications, although ongoing work, such as that by ITU-T Study Group 12, aims to address this(ITU-T, 2025; Kojić T. et al., 2021) .In practice, researchers draw on a variety of multi-item instruments, each targeting specific aspects of the experience. For example, the System Usability Scale (SUS) (Brooke, 1996)

measures usability and satisfaction, while the Self-Assessment Manikin (SAM) (Bradley & Lang, 1994) evaluates affective states along valence, arousal, and dominance dimensions.

However, several instruments commonly used in multimedia or gaming contexts, such as the Game Experience Questionnaire (GEQ) (IJsselsteijn et al., 2013; Poels et al., 2007) or the Player Experience Inventory (PXI) (Abeele et al., 2016, 2020), do not fully capture VR-specific aspects such as discomfort, fatigue, or cybersickness. This gap has led to the creation of VR-specific tools like the Virtual Reality Neuroscience Questionnaire (VRNQ) (Kourtesis et al., 2019), which includes items on user experience, game mechanics, and VR-induced symptoms, and the more general VR UX Questionnaire developed by (Tcha-Tokey et al., 2016), which includes subscales for presence, flow, engagement, fatigue, and technology adoption. Simulator Sickness Questionnaire (SSQ) (Balk et al., 2013; Kennedy et al., 1993) is also widely used to assess discomfort related to VR-induced motion symptoms.

In addition to these multi-item instruments, some studies use Absolute Category Ratings (ACR) (*ITU-T P. 809: Standardization Activities Targeting Gaming Quality of Experience – ACM SIGMM Records*, 2018) to evaluate specific aspects such as visual clarity, control responsiveness, and comfort. These scalar judgments are typically aggregated into Mean Opinion Scores (MOS) (ITU, 2017; Streijl et al., 2016), a method commonly applied in multimedia and VR research, to provide a concise and interpretable summary of perceived quality on a standardized five-point scale.

Nevertheless, subjective self-reports are cognitively mediated, and thus vulnerable to several forms of response bias. These include central tendency bias, acquiescence bias, and social desirability effects, especially when responses are recorded by an administrator during HMD use (Bowman et al., 2002; Vlahovic et al., 2022). Furthermore, question interpretation, questionnaire fatigue, and recall limitations may distort the accuracy of user feedback, particularly in long or cognitively demanding experiences.

As a result, researchers increasingly recommend integrating questionnaires directly within the virtual environment (in-VR assessments) (Regal et al., 2018). This method reduces interruptions, avoids experimenter bias, and captures immediate reactions without compromising immersion. However, implementing in-VR questionnaires adds technical and design complexity and was beyond the scope of the present study.

In summary, subjective QoE assessment/evaluation methods are indispensable for capturing user perception, but they must be carefully selected and interpreted. The development of VR-specific instruments is a promising trend that addresses platform-specific challenges. For comprehensive

evaluation, subjective methods should be integrated with complementary implicit and behavioural metrics, as discussed in the following section.

2.3.3.2.2 Implicit QoE assessments methods and challenges

To complement subjective measures, QoE studies also can incorporate implicit assessment techniques that capture involuntary or non-verbal indicators of user state. These approaches are particularly relevant in assistive and clinical settings, where verbal self-reporting may be constrained by fatigue, cognitive limitations, or communicative impairments.

Physiological measures and QoE considerations

As discussed earlier, physiological signals offer real-time, involuntary markers of user state, supporting inferences about arousal, fatigue, engagement, and mental effort (da Silveira et al., 2024; Vlahovic et al., 2022). Commonly used signals include heart activity, galvanic skin response (GSR), and brain activity (Timmerer et al., 2015; Vlahovic et al., 2022). These measures might provide continuous, less biased insights into internal states, enhancing QoE models with user-related factors.

However, their integration into QoE research present multiple challenges. Firstly, medical-grade devices can be intrusive, potentially degrading user experience and influencing the QoE they aim to measure (Timmerer et al., 2015; Vlahovic et al., 2022). As a result, researchers increasingly use less invasive tools, such as wearables like smartwatches and fitness bands.

Second, another challenge concerns the context-dependent nature and inter-subjective variability of physiological responses in emotion recognition, as these may reflect a range of overlapping psychological states. For instance, elevated heart rate or increased skin conductance can occur across a range of affective conditions, from surprise, excitement or anxiety, since these markers primarily reflect arousal intensity and is non-specific to the emotion valence (Ahmad & Khan, 2022; Giannakakis et al., 2022; Kleiman et al., 2021; Wilhelm et al., 2006). This multimodal characteristic highlights the importance of contextualising physiological signals, rather than interpreting them in isolation.

To support this perspective, the 2D (valence-arousal) and 3D (valence-arousal-dominance) emotion models are illustrated in Figure 2.5, offering a visual representation of emotion dimensions from (Ahmad & Khan, 2022; Russell, 1980).

Furthermore, some physiological methods are susceptible to motion artifacts, particularly in VR where head and hands movement is frequent. EEG is especially vulnerable to such artifacts (Murphy & Higgins, 2019), while techniques like fMRI requires stillness, making it incompatible with natural VR use.

Given these challenges, the interpretation of physiological signals requires careful contextualisation. Instead of relying on these measures in isolation, they should be integrated into a broader multimodal QoE assessment framework that combines behavioural observations, performance metrics, and subjective reports.

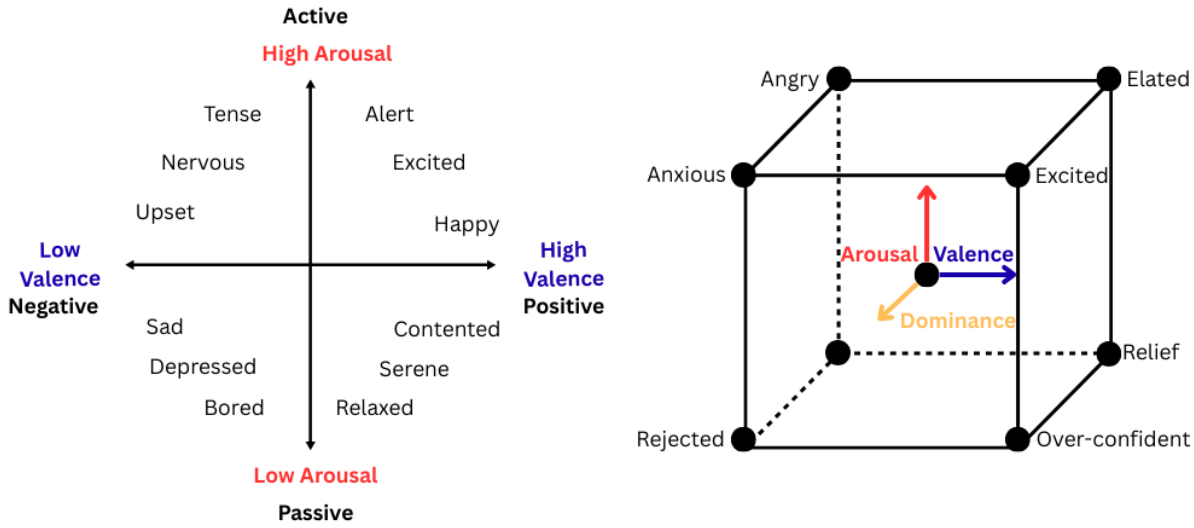


Figure 2.5: 2D (valence-arousal) and 3D (valence-arousal-dominance) emotion models adapted from (Ahmad & Khan, 2022; Blanco-Ríos et al., 2024; Russell, 1980).

Behavioural indicators

Behavioural responses offer indirect but valuable insights into user presence, engagement, and comfort in VR. These indicators include both reflexive reactions to virtual stimuli and adaptive behaviours in response to environmental cues.

Notable measures include user responses to conflicts between real and virtual cues, such as flinching or stepping back from perceived threats, as well as reflexive gestures to unexpected events (e.g., ducking or sidestepping) (Murphy & Higgins, 2019; Sheridan, 1996; Vlahovic et al., 2022). Such reactions are often interpreted as evidence of sensorimotor immersion and realism in the virtual environment.

As with physiological data, behavioural cues are most meaningful when interpreted within context and in combination with other data streams. Integrated into a broader QoE framework, they help capture non-verbal feedback that may signal discomfort, disorientation, or spatial awareness, especially in users who may not be able to verbalize their experiences.

Performance Metrics

Task-based measures such as completion time, control smoothness, spatial accuracy, and response latency objectively quantify user performance. These metrics capture ease of use, motor efficiency, and interaction quality, and are especially relevant in evaluating learning curves and system usability in VR applications (A. R. de Sá et al., 2022).

Cybersickness with Physiological and Behavioural Metrics

Combining physiological and behavioural data can also help detect cybersickness, a common concern in immersive simulation. For instance, psychophysiological responses have been associated with simulator sickness through SSQ (Dennison et al., 2016; Iskander et al., 2018), and eye-gaze patterns (behaviour) has been used to infer visual fatigue, attentional allocation, and reaction to cybersickness triggers (Iskander et al., 2018; Vlahovic et al., 2022).

Therefore, study in the literature exploring deep neural networks, such as CNN-LSTM models, applied to users' physiological signals (heart rate and galvanic skin response), have demonstrated the ability to detect and predict cybersickness using few minutes of data, achieving accuracies of 97.44% and 87.38%, respectively (Islam et al., 2020).

Summary and Implications

A key limitation of implicit QoE assessment lies in the non-specificity of physiological markers. As emphasized in literature, no single physiological signal maps unambiguously to a psychological state, given that multiple mental or emotional conditions may trigger similar autonomic responses. This reinforces the need for multimodal fusion, combining physiological, behavioural, and subjective inputs to achieve a more accurate and ecologically valid understanding of user experience.

Ultimately, integrating implicit metrics into QoE evaluation frameworks is particularly beneficial in assistive technology applications. These measures not only provide continuous, non-intrusive feedback on user state, but also can support the development of adaptive systems that dynamically adjust to user needs, enhancing comfort, engagement, and accessibility. The following section connects these QoE principles to core healthcare quality domains, patient experience, clinical effectiveness, and safety, highlighting their relevance in digital health applications.

2.3.4 Summary

This chapter established the theoretical and methodological foundation for assessing user responses in immersive powered wheelchair simulators. By integrating Cognitive Load Theory (CLT) and Quality of Experience (QoE), it proposed a multidimensional evaluation strategy that extends beyond traditional performance metrics to include both subjective and objective measures.

Section 2.3.2 introduced CLT, outlining the types of cognitive load (intrinsic, germane, extraneous) and justifying a multimodal assessment approach that combines self-reports, behavioural observations, and physiological signals to evaluate mental workload.

Section 2.3.3 explored the QoE framework, focusing on dimensions such as usability, emotional response, immersion, and system tolerance. The integration of subjective instruments (e.g., SUS, IPQ, SAM, SSQ) with implicit signals (e.g., EDA, HRV, eye tracking, and task performance) was emphasized as essential for capturing user experience, particularly in assistive settings where verbal feedback may be limited.

Overall, the chapter supports a user-centred, inclusive, and clinically grounded evaluation model. This foundation informs the next chapter, which details the experimental design, participant recruitment, data collection procedures, and analysis plan for applying the proposed multidimensional framework.

Part III METHODOLOGY

Chapter 3 Research Methodology

3.1 Introduction

This chapter presents the methodology adopted across three experimental studies conducted during this PhD research. The overall approach integrates explicit and implicit data to assess users' Quality of Experience in a powered wheelchair simulator, using a mixed-methods framework to both lab and field settings. Iterative refinements were made across studies to enhance ecological validity and inclusivity, ensuring relevance to diverse user profiles.

3.2 Studies Overview

This research involved three iterative studies designed to develop, test, enhanced and validate E-WATS system, a power wheelchair simulator. Each study built upon the previous one, contributing to a structured assessment and training framework grounded in user-centred and human-centred design principles.

- **Study 1** assessed initial simulator use in controlled settings, focusing on usability, cognitive workload, and user satisfaction.

  Contribution 1: Established a QoE assessment approach that combines subjective instruments with wearable-derived physiology (e.g., EDA, HR/HRV) alongside performance metrics. To the best of the author's knowledge, this represents an early application of such multimodal QoE methods in wheelchair simulation and has been cited in subsequent work as a rationale for using wearables in this domain.
- **Study 2** examined the impact of motion profiles (e.g., low vs. high jerk) on cybersickness and task performance.

  Contribution 2: Demonstrated that adjustable motion settings (rather than a single "most realistic" profile) can be used to tailor immersion to user tolerance and to evaluate whether software-level adjustments mitigate cybersickness while preserving performance.
- **Study 3** evaluated the simulator in a real-world clinical setting with powered wheelchair users, focusing on feasibility, adaptability, and acceptability.

  Contribution 3: Developed and piloted a protocol that maps simulator tasks to recognised clinical frameworks (e.g., WST/PMRT alignment) and integrates subjective QoE with wearable physiology

in collaboration with clinicians and end-users. This extends prior interventions summarised in Chapter 2 by embedding multidimensional QoE into a clinically interpretable workflow.

- **Synthesis across studies:**

  Contribution 4: Formulated EMPOWER-SIM, as a set of practical guidelines and methodological recommendations, informed by findings across the studies. It represents an initial step toward a validated framework for clinical integration

These contributions collectively support the integration of the simulator into clinical and training contexts.

3.3 Wheelchair Simulator System Overview

Designing a wheelchair simulator that accommodates a range of user needs, varying in motor ability, cognitive function, and experience with power wheelchair controls, requires a flexible and adaptable approach. To address this, a user-centred design (UCD) methodology was adopted (Figure 3.1), allowing iterative refinement based on user feedback.

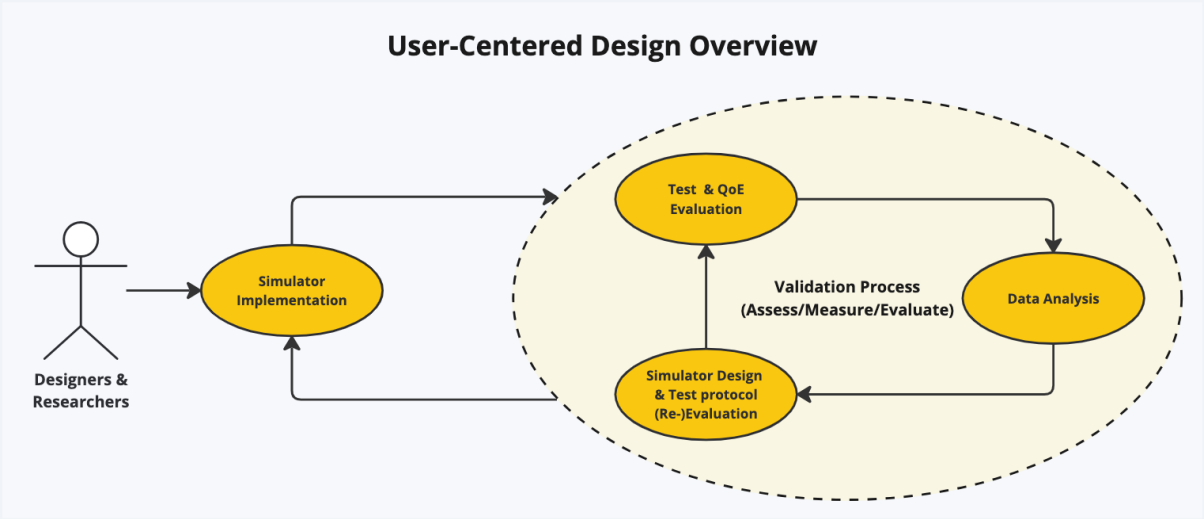


Figure 3.1: User-centred design overview.

Figure 3.1 illustrates the iterative UCD cycle used to guide simulator development and evaluation. The process begins with initial system implementation, followed by a validation phase focused on usability testing and QoE evaluation. Data collected during this phase are analysed to identify strengths and limitations, which then inform revisions to the simulator design and test protocol. This feedback loop ensures continuous improvement and alignment with evolving user requirements.

A modular design approach supports flexibility while maintaining a consistent assessment structure. This modularity allows the simulator to be tailored to various user profiles through configurable

software settings and hardware components. In practice, this means that individual subsystems can be modified or replaced (e.g., swapping input devices, updating rendering engines, or adding physiological sensors) without compromising the integrity of the overall framework. Such adaptability enhances accessibility, ensures sustainability, and supports ongoing evolution in response to user feedback.

The QoE-based evaluation framework was designed in parallel with the simulator and combines implicit and explicit measurements. The modular components are described in the following section.

On the hardware side, the simulator can integrate different input and output devices, including joysticks, eye trackers, head-mounted displays, haptic feedback coupled with the joystick, and wearable sensors for physiological monitoring. On the software side, a layered architecture was implemented (Unity for the simulation environment, Lab Streaming Layer for synchronisation (Kothe et al., 2025), OpenFace (Baltrusaitis et al., 2016) for video capture, and analysis in Python/MATLAB), allowing individual components to be replaced or extended. This approach ensures that the system is not a fixed prototype but rather a platform that can evolve in line with standards for hardware, software, and medical device interoperability.

In terms of standards, the simulator was formally registered as software with the Instituto Nacional da Propriedade Industrial (INPI) in Brazil in 2019, securing intellectual property protection. By its intended use, supporting wheelchair assessment and training, it falls within the definition of software as a medical device (SaMD). At the time of registration, regulatory frameworks for SaMD were still emerging; the FDA had issued its first SaMD guidance in 2017 and was in a transition phase, the European MDR was adopted in 2017 but only came into force in 2021, and ANVISA introduced RDC 657/2022. While the system has not been submitted for regulatory approval, its modular design allows future alignment with these standards, particularly in relation to interoperability protocols such as HL7/FHIR.

3.3.1 Core Modular Components

The simulator is structured using a System-of-Systems (SoS) architecture, with three key subsystems working independently yet cohesively (Figure 3.2):

- **Engine System**: Generates the virtual environment, supports physics-based interaction, and allows customization of training scenarios and task rules.
- **User Interface System**: Manages user interaction through visual, auditory, and haptic feedback. It includes graphical rendering and input/output devices like joysticks and screens.

- **Data Collection System**: Logs performance metrics, task outcomes, and environmental variables in real-time, with synchronized data streams to support multimodal analysis.

The virtual environment was developed using Unity 3D (versions 2014.9.18f1, 2017.2.0f3, and 2021.3.24f1) and ran on a Windows 10 Enterprise PC equipped with an Intel Core™ i7-8700 CPU, 16GB RAM, and an NVIDIA GeForce GTX 1080 GPU. The simulator was designed as a safe training and assessment tool, offering novice power wheelchair users a way to develop skills without real-world risk.

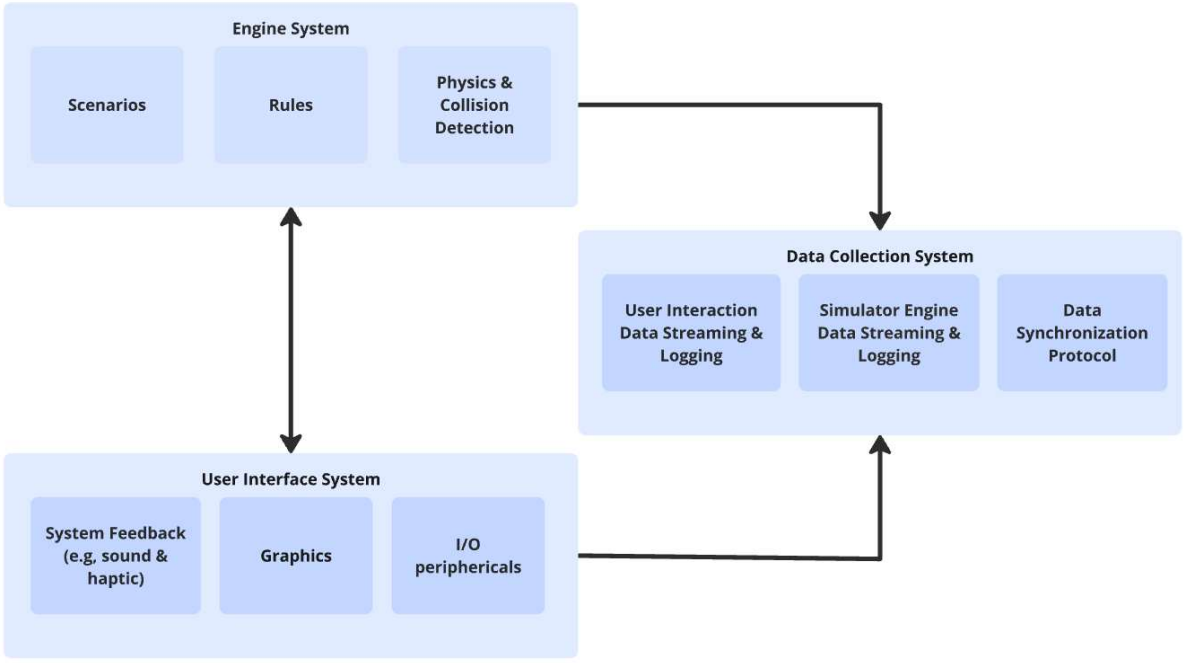


Figure 3.2: Wheelchair Simulator components (subsystems) that define the architecture of the system proposed.

3.3.2 User Feedback Integration: Sense module

To capture user responses, the simulator includes a Sense Module (see Figure 3.3), designed to collect both implicit and explicit feedback. This includes physiological signals from wearable devices, eye and head tracking data (camera based), as well as a single-channel EEG signal (frontal electrodes location) for detecting eye movement, specific blinks. Explicit responses are gathered through questionnaires and audio-recorded interviews. Also, it is important to note that not all data streams were equally reliable or continuously available in every session, and integration across devices required manual calibration and post-processing.

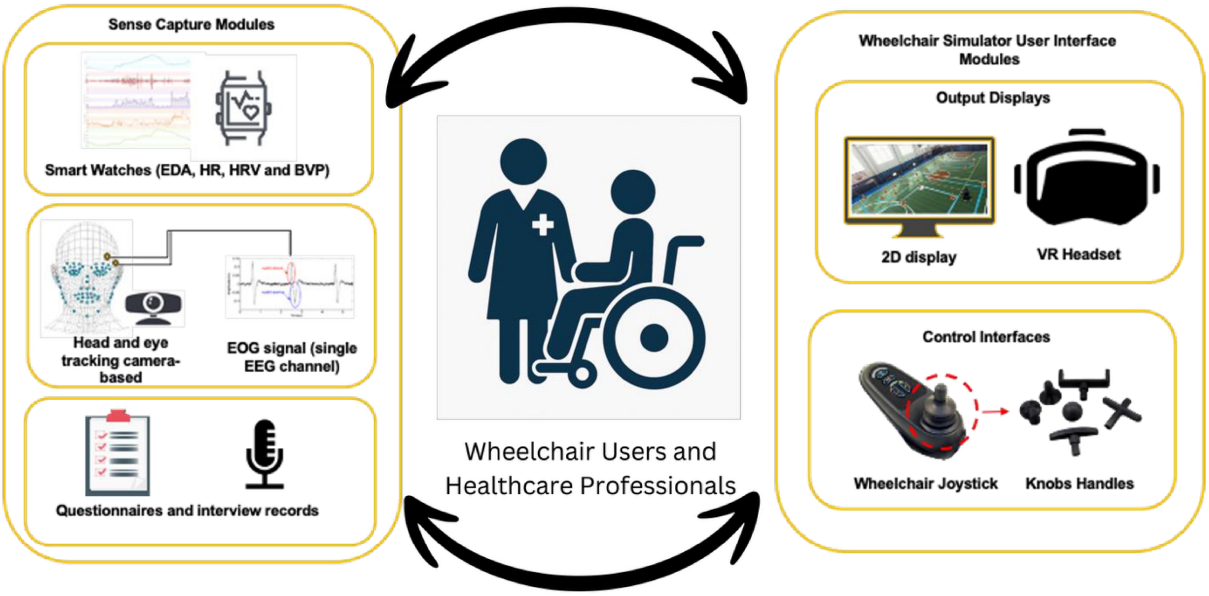


Figure 3.3: User interface and sense capture modules overview.

The integration of wearable technologies with synchronized data collection enables monitoring of users' cognitive and emotional states. This multimodal feedback provides valuable insights into user behaviour and system interaction, supporting evidence-based design adjustments. By incorporating these inputs, the simulator becomes progressively more adaptive and user-centred, improving its effectiveness across diverse user profiles.

3.4 Data Synchronisation Framework

The data synchronisation framework consists of the Local Machine, which runs the Unity-based simulator and records data streams, and the external devices, which provide physiological and behavioural signals. This framework ensured temporal alignment across multiple input streams, including EEG, joystick events, Empatica wristband data, OpenFace head-pose tracking, and simulator performance metrics.

Two complementary methods were implemented to accommodate this diverse range of devices and tools. The first method employed the Lab Streaming Layer (LSL)(Kothe et al., 2025), an open-source framework designed to provide sub-millisecond accuracy for time-synchronised data acquisition. LSL operates as a middleware layer that corrects for network jitter and latency, aligns device clocks, and supports high-throughput multi-stream recording using LSL dynamic library (liblsl). The second method combined TCP/IP event tagging with the OpenViBE acquisition server(Renard et al., 2010). OpenViBE is an open-source platform that can receive data through TCP/IP sockets and republish them into LSL. In this way, OpenViBE acted as middleware whenever devices, such as the EEG headset,

lacked native LSL support, thereby ensuring that all streams were harmonised within a single temporal framework.

Figure 3.4 provides an overview of this architecture, illustrating how LSL establishes a network connecting acquisition devices, storage components, and processing tools. Within this network, LSL outlets publish data streams that LSL inlets can subscribe to, while LSL Data Recorder supervises the recording of multiple outlets with minimal overhead. Clients may include device integrations, visualisation modules, real-time analysis tools, or stimulus–response mechanisms. This structure provided the foundation for the data synchronisation approach used throughout the thesis.

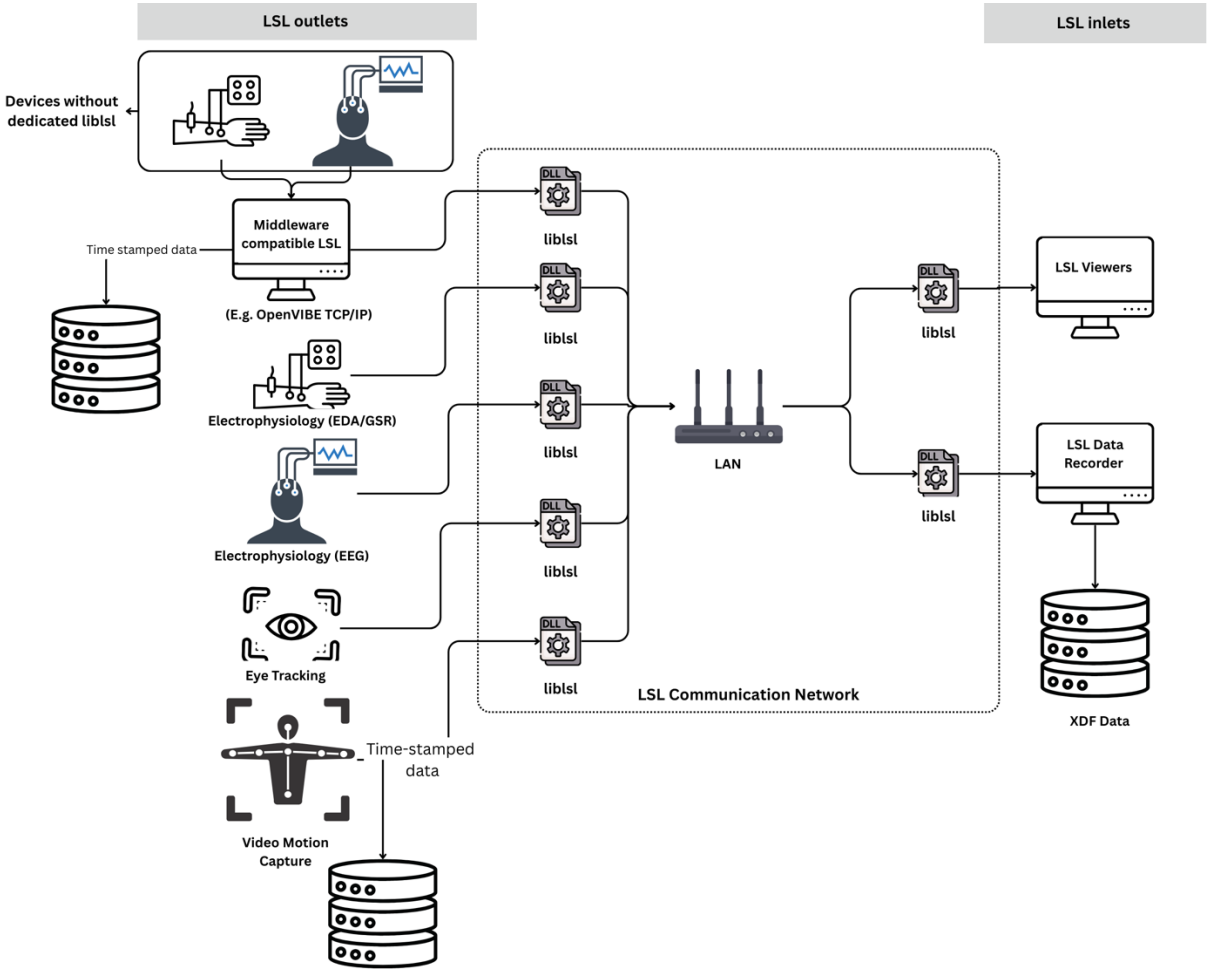


Figure 3.4: Data Synchronisation system overview. Imaged adapted from (Kothe et al., 2025).

3.4.1 Data Synchronisation Pipeline

The pipeline integrated core components implemented in different programming languages but connected through common bindings. Participants operated the Unity-based simulator, implemented in C#, using a joystick. Simulator events, including joystick actions and collisions, were transmitted

directly into LSL via the LSL4Unity-liblsl interface. On the local machine, this ensured sub-millisecond accuracy, with reported latencies of less than 0.1 milliseconds. Within the pipeline, Unity also published structured markers for the main experimental events: (a) Experiment Start; (b) Experiment Stop; (c) Trial Start; (d) Trial Stop; (e) Baseline Start; (f) Baseline Stop; (g) Button pressed; (h) Button released; and (i) Collision (user mistake).

For devices unable to connect natively to LSL, the pipeline incorporated TCP/IP event tagging through OpenViBE. In this configuration, Unity continued to publish simulator events to LSL, while devices such as the EEG headset streamed data to OpenViBE via TCP/IP. OpenViBE subsequently republished these signals into LSL, thereby serving as a bridge between device outputs and the synchronised environment. Figure 3.5 illustrates this integration, showing how simulator events (joystick inputs, collisions, and task markers) and external device data were unified through the LSL framework.

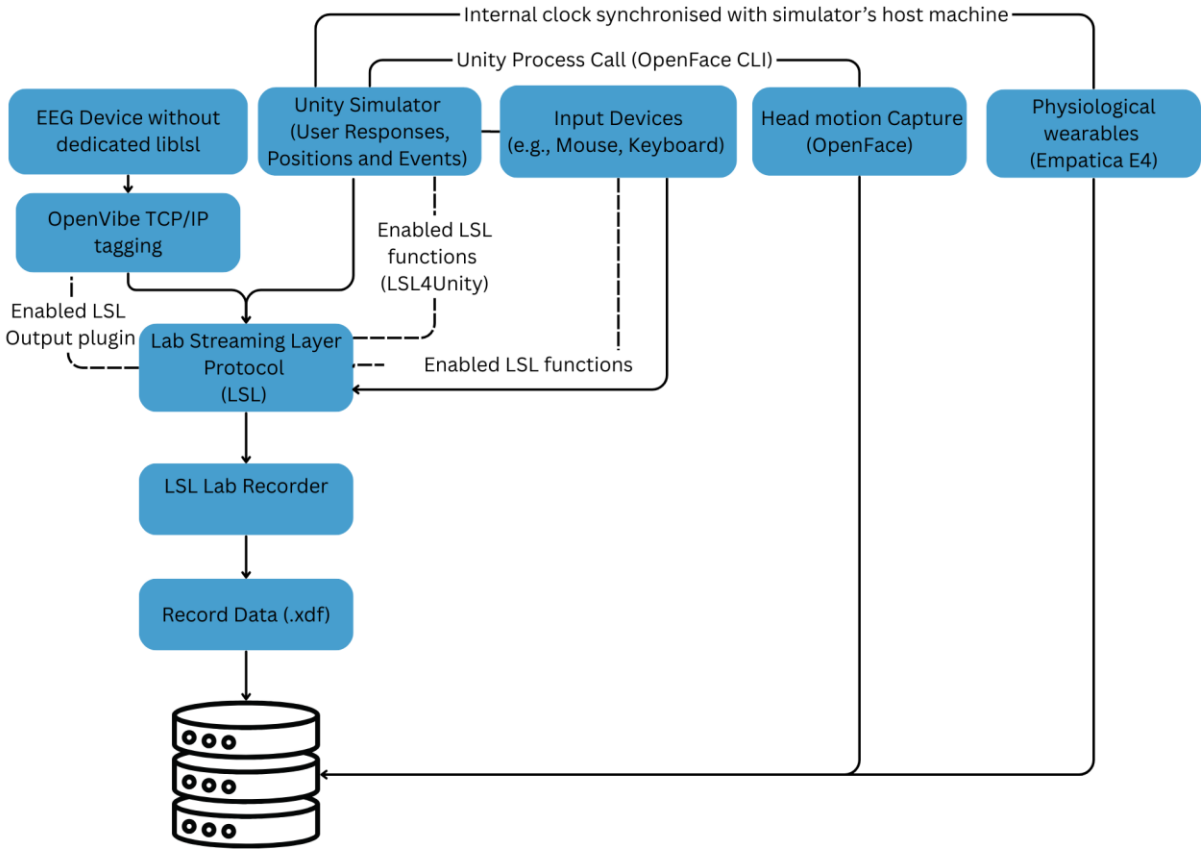


Figure 3.5: Data synchronisation pipeline using Lab Streaming Layer and OpenVIBE TCP/IP.

The Empatica wristbands were synchronised using two mechanisms. First, the wristband itself supports button markers that are stored in the raw data whenever the device button is pressed. These markers were used during the experiments at key task boundaries (e.g., start or end of a trial) to

provide reference points for alignment with simulator events. Second, the wristband's internal clock was synchronised with the host computer's system time at the beginning of each session.

Physiological devices required additional handling. The Empatica wristband supported two synchronisation mechanisms. First, button markers embedded within the raw data at key task boundaries provided reference points for alignment with simulator events. Second, the wristband's internal clock was synchronised with the host computer at the beginning of each session.

OpenFace, an open-source toolkit for head-pose and facial landmark tracking, was launched from Unity by invoking a terminal command through the C# process interface. Unity logged a marker at the onset of recording, while OpenFace generated frames with timestamps initialised at zero. To correct for this, a post-processing step compared the file creation time and frame rate against Unity's timeline, enabling accurate offset correction. Although this additional alignment procedure was sufficient for the present studies, recent extensions of cameras captures (e.g., TimeShot application) and python LSL library (pylsl) would allow OpenFace outputs to be streamed directly into LSL in future iterations, eliminating the need for post-hoc correction

3.4.2 Software Packages and Data Analysis Stack

All synchronised data streams were stored on the Local Machine in extensible data format (XDF) files for offline analysis. At the time of system development, MATLAB offered stable bindings for LSL and XDF and therefore served as the primary analysis environment. Its established signal-processing toolboxes facilitated reliable preprocessing, filtering, and statistical operations without requiring additional implementation overhead. However, the framework is not dependent on MATLAB. Equivalent pipelines are now available in Python, including pyxdf for loading recordings, SciPy and MNE-Python for signal processing, and NumPy and pandas for statistical analysis. Core functionality is also supported in GNU Octave, which provides MATLAB-compatible syntax for most basic operations.

The portability of the analysis stack is illustrated in Table 3.1 and Table 3.2. Table 3.1 summarises the core software components of the synchronisation framework, their primary implementation language, and the bindings used in this project. Table 3.2 compares the MATLAB functions applied in the analysis with their equivalents in Python and Octave, demonstrating that the entire workflow can be replicated using free and open-source tools. This ensures that the framework remains both accessible and sustainable, supporting its use in future research and clinical applications without dependence on proprietary software.

Table 3.1: Software components and bindings used in the data synchronisation framework

Component	Primary language	Bindings / Interfaces Used
Unity simulator	C#	LSL4Unity liblsl
Lab Streaming Layer	C++	MATLAB API; C++ (lab recorder and viewer)
OpenFace	C++(OpenCV)/ MATLAB	Called from Unity process (C#); Python API (pylsl)
OpenViBE	C++	TCP/IP and LSL plugins (bridges)
Empatica E4	Proprietary SDK	CSV export; MATLAB post-processing

Table 3.2: MATLAB functions for LSL and XDF data preprocessing and their open-source equivalents

Task	MATLAB function/toolbox	Python	GNU Octave
Load XDF recordings	load_xdf (LSL-MATLAB API)	pyxdf.load_xdf	No direct XDF; import CSV/EDF
Read CSV (Empatica/OpenFace)	readtable, readmatrix	pandas.read_csv	csvread, dlmread, readtable
Degrees ↔ Radians	deg2rad, rad2deg	np.deg2rad, np.rad2deg	deg2rad, rad2deg
Euler → Quaternion (XYZ)	eul2quat(eul,'XYZ') or quaternion(eul,'eulerd','XYZ','frame')	Rotation.from_euler('xyz', euls).as_quat() (returns [x,y,z,w])	No built-in; small helper (compose per-axis quats)
Quaternion multiply / inverse	quatmultiply, quatinv or q1*q2 (quaternion class)	custom funcs or Rotation ops: R2*R1.inv()	small helpers (quatmul, quatinv)
Axis–angle / rotvec from quaternion	quat2axang or rotvec (q) (quaternion class)	Rotation.as_rotvec() *(axis×angle)*	helper: angle = 2*acos(w); axis = v/sin(angle/2)
Angular velocity from quaternion time series	angvel(quat, dt, "frame")(Robotics/Aerospace)	(R[k+1]*R[k].inv()).as_rotvec()/dt*(frame/body rates)*	helper using quat Δ → axis–angle → ω/dt
World ↔ Body (frame) components	angvel(...,"point") for world; or rotate: rotmat(q)*ω_body	R.apply (ω_body) → world components	multiply by rotation matrix from q
Resampling	resample, retime (timetables)	scipy.signal.resample, mne.filter.resample, pandas.resample	resample
Filtering (e.g., BVP)	butter,filtfilt(Signal)	scipy.signal.butter, scipy.signal.filtfilt	butter, filtfilt
Interpolation / sync to markers	interp1, synchronize(timetables)	np.interp, pandas.merge_asof	interp1
Descriptive statistics	mean,std,median,iqr	numpy,scipy.stats,pandas	same functions
Non-parametric statistics	Ranksum (Wilcoxon rank-sum)	scipy.stats.ranksums	ranksum

3.5 Lab Settings and Protocol Guidelines

The Study 1 and Study 2 were conducted under the same lab-controlled condition. This lab space was inspired by the guidelines outlined in ISO 8589:2007 (ISO, 2007). ISO 8589:2007 provides guidelines for test room design that focus on sensory analysis.

3.6 Participants

Lab-based studies

A convenience sampling approach was used to recruit 62 participants for the lab-based studies. However, a subset of 57 participants was selected for analysis in this thesis. Exclusions were based on eligibility criteria during screening or incomplete/inconsistent data that rendered certain cases unsuitable for analysis. Participants who required prescription glasses were allowed to wear them during the experiments to ensure comfort and usability. Participants were assigned into three experimental groups: Desktop group (12 males and 12 females), Headset 1 Group (10 males and 7 females) and Headset 2 Group (8 males and 8 females).

A power analysis was performed using G*Power 3.1 (Erdfelder et al., 2009) to determine the minimum sample size required per group. The analysis indicated that a sample size of 26 participants per group was optimal to detect a large effect size (Cohen's $d > 0.8$), with $\alpha = 0.05$ and power = 0.80 ($1 - \beta$). While the final group sizes ranged between 16 and 24, this range still allowed for the detection of statistically significant differences under the specified parameters.

Field-based Study

For the field-based study, a total of 17 wheelchair users were initially recruited. However, due to technical issues, time constraints, or voluntary withdrawal, only 10 individuals (6 females and 4 males) completed all the virtual tasks and post-assessments. To support comparison, a matched control group of 10 non-disabled participants was also recruited, yielding a final sample of 20 participants (10 wheelchair users and 10 control participants).

A priori power analysis was conducted to determine the minimum sample size required for meaningful statistical analysis. For correlation analyses (e.g., between simulator metrics and reference standards such as WST-Q and MoCA), a total of 29 participants is required to detect a large effect size ($r = 0.5$) in with 80% power at $\alpha = 0.05$ (two-tailed). For group comparisons, an independent samples t-test, requires at least 26 participants per group (52 total) is needed to detect a large effect size ($d = 0.8$) under the same parameters.

Given the practical constraints of field implementation, a target sample range of 10 to 15 individuals per group was selected. This range ensured study feasibility while enabling the capture of meaningful variation in mobility experience and cognitive function. The selected sample size also allowed for exploratory insights and expert feedback from healthcare professionals, while laying a baseline for future validation efforts. This sample size aligns with recommendations for pilot and feasibility studies (Hertzog, 2008; Johanson & Brooks, 2009), balancing between manageability and meaningful data collection goals.

3.7 Ethics Approval

Ethical approval was obtained from the institute's ethics committee in advance of the experiments. The ethics application consisted of a research proposal, a risk assessment, a study design and a protocol. Consent was obtained in accordance with Declaration of Helsinki and ethical approval was obtained from the Technological University of the Shannon (previously known as Athlone Institute of Technology) Research Ethics committee (REC).

3.8 QoE-Based Assessment Approach Guidelines

The overall QoE-based assessment approach was implemented through a three-phase structure: Pre-Test, Test, and Post-Test, each targeting different aspects of the user experience.

These phases were designed to be flexible across lab and field settings while maintaining methodological consistency. The process, as illustrated in Figure 3.6, begins with participant briefing and consent, followed by screening and baseline data collection. During the Test phase, users first undergo a tutorial or free practice session before performing the experimental tasks. Finally, the Post-Test phase captures both structured questionnaire responses and qualitative feedback through open-ended questions.

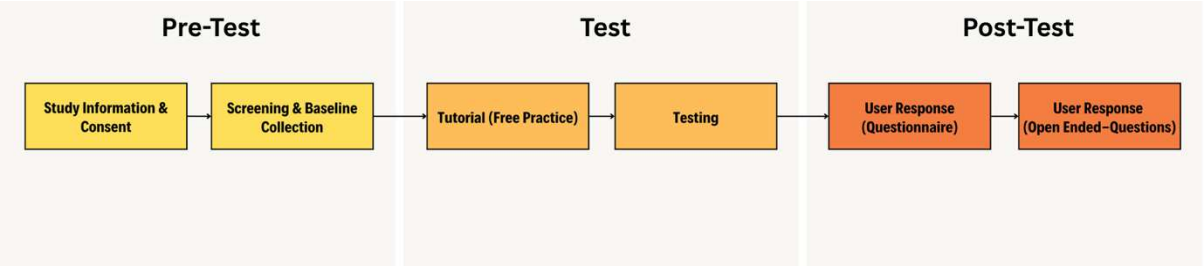


Figure 3.6: Overview of the QoE-based assessment procedure applied in this study.

This structured approach provided a framework for data collection across studies contexts, supporting the multidimensional evaluation of user experience. The following section outlines each component in detail.

3.8.1 Pilot Testing

Prior to the commencement of the main experimental trials, a series of pilot tests were conducted to evaluate the feasibility, clarity, and applicability of the proposed protocol methodologies across the three studies. These sessions served as iterative validation steps, allowing the research team to refine experimental procedures, technical configurations, and participant interaction protocols based on real-world feedback.

The pilot testing involved small samples of participants, including QoE researchers, assistive technology specialists, and representative end-users (e.g., wheelchair users or individuals with cognitive or motor impairments). These stakeholders contributed valuable insights regarding task design, simulator interface usability, duration and sequencing of assessments, and the acceptability of physiological sensors and feedback instruments.

The feedback gathered during pilot testing was instrumental in shaping a participant-centred, context-aware assessment flow, which improved the clarity of instructions, check the robustness of data collection procedures, and ensured the overall accessibility and acceptability of the simulator experience across diverse user profiles.

3.8.2 Pre-Test Methods

In the pre-test phase, participants were first provided with detailed study information and asked to give informed consent, in accordance with the approved ethical protocols.

Lab-based Studies

Lab-based studies followed a structured screening procedure to verify participant eligibility. The initial screening assessed exclusion criteria, including a history of epilepsy, inadequate sleep (defined as fewer than six hours of sleep the previous night), suspected pregnancy, and alcohol consumption within the past 24 hours. Subsequently, participants underwent visual acuity and colour perception tests.

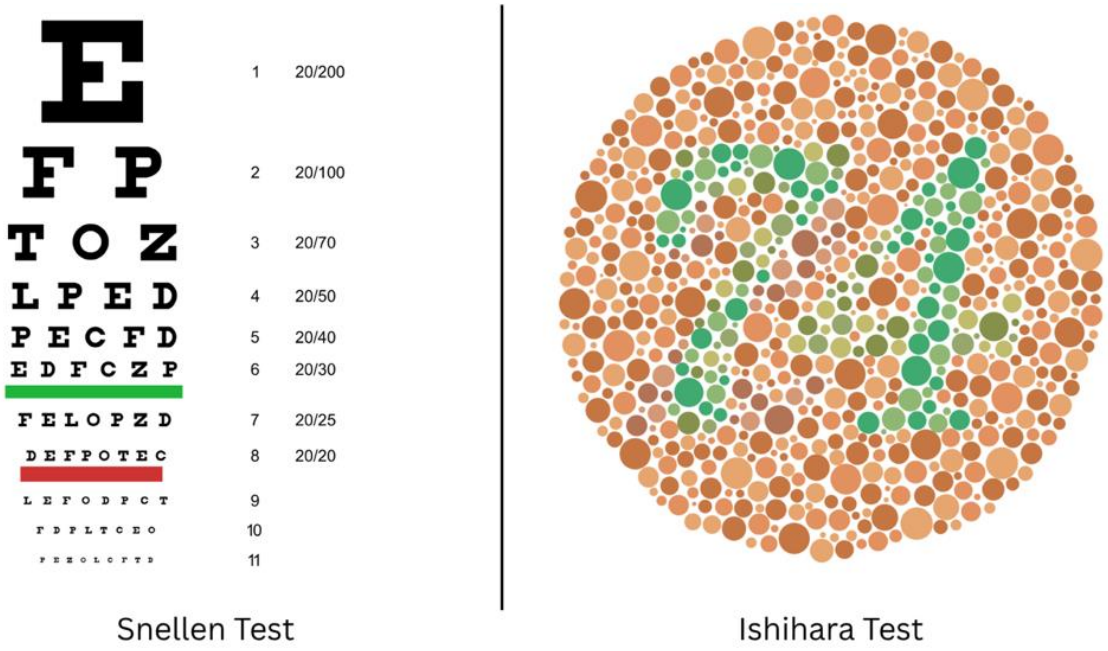


Figure 3.7: Snellen (left) and Ishihara (right) Tests. Images from (Hoffmann & Menozzi, 1999; Sue, 2007).

Visual acuity was evaluated using a Snellen chart (Sue, 2007), with a minimum requirement of 20/20 vision to pass (Figure 3.7 (left)). Colour perception was assessed using the Ishihara test (Hoffmann & Menozzi, 1999), which consists of 38 coloured plates designed to detect red-green colour deficiencies (Figure 3.7 (right)). Participants were permitted a maximum of four errors on this test to be considered eligible.

Field-based Study

In the field-based study, pre-test procedures were integrated into participants' routine clinical activities to accommodate their physical and cognitive conditions. Pre-screening and screening were conducted in close collaboration with clinical team at the IWA centres. Participants were assessed against the inclusion and exclusion criteria, and cognitive screening using the Montreal Cognitive Assessment (MoCA) (Nasreddine et al., 2005) was carried out for both control participants and wheelchair users

The Wheelchair Skills Test Questionnaire (WST-Q) for power wheelchair users (Mortenson et al., 2018; Rushton et al., 2016) was also administered during this pre-assessment stage to both groups. For control participants, who had no prior experience using a wheelchair, the WST-Q was repeated after the simulation session to explore whether their confidence or perceived skills changed following the virtual wheelchair experience.

Baseline measurements

Baseline data collection was conducted across all studies. However, in the field-based setting, a five-minute resting phase conducted in isolation was not feasible. Instead, baseline physiological signals were recorded passively while participants provided demographic information, seated in as relaxed a position as possible during the initial part of the session.

3.8.3 Test Methods

The test phase focused on participants' interaction with the virtual wheelchair simulator. In the lab-based study, this phase included a free-practice session followed by guided familiarization with the virtual environment. These steps ensured that participants could effectively navigate the virtual space and operate the control interface prior to the experimental task.

Lab-based Studies

Participants then completed a predefined route, designed based on tasks from the Power Mobility Training and Assessment Tool (PMRT) (Massengale et al., 2005). This task was performed under display conditions (desktop or immersive) and motion profiles (high or low jerk), allowing for comparative analysis of motion perception, usability, and physiological responses.

Field-based Study

In contrast, the field-based study emphasized ecological validity and user-led adaptation. Participants were introduced to the desktop version of the simulator and allowed to repeat parts of the free-practice session as needed. Joystick acceleration and speed settings were adjusted according to individual preferences to improve comfort and control. Tasks execution followed a clinically meaningful sequence based on WST (Kirby, 2017; Rushton et al., 2016) and PMRT (Massengale et al., 2005) task structures, reflecting real-world training and assessment goals.

Throughout the test phase across all studies, continuous physiological monitoring (e.g., electrodermal activity, heart rate) and system-based performance metrics (e.g., task duration, number of commands, collision count) were recorded. These data sources are described in more detail in the following sections.

3.8.4 Post-Test Methods

The post-test phase gathered both quantitative and qualitative feedback on participants' experience with the simulator. All participants completed a set of validated questionnaires designed to assess key

dimensions of Quality of Experience (QoE), including usability, immersion, emotional response, cybersickness and cognitive workload.

Lab-based Studies

In the lab-based studies, questionnaires were administered in a controlled setting immediately after the simulation session, allowing for the collection of uninterrupted and timely self-reported feedback.

Field-based Study

In the field-based study, post-test assessments were also conducted immediately after the simulator experience. However, to minimize fatigue and cognitive burden, adjustments were made during administration. For some participants, the questions were read aloud, and participants responded verbally, allowing for a more accessible and supportive feedback process adapted to individual abilities.

In addition to the questionnaires, qualitative feedback was gathered through open-ended responses from wheelchair users' group. These insights proved valuable in identifying contextual barriers, individual adaptation needs, and broader reflections on the simulator's perceived relevance for training and assessment in real-world clinical contexts.

In summary, the procedures outlined across the pre-test, test, and post-test phases enabled comprehensive data collection. The following section details the explicit and implicit measures used to evaluate participant experience, performance, and physiological responses across study conditions.

3.9 Explicit Measures

This section presents explicit instruments (questionnaire and rating-scales) used across the studies. These measures were selected to evaluate user responses across multiple dimensions: usability, immersion, emotional reaction, perceived workload, cognitive function, and mobility skills. A suite of validated self-report and performance-based instruments was employed.

To evaluate usability, a five-item version of the System Usability Scale (SUS) (Brooke, 1996) was used, where participants rated their agreement with statements about the simulator interface on a 5-point Likert scale. Higher scores indicate better perceived usability. Sense of Presence was assessed using a short-form version of the Igroup Presence Questionnaire (IPQ) (Igroup Project Consortium, 2015), with one item each for spatial presence, involvement, realism, and general presence, all rated on a 5-point scale.

Table 3.3 provides an overview of the instruments used and the studies in which they were applied.

To evaluate usability, a five-item version of the System Usability Scale (SUS) (Brooke, 1996) was used, where participants rated their agreement with statements about the simulator interface on a 5-point Likert scale. Higher scores indicate better perceived usability. Sense of Presence was assessed using a short-form version of the Igroup Presence Questionnaire (IPQ) (Igroup Project Consortium, 2015), with one item each for spatial presence, involvement, realism, and general presence, all rated on a 5-point scale.

Table 3.3: Explicit Measures Instruments

Instrument	Assessed Dimension	QoE Influencing Factors	Study 1	Study 2	Study 3
SUS (Brooke, 1996) – short version	Usability	System	✓	✓	✓
IPQ (Igroup Project Consortium, 2015) – short version	Sense of Presence (SoP)/ Immersion	System, Human	✓	✓	✓
SAM (Bradley & Lang, 1994)	Emotional Response	Human	✓	✓	✓
NASA-TLX (Hart, 2006; Hart & Staveland, 1988)	Cognitive Workload (task/session level)	Human, System	✓	✓	✓
SSQ (Balk et al., 2013; Kennedy et al., 1993)	Simulator-induced discomfort	Human		✓	
Paas (F. Paas et al., 2003; F. G. W. C. Paas & Van Merriënboer, 1994a)	Mental Effort (task-level)	Humna, Task			✓
MoCA (Nasreddine et al., 2005)	Cognitive abilities	Human, Context			✓
WST-Q (Mortenson et al., 2018; Rushton et al., 2016)	Confidence in mobility skills	Human, Context			✓
Custom Usability and Experience Questions	Quality of experience related to user- simulator interaction and general service level	Human, System			✓

For emotional response, the Self-Assessment Manikin (SAM) (Bradley & Lang, 1994) was administered to measure affective responses across three dimensions: valence (pleasure-displeasure), arousal (calm-excited), and dominance (controlled-in control). It was administered after simulation session to assess emotional state during the experience.

The Simulator Sickness Questionnaire (SSQ) was applied only in Study 2, where VR headsets were used. Participants rated symptoms across nausea, oculomotor discomfort, and disorientation categories, each scored from 0 (none) to 3 (severe). Total and subscale scores were computed using standard SSQ formulas, excluding the general symptom category (Balk et al., 2013; Kennedy et al., 1993).

Perceived workload was measured using the NASA Task Load Index (NASA-TLX), which asks participants to rate six dimensions: mental demand, physical demand, temporal demand, performance, effort, and frustration. The weighted average of these components provides an overall workload score. Paper-based administration followed standardized TLX scoring procedures. (Hart, 2006; Hart & Staveland, 1988).

In Study 3, the Paas Mental Effort Scale was also used to complement NASA-TLX. This single-item scale asks participants to rate their mental effort on a 9-point scale, from “very, very low” to “very, very high” (F. Paas et al., 2003; Reedy, 2015).

To evaluate baseline cognitive function, the Montreal Cognitive Assessment (MoCA) (Nasreddine et al., 2005) was administered in Study 3. This standardized screening tool evaluates a range of cognitive domains, including attention, memory, visuospatial ability, executive function, and language. One example of a task included in the MoCA is the clock-drawing exercise, which assesses visuospatial and executive skills (see Figure 3.8). The MoCA was used to screen for potential cognitive impairments and to explore associations between cognitive ability and simulator performance.

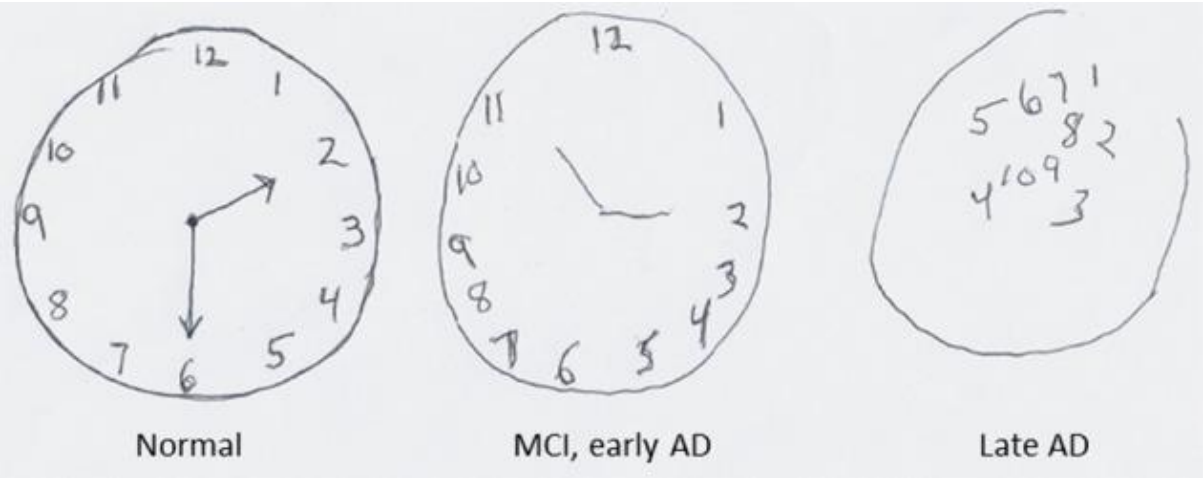


Figure 3.8: Example of one activity from MoCA (clock drawing task). Image from (Mattson, 2014).

Finally, to assess confidence in mobility skills, the Wheelchair Skills Test Questionnaire (WST-Q) (Mortenson et al., 2018; Rushton et al., 2016) was used in Study 3. Administered to both wheelchair users and control participants, it aimed to evaluate perceived skill level and, for controls, detect any change in confidence after simulator exposure.

3.10 Implicit Measures

This section describes the collection and use of physiological signals as implicit indicators of participants’ emotional and cognitive responses during simulator use.

3.10.1 Physiological metrics

Throughout the experiment, a continuous stream of physiological data was collected to capture users' implicit responses across all phases of simulator interaction. These data were acquired using Empatica wearable devices, including the E4 wristband and, in the third study, the EmbracePlus device. Both devices are medically certified, with CE marking and FDA-clearance as part of the Empatica Health Monitoring Platform, supporting their use in research and clinical-grade data collection.

Figure 3.9: Empatica wristband devices. Images from (Empatica, 2025b).

The E4 wristband includes four sensors to determine the blood volume pressure (BVP) at a sample rate of 64 Hz, inter-beat interval (IBI), Heart Rate (HR) in a sample rate of 1 Hz, electrodermal activity (GSR/EDA) at a sample rate of 4 Hz, XYZ raw acceleration at a sample rate of 32 Hz and the skin temperature at a sample rate of 4 Hz. The skin conductance response (SCR) was extracted from the EDA during the experience. The peripheral skin temperature was used to check if the EDA signal oscillations were not affected by external temperature changes.

The EmbracePlus device used in Study 3 featured upgraded sensors, including a ventral EDA sensor, multi-wavelength photoplethysmography (PPG), an accelerometer and gyroscope, and a digital temperature sensor. Data were logged and synchronized via the Empatica Care platform. With improved sensor and the same certified medical status as the E4, EmbracePlus enabled regulatory-grade physiological monitoring under ecologically valid conditions.

During Study 3, participants wore the EmbracePlus on the non-dominant hand and the E4 on the dominant hand. This configuration was chosen to allow for exploratory analysis of hand/wrist control behaviour using the E4's 3-axis accelerometer.

The adoption of wearable physiological sensors enabled continuous, non-intrusive monitoring of autonomic responses, supporting the investigation of user experience and physiological engagement under realistic simulation conditions.

3.10.1.1 Rationale to capture physiological metrics

The autonomic nervous system (ANS) is key regulatory component of human physiology that operates largely unconsciously to maintain homeostasis and adapt to environmental demands. It modulates vital functions such as heart rate, respiration, and glandular activity, and plays a central role in the body's physiological response to stress or cognitive demand.

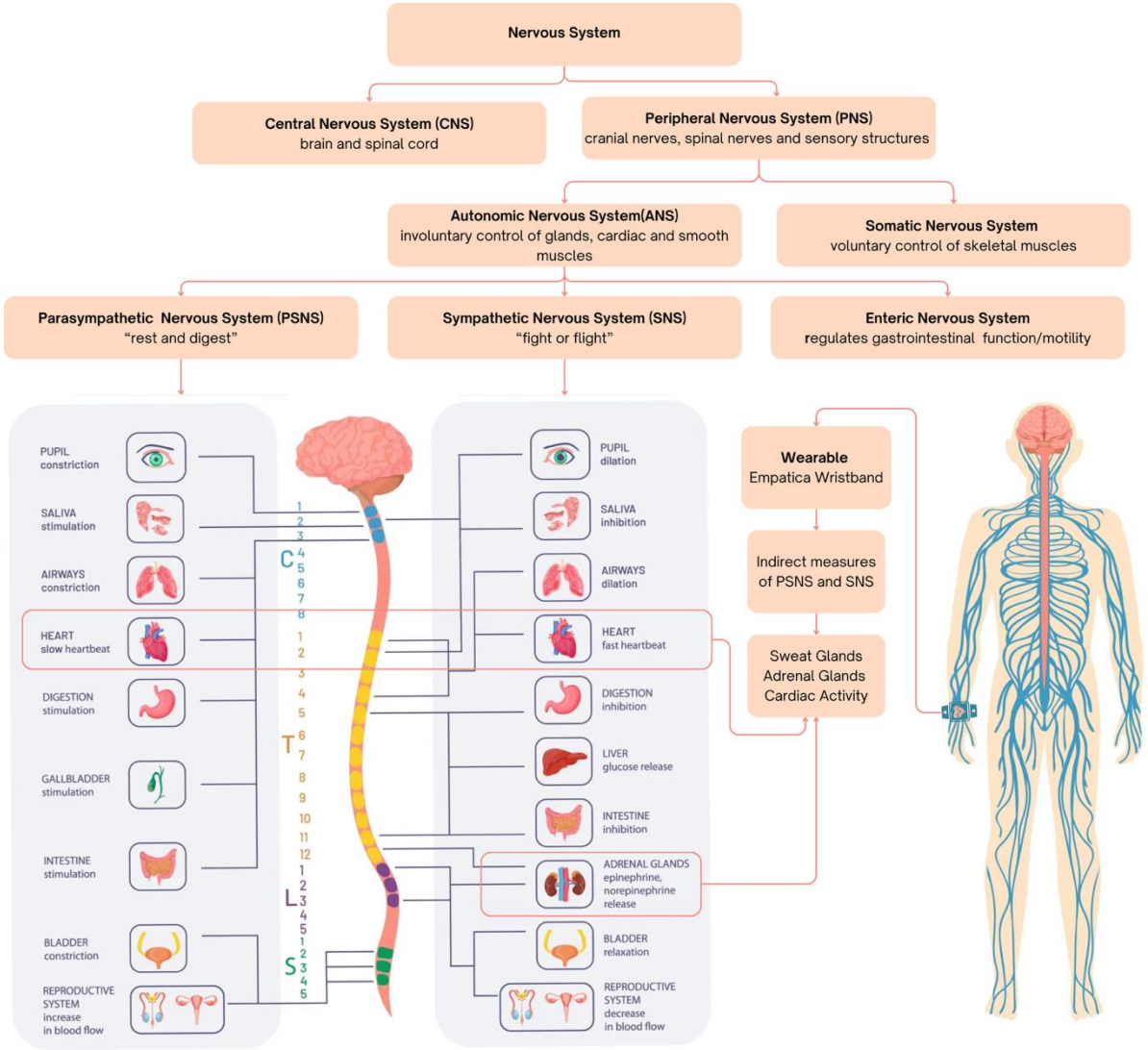


Figure 3.10: Overview of the human nervous system with emphasis on the autonomic nervous system (ANS). Image adapted from (Guy-Evans & McLeod, 2025).

The ANS comprises two complementary branches (see Figure 3.10): the sympathetic nervous system (SNS), responsible for activating the "fight or flight" response, and the parasympathetic nervous system (PSNS), which promotes "rest and digest" functions and recovery states (Athif et al., 2020; Giannakakis et al., 2022). There are many studies related to the Autonomic Nervous System (ANS) to observe the human behaviour (Banu & Nagaveni, 2023; Giannakakis et al., 2022).

During periods of heightened arousal, cognitive effort, or emotional stress, sympathetic activation leads to increases in heart rate (HR), respiration rate, and electrodermal activity (EDA), as sweat glands become active. Once the stressor subsides, parasympathetic activity predominates to reduce arousal and facilitate physiological recovery. As such, monitoring cardiac and electrodermal signals provides a non-invasive approach to assessing autonomic nervous system activity and can offer insights into user states during interaction with virtual environments or training simulators (Gullett et al., 2023; Lima et al., 2020; Ronca et al., 2023).

Figure 3.10 provides a structured overview of the nervous system, with emphasis on the ANS and its influence on key organ systems. The left side of the figure depicts how the SNS and PSNS differentially regulate organ responses (e.g., heart rate, pupil dilation, respiratory rate), while the right-side maps these physiological changes to biometric signals that can be captured using wearable sensors, such as the Empatica wristbands employed in this study. These include heart rate (HR), heart rate variability (HRV), inter-beat interval (IBI), and electrodermal activity (EDA).

To contextualize how these signals reflect autonomic activity during simulator use, Table 3.4 summarizes the physiological metrics captured in this study and their associations with sympathetic and parasympathetic activation. These mappings are grounded in neurophysiological literature and help justify the use of EDA and HR-derived metrics (HRV, IBI) as proxies for emotional and cognitive state monitoring in virtual tasks.

Table 3.4: Mapping of physiological metrics to SNS and PSNS activities.

Physiological Metrics	Sympathetic Nervous System("Fight or Flight")	Parasympathetic Nervous System ("Rest and Digest")
Hear Rate (HR)	Increases heart rate as a response to stress, arousal, or cognitive demand.	Decreases heart rate as the body returns to a relaxed or recovered state.
Heart Rate Variability (HRV)	Decreases HRV, indicating reduced vagal tone and heightened stress/arousal.	Increases HRV, reflecting greater vagal tone and relaxation or task disengagement.
Electrodermal Acitivity (EDA)/ Skin Galvanic Response (GSR)	Increases due to sweat gland activation from heightened sympathetic arousal (e.g., stress, emotional intensity, cognitive load)	No significant change; parasympathetic state does not activate sweat glands

Although HF HRV analysis, typically defined as the 0.15–0.4 Hz frequency band, is a validated method for assessing parasympathetic modulation, it was not applied in this study. The decision was based on known limitations of wrist-worn photoplethysmography (PPG)-derived inter-beat interval (IBI) data, especially during active simulator use . Such signals are prone to motion artifacts, irregular sampling, and occasional signal loss, which compromise the reliability of frequency-domain HRV metrics such as HF power (Banu & Nagaveni, 2023; Gullett et al., 2023; Lima et al., 2020; Van Voorhees et al., 2022).

Similarly, while electroencephalography (EEG) is widely used to assess cognitive states, its application in ecologically valid simulation settings introduces practical challenges. As noted by Duncan et al. (Duncan et al., 2009), EEG recordings are prone to contamination from facial muscle movements, eye blinks, and electromagnetic interference. Moreover, clinically accepted EEG procedures often involve the use of wet or gel electrodes and extensive scalp preparation, which can be intrusive, fatiguing, and reduce participant comfort. EEG setups may also limit physical movement, potentially diminishing the user's Quality of Experience (QoE) during interactive tasks (Vlahovic et al., 2022).

Given these constraints, this study prioritized the use of minimally intrusive and low-movement-compatible physiological signals. Specifically, electrodermal activity (EDA), heart rate (HR), inter-beat interval (IBI), time-domain HRV metrics (SDNN and RMSSD), and 3-axis wrist acceleration were captured using wearable sensors. These measures were selected to enable continuous monitoring without restricting user mobility or compromising the ecological validity of the simulation environment. HRV metrics were computed over the full task duration and compared to baseline values to assess task-induced autonomic variation under realistic interaction conditions.

3.10.2 Simulator-performance based indicators

The performance metrics included joystick events, time to complete and the number of collisions and were captured inside the application (Unity platform). These metrics are widely used in the QoE and Quality of Service (QoS) evaluations(W. Li et al., 2016). Unlike real-world environments, simulated contexts allow for safe, repeatable measurement of quantitative indicators that reflect operational control, trajectory accuracy, and situational awareness during powered wheelchair use.

The selected indicators are grounded in prior research and adapted from the Quality Driving Metrics (QDM) proposed by (Kamaraj, 2020). These metrics have also been featured in other simulator studies, including the works of (A. R. de Sá et al., 2022; P. S. Archambault et al., 2011; F. Martins, 2022), among others. Given the constraints of a virtual indoor environment

and the goal of developing an efficient and scalable evaluation system, four key performance indicators were chosen for this study: number of collisions, task completion time, number of commands and root-mean-square error (RMSE) implemented later in Study 3**.** These indicators offer a balance between interpretability, computational efficiency, and relevance to driving proficiency.

I) Number of Collisions

Collisions are recorded as discrete events whenever the virtual wheelchair makes contact with obstacles (e.g., walls, cones, furniture) after task initiation. This metric reflects user safety and spatial awareness. Higher collision counts suggest poor control or limited anticipatory behaviour, while near-zero collisions indicate safe navigation. Simulation allows safe measurement of this safety-critical metric, which is widely used in the literature (John et al., 2018).

II) Time to Complete a Task

Task duration reflects navigation efficiency, decision-making speed, and confidence. It is measured from the first joystick input to task completion. Shorter times, particularly when paired with low collisions, suggest better control and goal-directed behaviour. This is one of the most frequently used metrics in simulator research (A. R. de Sá et al., 2022; Hafid & Inoue, 2006; Hernandez-Ossa et al., 2017; Mahajan, 2012; Morère et al., 2018; Zatla et al., 2015).

III) Number of Commands

This metric counts discrete joystick input changes, reflecting user control strategy. Erratic or excessive commands may signal uncertainty, poor planning, or impaired motor control. Commands include nine types: forward, reverse, left, right, their directional combinations, and stop. Continuous movement in one direction is counted as a single command. This metric has been used in studies on input behaviour (P. S. Archambault et al., 2011; Morère et al., 2018).

IV) Root-Mean-Square Error (RMSE)

RMSE measures the deviation between the user's actual driving path (nominal trajectory) and an ideal or pre-defined trajectory. It is calculated based on Cartesian coordinates (x, y) sampled at 60 Hz and uses the standard RMSE formula.

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left[(x_i - x_i^*)^2 + (y_i - y_i^*)^2\right]}$$

Where:

- (x_i, y_i) are user's actual positions
- (x_i^*, y_i^*) are the corresponding ideal trajectory points
- n is the total number of sampled points.

Higher RMSE values indicate greater deviations from the optimal trajectory, suggesting difficulties in steering accuracy and path planning. This metric is particularly valuable in assessing navigation precision and has been adopted in multiple simulation studies (Hafid & Inoue, 2006; Mahajan, 2012; Morère et al., 2018). Figure 3.11 illustrate the sequences of coordinates as nominal trajectory and ideal trajectory (pre-defined), The RMSE is calculated from comparing the nominal trajectory to the ideal one.

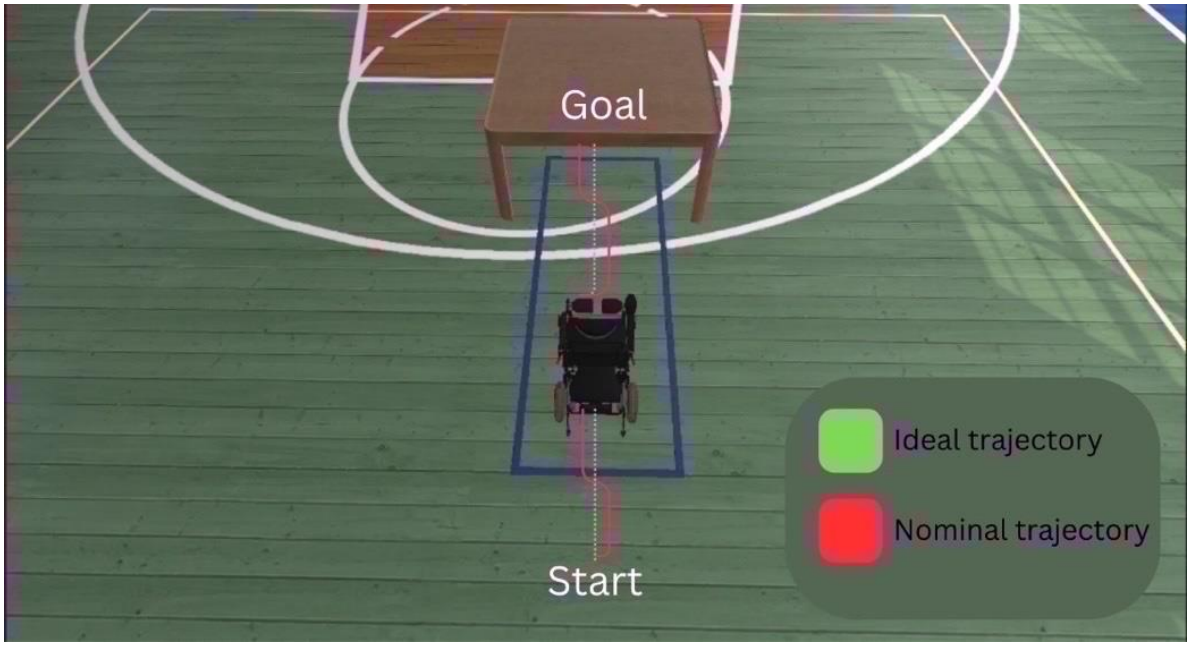


Figure 3.11: RMSE representation of difference between the ideal and nominal trajectory.

3.11 Data Analysis Pipeline Overview

To support data processing and analysis across the studies, a MATLAB-based tool named WheelSimAnalyser was developed. This tool was designed to streamline preprocessing and extract descriptive insights from the dataset generated in the two initial studies, referred to as the WheelSimPhysio-2023 dataset". The primary aim of this tool was to simplify the workflow and reduce manual handling by automating key steps from data import to visualization (Figure 3.12).

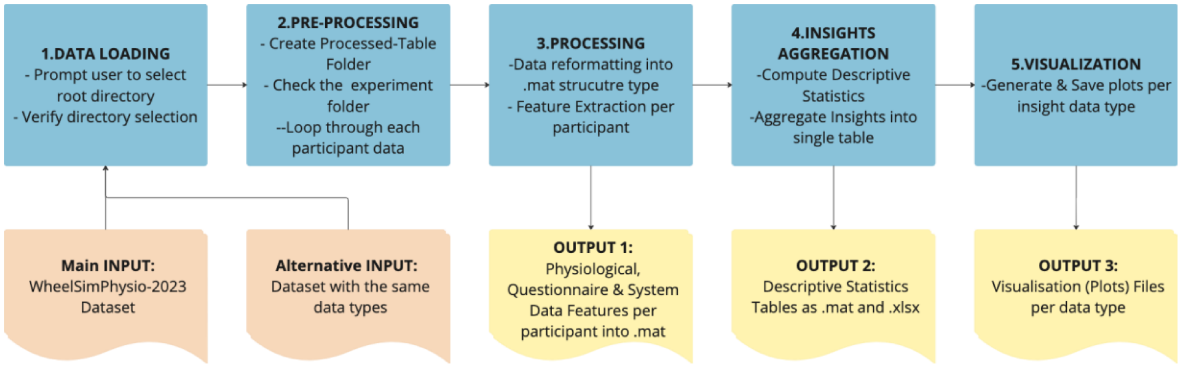


Figure 3.12: WheelSimAnalyser detailed summary of the processes and output involved in the data pipeline analysis.

The key main processes of data analysis are:

- **Data Import**: The user selects the root directory containing the dataset (e.g., WheelSimPhysio-2023). To run the tool, users clone or download the GitHub repository, execute *wheelSimAnalyser.m*, and follow prompts to select the appropriate folder. Detailed usage instructions are provided in the repository.
- **Data Pre-Processing**: After loading, a *Processed-Table* directory is generated to store cleaned data. The pipeline scans each participant folder and processes subdirectories for physiological signals, questionnaire data, and simulator logs. It flags missing data and organizes file paths and metadata automatically.
- **Data Processing (reformatting and feature extraction)**: This step extracts features from physiological metrics (e.g., heart rate, EDA), questionnaire responses (e.g., usability, emotional state, cognitive load), and performance metrics (e.g., task time, errors). The data processing workflow comprises four stages: I) File retrieval and synchronization across modalities; II) Extraction of relevant features; III) Difference calculations (e.g., test vs. baseline) and IV) Saving structured data into *.mat* files.
- **Data Aggregation:** Aggregated datasets are compiled into .mat and .xlsx files, containing physiological, questionnaire, and performance metrics along with descriptive statistics. A full list of extracted features is included in Appendix C.
- **Visualization**: To facilitate interpretation, the tool generates several visual outputs including box plots, violin plots, bar charts, and histograms (see Figure 3.13). The visualization pipeline includes I. extraction and formatting of key numerical data by experimental condition; II. combination of features across experiments for unified display; III. Plot generation using built-in and third-party tools (e.g., violin plots via (Holger Hoffmann, 2015)); IV. saving figures and exporting summary statistics to Excel for reporting.

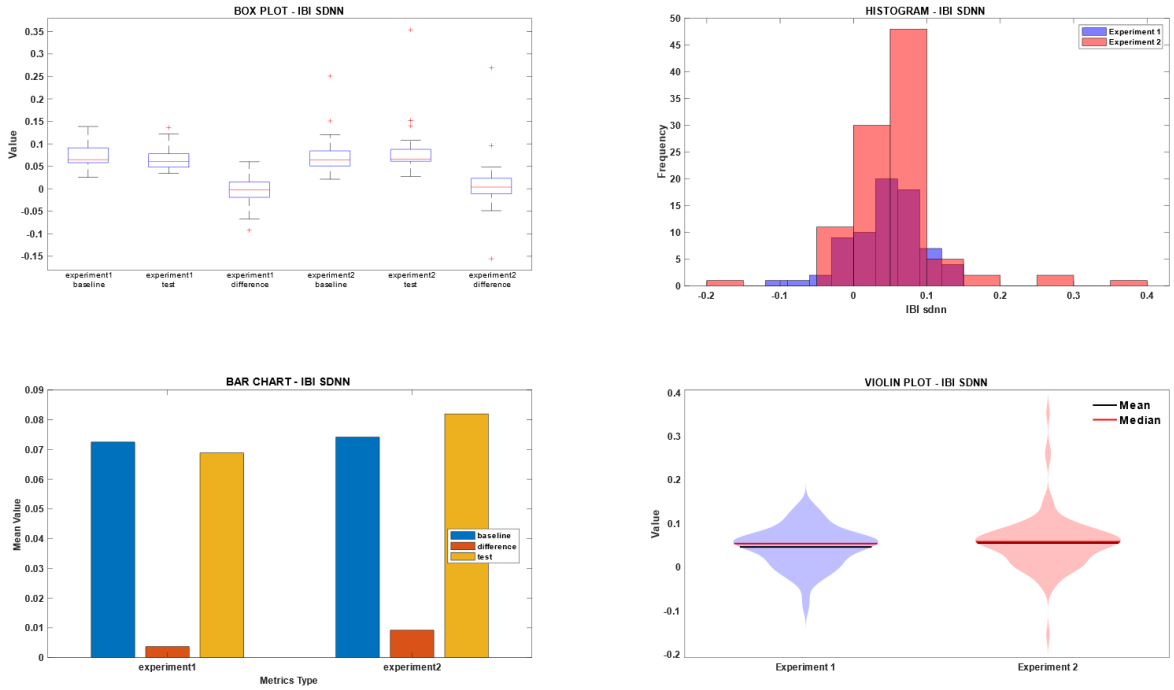


Figure 3.13: Mean values of SDNN (Standard Deviation of NN intervals) from inter-beat interval data (IBI) across desktop (experiment 1) and immersive groups (experiment 2).

3.12 Data Processing

This section outlines the specific procedures used for data filtering, feature extraction, and scoring applied to the physiological signals, simulator-derived metrics, and questionnaire-based assessments used in the presented thesis.

3.12.1 Questionnaires

Questionnaire data were scored according to the instructions provided by each instrument. For multi-item scales like the short versions of SUS, IPQ, and NASA-TLX, scores were averaged to produce a total value. Paas Scale was recorded after each task and SAM after the test. The MoCA was scored to reflect baseline cognitive function, and the WST-Q was completed before and after the simulation to assess perceived wheelchair skills. All scores were organized per participant and matched to their simulator session.

3.12.2 Physiological Data Processing and Feature Extraction

Physiological metrics were recorded during both a resting baseline phase and the subsequent simulator test phase. Signals included EDA, HR, and IBI, captured via wearable sensors. These signals were used to quantify autonomic nervous system responses during virtual simulator-based tasks.

To ensure interpretability and account for inter-individual physiological variability, a baseline correction approach was employed. This method involved subtracting and or dividing the participant-specific mean from the resting baseline period (5 minutes window) from values recorded in during the test phase or within specify event-related time windows. These techniques are commonly used in Quality of Experience (QoE) research in extended reality (XR) studies to quantify physiological changes relative to a resting state(Hynes et al., 2023; Rodrigues et al., 2022). Baseline correction is also a standard practice in broader psychophysiological research, including applications such as estimating cybersickness (Dennison et al., 2016) and assessing arousal (Azbel-Jackson et al., 2016). By anchoring the analysis to a resting state, baseline correction supports within-subject comparisons and facilitates between-group interpretation of task-induced physiological changes. The time-series data analysis was segmented into predefined event-related windows, including:

- **Baseline period**, use as reference for normalisation.
- **Test duration** period.
- **Collision periods**, defined as fixed 5-second segments following each simulator-detected collision, during which visual, auditory, and haptic feedback was delivered. To ensure signal independence, overlapping collision windows were excluded using a minimum temporal separation criterion of 7-seconds.
- **First collision period,** defined as 5-second window after the first collision.
- **Non-collision period**, defined as the first error-free window during the test with no detected collisions.

In the early stages of analysis, temporal binning was explored to visualize the evolution of physiological responses throughout the simulator task (see Figure 3.14). This involved dividing the full task duration into ten equal segments (each representing 10% of total time) and computing summary statistics (mean, standard deviation, variance, and standard error) for each bin. The bin width was calculated using the standard formula:

$$Bin\ Width = \frac{\max(t) - \min(t)}{N}$$

Where t is the task time vector and N(10 or 100) is the number of bins. This approach allowed for normalized descriptive comparisons across participants with varying task durations and helped identify general response trends over time.

Previous research has shown that binning physiological time-series data can enhance robustness in exploratory analyses, particularly by managing signal irregularities and reducing the impact of missing data from wearable sensors (Chakrabarti et al., 2023; Darji et al., 2024). Additionally, binning has been applied to structure irregular data for modelling or prediction tasks (Lee et al., 2024; Xiong & Dubin, 2010). While this study did not pursue modelling or imputation, the binning approach provided an initial framework for interpretability and signal exploration.

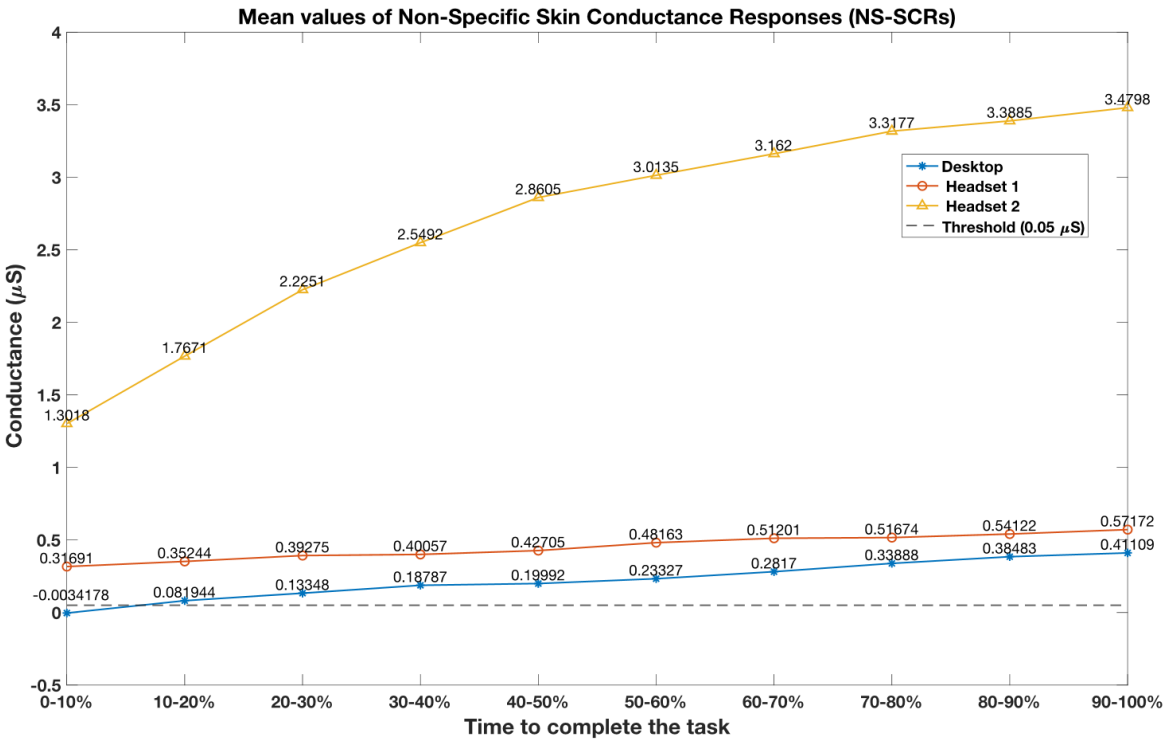


Figure 3.14: Skin Conductance Responses in binned format.

However, in the present thesis, this strategy was not retained in the final analysis. Instead, the adopted approach focuses on event-related analysis and baseline correction, which enable physiologic grounded basis analysis of short-term responses. Specifically, the final analysis concentrated on the baseline, test, and first collision segments. This decision was based on the observation that other segments, such as collision and non-collision windows, exhibited high variability in window size and number of samples per participant, which compromised consistency and interpretability in group-level comparisons.

By focusing on consistently structured and comparable segments, the analysis aimed to improve the reliability and interpretability of task-induced autonomic responses, while acknowledging the limitations posed by data variability. The following sections describe the specific processing and feature extraction procedures applied to each physiological.

3.12.3 Electrodermal Activity Signal (EDA)

Electrodermal activity (EDA) signals were collected using two different wearable devices across the studies. For the lab-based studies (Study 1 and Study 2), EDA was recorded using the Empatica E4 wristband. The following section describes in detail the signal pre-processing and feature extraction procedures applied to these datasets. In contrast, the field-based study employed the EmbracePlus device, a newer generation of Empatica sensors, and EDA data from this study were analysed based on pre-processed physiological biomarkers provided directly by the device.

EDA Signal Pre-processing

Electrodermal activity (EDA) signals were collected using a wearable sensor at a sampling frequency of 4 Hz. Prior to data collection, the physiological sensor and the virtual wheelchair simulator were synchronised to ensure accurate temporal alignment across datasets, as described in Section 3.4. The raw EDA data were imported and converted into numerical arrays for further processing.

A fifth-order Butterworth low-pass filter with a 1 Hz cut-off frequency was applied, using zero-phase forward and reverse filtering (filtfilt, MATLAB) to remove high-frequency noise while preserving the phasic components of the EDA signal (Ronca et al., 2023; Wu et al., 2025). An adaptive check was implemented to ensure the stability of filter coefficients, with filtering skipped if the signal length was insufficient for the given filter order. This filtering approach preserved the integrity of phasic signal components while minimizing artefactual fluctuations.

Following this, the filtered EDA signal was labelled into predefined, event-related time windows. All event-related windows were baseline-corrected by subtracting the participant-specific mean value from a 5-minute resting baseline period, enabling inter-subject comparison of event-related changes in skin conductance, serving as a proxy for phasic sympathetic arousal (Horvers et al., 2021).

Feature Extraction and Event-Related SCR Estimation

Skin conductance (SC) consists of two main components: tonic and phasic. The Tonic skin conductance level (SCL) refers to the slow-varying activity level of EDA. In contrast, phasic activity, also known as Skin conductance response (SCRs), represents the faster changing fluctuations of the EDA signal (Benedek & Kaernbach, 2010; Horvers et al., 2021). Phasic responses reflect short-term, stimulus-

evoked changes in sympathetic nervous system activity, distinguishing them from the more stable tonic component. Event-related SCRs are time-locked to identifiable stimuli or task events, while non-specific SCRs occur spontaneously, without a clearly identifiable external trigger.

In this study, event-related SCRs were estimated as the change of the EDA signal relative to baseline during the test and collision windows. A fixed threshold of 0.05 µS was used to interpret positive arousal responses. Although individual SCR peaks were not explicitly extracted, the use of mean amplitude of event-related windows provided a stimulus-locked approximation of sympathetic arousal, as commonly applied in immersive, QoE-based research settings. All pre-processing and feature extraction steps, and SCR estimations described above were applied for Study 1 and Study 2, allowing assessment of EDA-derived physiological responses under different levels of immersion and simulation design features.

3.12.4 Cardiac Features (HR and IBI)

Cardiac activity was monitored using the same wearable sensors employed for electrodermal activity acquisition. In the lab-based studies, the Empatica E4 wristband provided cardiac signals in the form of inter-beat intervals (IBI) and heart rate (HR). These data streams were aligned with the simulator timestamp recordings via prior synchronisation. In the field-based study, cardiac biomarkers were obtained from the EmbracePlus device, which provides outputs pre-processed physiological features.

IBI and HR Signal Description

Both HR and IBI were derived from BVP signal. HR values were reported in beats per minute (bpm) and recorded as averaged values over 10-second moving window, as described in the manufacturer's documentation (Empatica, 2025b) IBI values represented the duration between consecutive heartbeats and were recorded with a 1/64-second resolution (~15.625 ms). Each IBI entry was timestamped relative to the session's start time (t_o, UNIX time). IBI data were only recorded when the BVP signal quality was sufficient; missing segments were identifiable when the difference between consecutive timestamps did not match the expected interval (e.g., $t_2 - t_1 \neq IBI(t_2)$) (Empatica, 2025a). Both HR and IBI were analysed as time series aligned to the same event-related structure used for EDA.

Cardiac Signals Preprocessing

HR and IBI data were extracted directly from the E4's timestamped output. Artefactual IBI values, including NaNs and zero values, were excluded prior to analysis. Cardiac data were analysed using the same event-related time windows applied in the EDA analysis.

Feature Extraction and Temporal Profiling

The following features were computed from the cardiac signals:

- Mean HR and mean IBI values were calculated for each event-related segment.
- Heart Rate Variability (HRV) metrics were extracted as follows:
    - SDNN (Standard Deviation of NN intervals) was computed from IBI time series within each time duration window, indicating overall autonomic variability:

$$SDNN = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(IBI_i - \overline{IBI})^2}$$

    - RMSSD (Root Mean Square of Successive Differences) was used to capture short-term parasympathetic modulation, reflecting rapid fluctuations in heart rate:

$$RMSSD = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n-1}(IBI_{i+1} - IBI_i)^2}$$

Where:

- IBI_i denotes the inter-beat interval at time index I,
- $\overline{IBI}$ is the mean of all IBI values in the window,
- n is total number of IBI values in the window.

HRV features were computed only for the baseline and entire test duration periods. These longer, continuous intervals ensured reliable HRV estimation by minimizing the impact of signal loss and providing sufficient sampling density. HRV was not extracted for shorter or discontinuous windows (e.g., collision segments) due to their limited duration and the requirement for uninterrupted IBI sequences.

3.13 Head Movements Pre-processing and Feature Extraction

As an exploratory component of the analysis, head movements were examined to assess potential relationships with participants' subjective experiences, including QoE and cybersickness symptoms. Since head motion might be correlate to discomfort, a set of rotational movement features was extracted from both immersive and non-immersive conditions to support comparative analysis (Dennison et al., 2016). These features were not primary outcome variables but were included to explore individual differences in simulator responses.

3.13.1 Immersive Headset Data (Unity3D)

Head movement data during the simulator-based condition were captured using a head-mounted display (HMD) integrated within Simulator at the Unity3D engine platform. Unity records head orientation in both Euler angles (in degrees) and quaternions, based on a left-handed coordinate system.

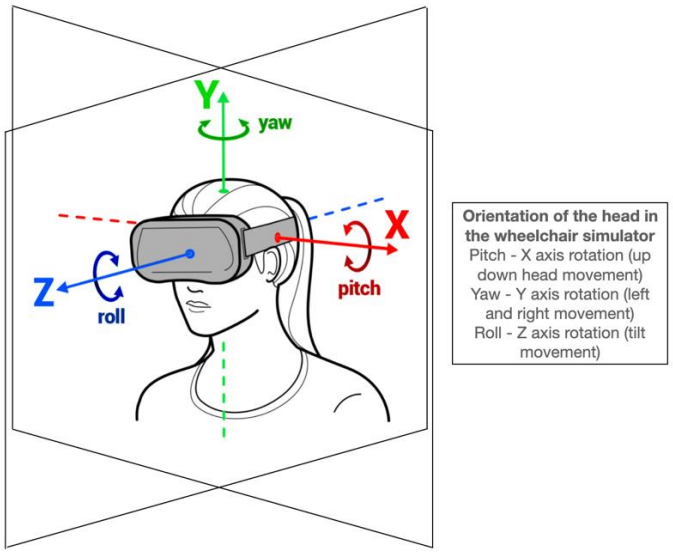


Figure 3.15: Orientation of the head in the wheelchair simulator environment, showing pitch (X-axis), yaw(Y-axis), and roll (Z-axis) rotational axes.

Euler angles represent three rotational degrees of freedom, pitch (X-axis: up/down), yaw (Y-axis: left/right), and roll (Z-axis: side tilt), as illustrated in Figure 3.15 and were converted to radians for consistency. All head tracking data were processed at a sampling rate of 10 Hz, corresponding to a time resolution of 100 milliseconds between samples.

Since the lab-based study involved a single structured task (a ramp navigation route), features were extracted over the entire test duration segment. Two categories of metrics were computed. First, the range of head rotation was calculated per axis by measuring the angular displacement over time. These features (rangePitch, rangeYaw, and rangeRoll) represent the total extent of head movement in radians

Second, angular velocity was computed from the quaternion time series using MATLAB's angvel function with the "frame" option, which estimates body-fixed angular velocity components. This approach calculates quaternion increments, extracts the instantaneous rotation vector, and normalises by the sampling period ($\Delta t = 0.1$s).The resulting mean angular velocities (meanAngVelX,

meanAngVelY, meanAngVelZ, in rad/s). These metrics reflect both the amplitude and dynamic intensity of head movements during the simulator task.

3.13.2 Desktop Non-Immersive Data (Webcam - OpenFace)

For the desktop (non-immersive) condition, head pose data were recorded using the OpenFace toolkit (Baltrusaitis et al., 2016, 2018) (see Figure 3.16), which estimates head position and rotation from real-time webcam video. OpenFace outputs both translation (in millimetres) and rotation (in radians), using a left-handed coordinate system and a composite world-based rotation convention ($R = R_x \cdot R_y \cdot R_z$), with the camera as the origin. Specifically, pitch, yaw, and roll correspond to rotations around the X-, Y-, and Z- axis rotations, respectively. As with the immersive condition, all data were processed at a sampling rate of 10 Hz.

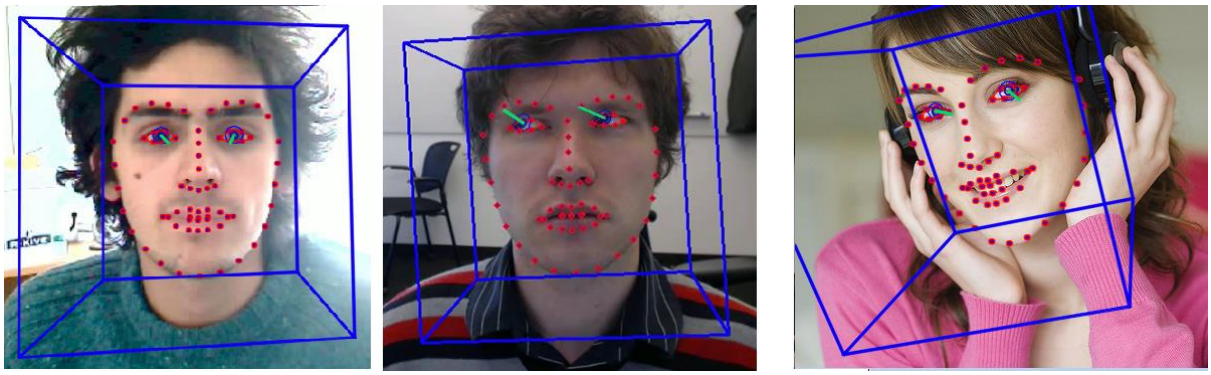


Figure 3.16: Facial and Gaze Tracking Features extracted through OpenFace Toolkit. Imaged From (Baltrusaitis et al., 2016, 2018).

Table 3.5: Summary of Extracted Head Movements Metrics

Metric	Description	Unit
rangePitch	Range of head pitch (up/down) rotation	Radians (rad)
rangeYaw	Range of head yaw (left/right) rotation	Radians (rad)
rangeRoll	Range of head roll (side tilt) rotation	Radians (rad)
meanAngVelX	Mean angular velocity around X-axis (pitch)	Radians/second
meanAngVelY	Mean angular velocity around Y-axis (yaw)	Radians/second
meanAngVelZ	Mean angular velocity around Z-axis (roll)	Radians/second

Head movement metrics were extracted from the test duration segment of the desktop task, matching the same analysis window as in the immersive condition. Ranges of rotation (rangePitch, rangeYaw, rangeRoll) were calculated directly from the Euler series. To quantify rotational dynamics, Euler angles were converted into quaternions consistent with the OpenFace convention, combining sequential axis

rotations as $q = q_x \otimes q_y \otimes q_z$. From this quaternion sequence, angular velocity was computed using the same axis–angle increment approach as in the immersive condition, again implemented in MATLAB with angvel. The mean angular velocities per axis (meanAngVelX, meanAngVelY, meanAngVelZ) were then extracted. This processing enabled a direct comparison of head movement behaviour between immersive (Unity) and non-immersive (OpenFace) test settings.

3.14 Statistical Analysis

Statistical analyses were performed using IBM SPSS and MATLAB. Prior to hypothesis testing, the normality of each variable was assessed using the Shapiro-Wilk test. Variables that violated normality assumptions were analysed using non-parametric methods.

Lab-based studies

In the lab-based studies, which involved between-group comparisons (e.g., Desktop vs. Headset-1 vs. Headset-2), different statistical approaches were applied depending on the distribution of the data. For non-normally distributed outcomes, Kruskal–Wallis H tests were used as a non-parametric alternative to one-way ANOVA to evaluate overall group differences. Pairwise post hoc comparisons were conducted using Mann–Whitney U tests, with Bonferroni correction applied to adjust the significance threshold in the case of multiple comparisons. For normally distributed outcomes, one-way ANOVA with Bonferroni-adjusted post hoc tests was used. When comparing only two groups under normality, independent samples t-tests were applied. Effect sizes were reported using Cohen's d and r (rank-biserial) or η^2(eta squared estimate), as appropriate (J. Cohen, 2013; Computation of Different Effect Sizes like d, f, r and Transformation of Different Effect Sizes: (Lenhard & Lenhard, 2022). Interpretation of the effect size followed (Cohen, 2013) guidelines (see Table 3.6).

Field-based study

In the field-based study, participants completed a series of 12 simulator tasks, each with distinct levels of complexity and type (e.g., obstacle avoidance, narrow turns, etc.). Since tasks were not repeated, the analysis was conducted per task, rather than as a repeated-measures design. Each task was analysed independently to examine differences between wheelchair users and non-users.

Depending on the distribution of the data, either Mann–Whitney U tests or independent samples t-tests were applied. Analyses focused on performance indicators such as task duration, number of collisions, and command frequency, as well as subjective responses including cognitive workload (NASA-TLX, PAAS), emotional dimensions (valence, arousal, dominance via SAM), presence (IPQ),

usability (SUS), and simulator sickness (SSQ). Bonferroni adjustments were used where multiple comparisons across tasks were conducted.

Correlation analyses were conducted using Spearman's rank coefficient (ρ) to examine associations between physiological signals (HR, EDA, HRV), behavioural measures (e.g., head movement), and self-reported metrics (e.g., workload, emotion, presence, usability, simulator sickness), the interpretation of the strength of ρ is presented at . In Study 3, additional correlations were explored between MoCA scores and simulator-based performance metrics (e.g., total collisions, RMSE), as well as subjective ratings from the WST-Q confidence questionnaire.

Table 3.6: Interpretation for difference effect sizes for inferential (between group) analysis from (Cohen, 2013; Lenhard & Lenhard, 2022)

d	r*	η^2	Interpretation sensu Cohen (1988)
< 0	< 0	-	Adverse Effect
0.0	.00	.000	No Effect
0.1	.05	.003	
0.2	.10	.010	Small Effect
0.3	.15	.022	
0.4	.20	.039	
0.5	.24	.060	Intermediate Effect
0.6	.29	.083	
0.7	.33	.110	
0.8	.37	.140	Large Effect
0.9	.41	.168	
≥ 1.0	.45	.200	
* Cohen (1988/2013) reports the following intervals for r: .1 to .3: small effect; .3 to .5: intermediate effect; .5 and higher: strong effect.			

Table 3.7: Interpretation of the relationship (correlation analysis) according to (Schober & Schwarte, 2018)

Absolute Magnitude of the observed correlation coefficient	Interpretation
0.00 – 0.10	Negligible correlation
0.10 – 0.39	Weak correlation
0.40 – 0.69	Moderate correlation
0.70 – 0.89	Strong correlation
0.90 – 1.00	Ver strong correlation

3.15 Summary

This chapter presented the methodological foundations that guided the three experimental studies in this thesis. A QoE-based, mixed-methods approach was adopted, combining explicit measures (e.g., usability, sense of presence/immersion, emotional response, cognitive load) with implicit physiological metrics and simulator-based performance data to evaluate user experience in a powered wheelchair simulator.

The methodology evolved iteratively across the studies:

- Study 1 explored initial QoE responses and simulator usability in lab conditions.
- Study 2 refined the simulator's motion profile and assessed discomfort and sense of presence/immersion.
- Study 3 translated the approach to a real-world setting with wheelchair users, focusing on feasibility and adaptability.

Physiological signals, including electrodermal activity (EDA), heart rate (HR), and inter-beat interval (IBI), were captured using medical-grade wearable devices and processed using event-related segmentation and baseline deviation analysis to support within-subject and between-group comparisons. A temporal binning approach was also explored during early stages for trend visualization but was replaced by more targeted, event-driven analyses in the final framework.

To ensure transparency and reproducibility, a MATLAB-based tool, WheelSimAnalyser, was developed to process the dataset, extract features, and generate descriptive and visual summaries. This tool was applied to the WheelSimPhysio-2023 dataset, a curated dataset compiled from the lab-based studies and made available for public use to support further research.

Simulator-derived metrics such as task time, collisions, joystick commands, and RMSE were aligned with QoE dimensions, allowing for triangulated evaluation of user interaction quality. Validated questionnaires, including short-form SUS, IPQ, NASA-TLX, SAM, Paas, and performance-based tools such as MoCA and WST-Q, provided structured self-report data.

These refined methods enabled a multidimensional, QoE-driven assessment framework, grounded in human-centred design and tailored for both research and clinical settings. The next chapters apply this methodological framework to analyse user performance and experience under varying simulator configurations and participant profiles.

Part IV EXPERIMENTAL STUDIES

Chapter 4 Lab-based Studies – Exploring Simulator Use through Quality of Experience and User Performance

4.1 Introduction

This chapter addresses Sub-Research Question 1 (SRQ1):

> ***"How can a virtual wheelchair simulator be designed and tested in a controlled environment to establish a clinically relevant proof of concept that supports multidimensional assessment, incorporating immersive technologies, physiological signals, subjective feedback, and Quality of Experience (QoE) evaluation?"***

To answer SRQ1, two controlled lab studies were conducted to explore the simulator's experiential, technical, and design dimensions. These studies contributed to the development of a structured evaluation framework for immersive simulator use, aligning with the following thesis objectives:

- **Objective 1.1**: Evaluate QoE in wheelchair simulator use by combining subjective ratings and physiological signals (e.g., electrodermal activity and heart rate).
- **Objective 1.2**: Assess the influence of immersive technology design, particularly head-mounted displays, on usability, performance, cognitive workload, and simulator-induced discomfort (cybersickness).
- **Objective 1.3**: Examine how virtual motion settings, including acceleration and deceleration profiles, affect user experience and system tolerability.
- **Objective 1.4**: Design an initial evaluation framework for potential clinical use, based on the integration of findings across usability, immersion, workload, and physiological response dimensions.

Study 1 addressed Objective 1.1 by establishing a multidimensional QoE model that combined validated questionnaires with physiological monitoring to capture emotional response, usability perception, immersion, and cognitive workload under two system configurations (non-immersive vs immersive with high motion intensity).

Study 2 addressed Objectives 1.2 and 1.3, expanding the protocol to evaluate the effects of immersive display settings and motion design parameters (e.g., high vs low jerk) on user experience and simulator

tolerance. A third configuration featuring smoother motion dynamics was introduced to reduce cybersickness and enhance usability.

Together, these studies contributed to Objective 1.4, by validating the feasibility and value of combining subjective, behavioural, and physiological data to inform simulator evaluation in clinical settings.

The findings from both studies underpin Contribution 1, which developed and applied a multidimensional approach to assess users' QoE in powered wheelchair simulators through controlled studies combining subjective and physiological data, and Contribution 2, which explored how immersive features, such as display types and motion profiles, affect usability, cognitive load, and simulator-induced discomfort, informing design choices such as the adoption of smoother motion profiles and display configurations suitable for clinical settings. These contributions support the simulator's adaptation for clinical assessment and training workflows.

Study 1 – Foundational QoE Assessment of an Immersive Wheelchair Simulator in Controlled Settings

4.1.1 Study 1: Introduction

4.1.1.1 Introduction and motivation

Study 1 was designed to investigate user experience across different system configurations, with a particular focus on identifying QoE indicators relevant to simulator use in mobility contexts. The intention was to assess how immersive features influenced usability, emotional engagement, and cognitive workload, factors of growing importance in the development of virtual reality assistive technologies.

In traditional QoE assessments, post-task questionnaires are widely used but offer only a partial view of user experience. These methods often fail to capture unconscious or dynamic responses and may lack ecological validity in small-sample studies (Vlahovic et al., 2022). In the past years, research underscores the importance of considering Human Influencing Factors (HIFs), such as arousal, stress, and workload, which play a critical role in shaping perceived usability and system tolerance in immersive environments.

To address these limitations of relying solely on self-reported measures, Study 1 adopts a multidimensional QoE-based evaluation approach, building on previous work exploring implicit physiological metrics as indicators of user response during non-immersive and immersive interactions (Eoghan Hynes et al., 2023; Rodrigues et al., 2022). Therefore, this decision was informed by pilot testing discussed at Section 3.8.1, which suggested that certain immersive configurations could induce discomfort and adaptation challenges that were not always reflected in subjective feedback.

4.1.1.2 Study 1: Aim and Hypotheses

The aim of this study was to evaluate QoE through four key experiential dimensions: usability, sense of presence (SoP), emotional response and cognitive workload. These were measured through validated questionnaires (SUS, SAM, IPQ and NASA-TLX). In parallel, physiological responses (EDA, HR, IBI) were recorded using the Empatica E4 wristband to provide implicit indicators of arousal response and cognitive workload.

This study aligns with Objective 1.1 of the thesis and represents the first attempt to triangulate wheelchair simulator usability, immersion, and physiological tolerance in a lab-based evaluation

setting. To explore difference across system configurations (immersive vs non-immersive), the following hypotheses were tested:

- H_1: Participants using the immersive simulator (Headset 1 group) will report higher QoE scores— covering usability (SUS), emotional response (SAM), sense of presence/immersion (IPQ), and workload (NASA-TLX)—than those using the non-immersive (Desktop group) version of Wheelchair Simulator.
- H_2: Physiological responses (EDA, HR, IBI) will significantly differ between the immersive and non-immersive groups, indicating variation in arousal and engagement.

4.1.2 Study 1: Experimental Setup

Study 1 was designed as an exploratory, between-groups experiment to evaluate how different display modalities influence user experience in an immersive wheelchair simulator. A between-groups mixed-methods experimental design was implemented, involving two configurations: (a) a non-immersive desktop system and (b) an immersive setup using a head-mounted display (HMD).

Both configurations used the same virtual task environment and motion control profile to ensure consistency. The system architecture, including input devices, display setups, and data collection components, is summarised in Figure 4.1. This configuration enabled uniform task execution and facilitated comparison between immersive and non-immersive conditions.

Participants completed a predefined navigation task within the simulator (see Figure 4.2), after which they filled out post-experience questionnaires. Throughout the task, physiological signals were recorded to assess arousal, immersion, and cognitive workload, while performance data were simultaneously collected. This initial integration of physiological monitoring within the simulator context informed subsequent methodological refinements, such as the structured inclusion of simulator sickness measures in Study 2.

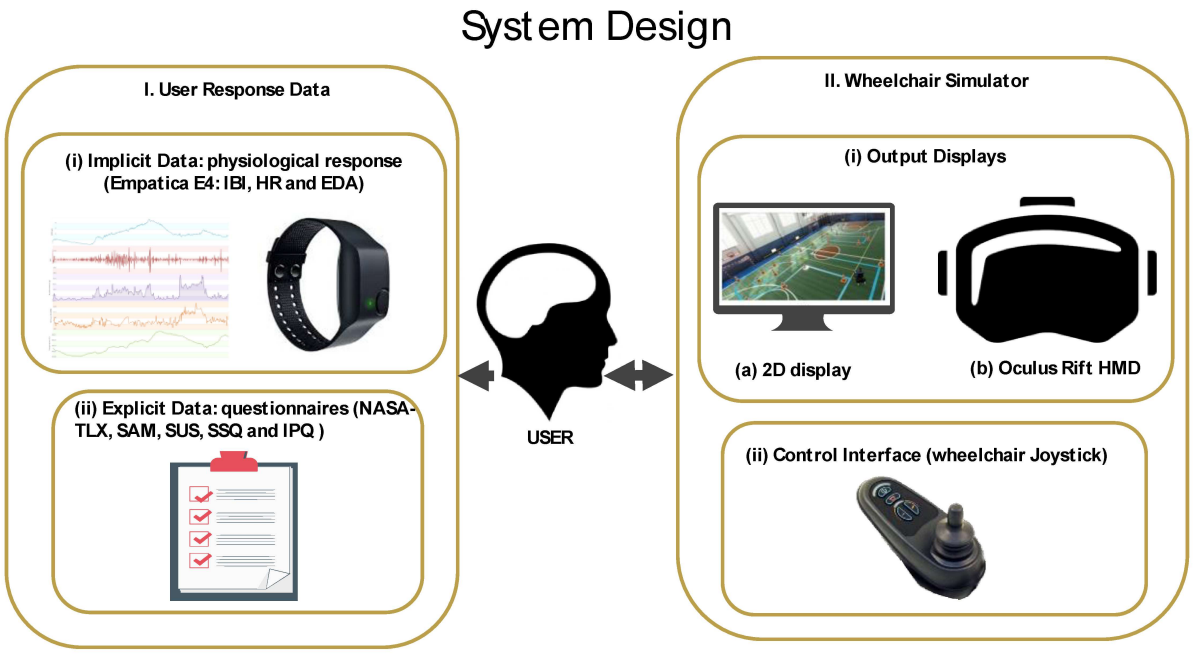


Figure 4.1: Study 1 system design configuration.

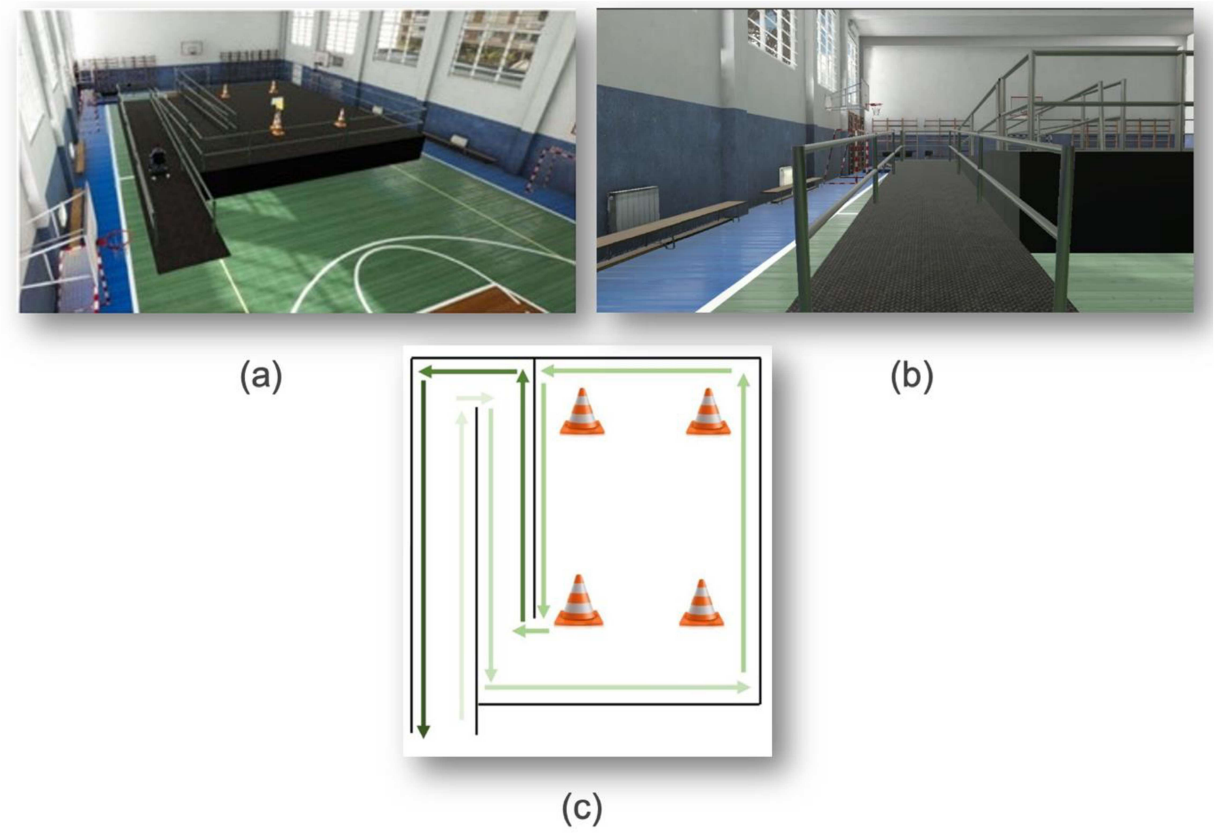


Figure 4.2: Lab-based studies ramp route task view.

4.1.3 Study 1: Powered Wheelchair Simulator System and Task Design

The simulator used in this study was developed in Unity3D (version 2017.2.0f3) to replicate common powered wheelchair navigation scenarios in a virtual environment. Its goal was to offer an immersive yet safe platform for assessment and training, supporting both desktop and VR configurations.

The virtual environment included three task scenarios: an obstacle course, an accessibility ramp, and an elevator navigation task. To ensure consistency and isolate the effect of display modality, only the ramp navigation scenario was used in this study (see Figure 4.2). This task involved ascending and descending a virtual ramp while following directional cues, requiring precise joystick control and spatial awareness.

The system supported immersive and non-immersive configurations: a 22-inch LCD monitor (desktop) and the Oculus Rift DK2 (immersive), both using a first-person viewpoint. In both cases, participants operated the simulator using a physical VR2 joystick configured to closely emulate powered wheelchair control. The immersive condition featured stereoscopic rendering and head tracking to enhance spatial perception. Although the system also supported real-time display of physiological and performance data, this functionality was used only during pilot testing. All data for Study 1 were recorded and analysed offline.

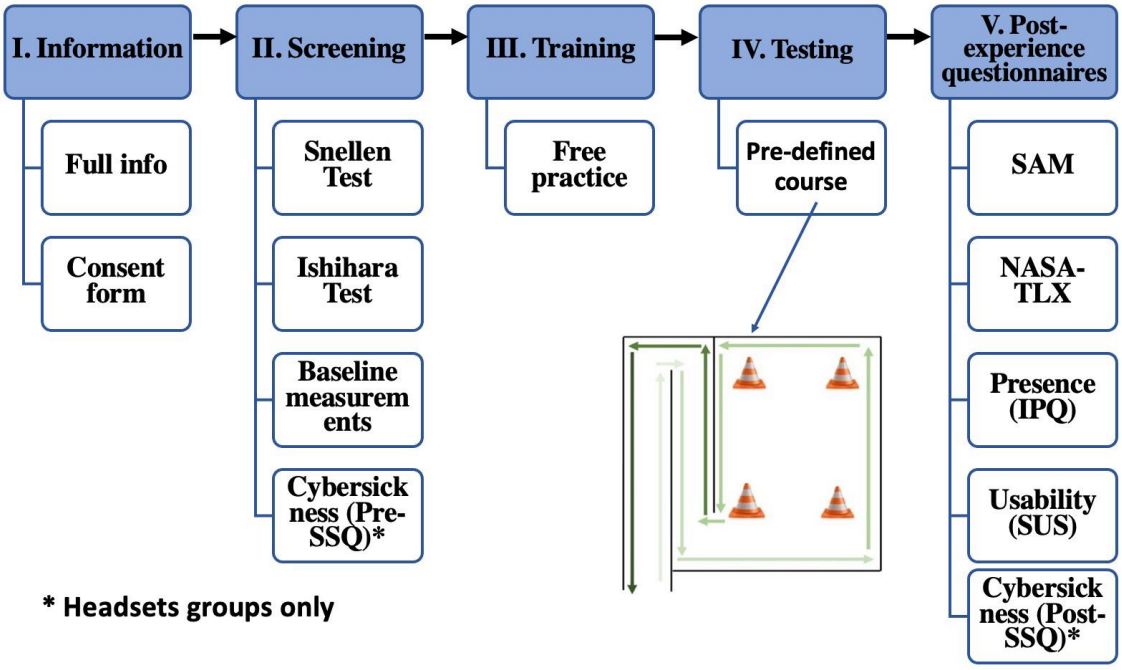


Figure 4.3: Lab-based studies assessment protocol.

4.1.4 Study 1: Assessment Protocol

The assessment protocol adhered to the three-phase structure described in Chapter 5 (Pre-test, test, and Post-test), and was operationalized in this study through five structured steps to ensure consistency as illustrated in Figure 4.3:

1. **Information Phase** – Participants received a detailed briefing (in person or virtually) along with an information sheet and signed informed consent prior to participation.
2. **Screening Phase** – Participants were screened for exclusion criteria, including recent alcohol consumption, insufficient sleep (<6 hours), or suspected pregnancy. Baseline physiological data were then recorded over a five-minute period using wearable sensors.
3. **Training Phase** – Participants engaged in a joystick-controlled familiarization session within a virtual, obstacle-free environment to practice navigation. This phase lasted approximately five minutes and allowed users to adjust to the interface and motion control.
4. **Testing Phase** – Participants completed the ramp navigation task (Figure 4.2), during which joystick activity, physiological signals, and simulator metrics were continuously recorded.
5. **Post-Experience Phase** – Participants completed the System Usability Scale (SUS), Igroup Presence Questionnaire (IPQ), Self-Assessment Manikin (SAM), and NASA Task Load Index (NASA-TLX). For participants in the immersive condition, the Simulator Sickness Questionnaire (SSQ) was also administered.

4.1.5 Assessment Tools

4.1.6 Usability (SUS) and Presence (IPQ)

A shortened version of the System Usability Scale (SUS) and the Igroup Presence Questionnaire (IPQ) was used in the lab-based studies. Items 1 to 5 assess usability, while items 6 to 10 assess presence. Participants rated each item using the Absolute Category Rating (ACR) method described in (ITU-T P.913, 2018), which is uses a five-point Likert-type scale to determine if a user agreed or disagreed with the statements. The full list of items and their corresponding dimensions presented in Table 4.1.

Table 4.1: SUS and IQP items with associated dimensions.

	Questions	Instrument	Dimension Assessed
1	I found the system unnecessarily complex.	SUS	System complexity
2	I thought the system was easy to use.	SUS	Ease of use
3	I would imagine that most people would learn to use the system very quickly.	SUS	Learnability
4	I found the system very cumbersome to use.	SUS	Navigation / intuitiveness
5	I need to learn a lot of things before I could get going with this system	SUS	Ease of learning
6	In the computer-generated world, I had a sense of "being there".	IPQ	General presence (PRES)
7	I had a sense of acting in the virtual space, rather than operating something free outside.	IPQ	Acting in virtual environment (SP)
8	I felt present in the virtual space.	IPQ	Sense of being present in VE (SP)
9	How aware were you of the real-world surroundings while navigating in the virtual world?	IPQ	Involvement (INV)
10	How much did your experience in the virtual environment seem consistent with your real-world experience?	IPQ	Realism (REAL)

4.1.7 Emotion (SAM)

The Self-Assessment Manikin (SAM) is a non-verbal, pictorial scale used to assess affective responses across three dimensions: valence (pleasure), arousal, and dominance (control-submission). Participants were briefed on these dimensions prior to use. A 9-point version of the SAM was administered at the end of the session to evaluate participants' emotional state (Bradley & Lang, 1994).

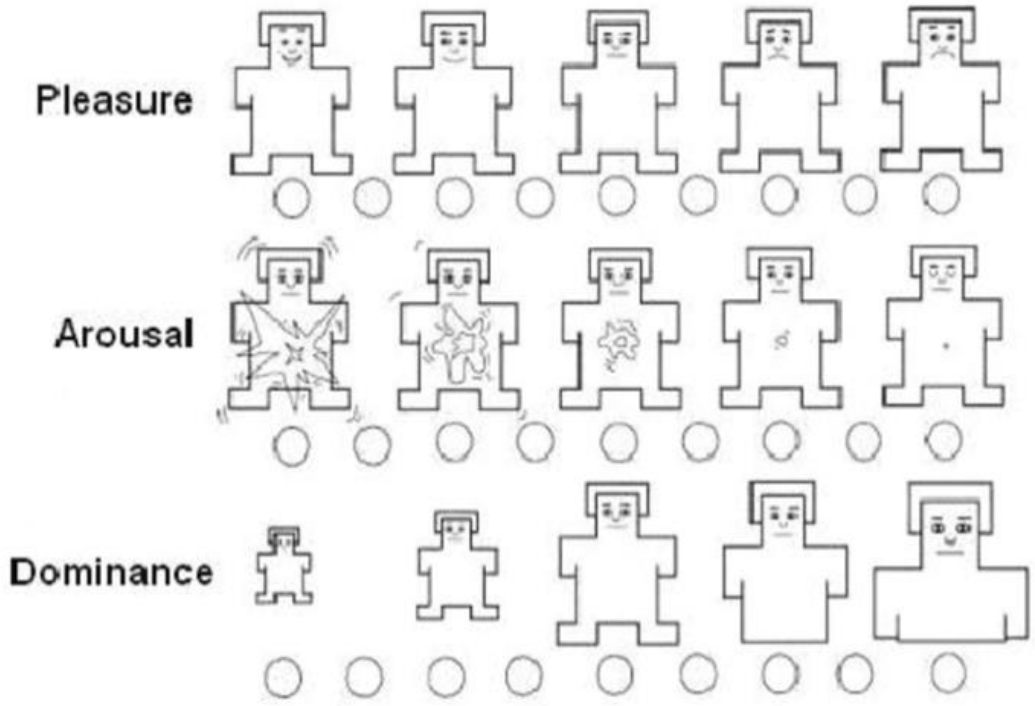


Figure 4.4: 9-point SAM scale.

4.1.8 NASA-TLX

The NASA Task Load Index (NASA-TLX) is a widely used subjective workload assessment comprising six dimensions: mental demand, physical demand, temporal demand, performance, effort, and frustration. Participants rate each dimension on a 0–100 scale using 5-point increments. Two types of scores are computed: the raw TLX score, as the unweighted average of all six ratings, and the weighted TLX score, which is derived from a pairwise comparison task. In this task, participants compare the six dimensions in all possible pairs (15 pairs total) and select the dimension that is more significant to their workload experience. The number of times each dimension is selected is used to generate a unique weighting for that participant. These individual weights are then applied to the participant's original dimension ratings to calculate the final weighted score (Hart, 2006; Hart & Staveland, 1988).

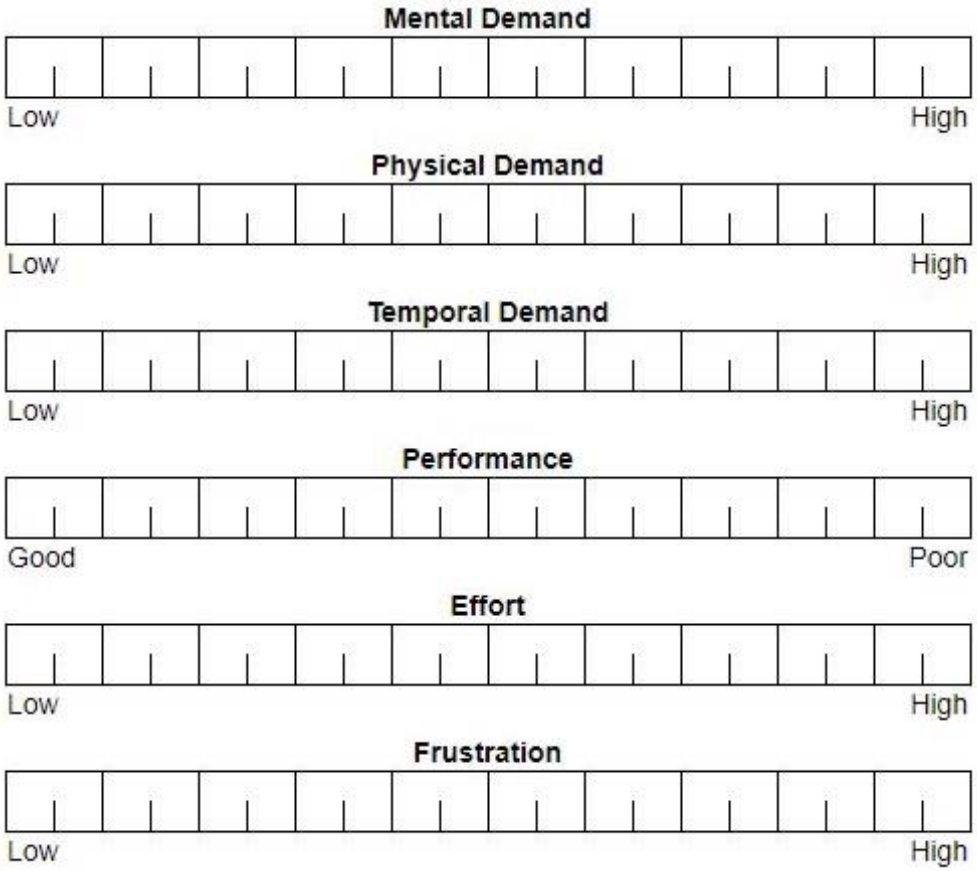


Figure 4.5: NASA-TLX rating scale.

4.1.9 SSQ

The Simulator Sickness Questionnaire (SSQ) was administered to participants after exposure to the virtual reality environment. The SSQ is a standardized tool designed to assess symptoms of motion sickness, and its item groupings are based on established guidelines. The questionnaire comprises 16 items, which are organized into four symptom categories: Nausea (N), Oculomotor (O), Disorientation (D), and General discomfort. Each symptom is rated on a 4-point Likert scale, ranging from 0 (None) to 3 (Severe).According to the scoring guidelines developed by Kennedy et al. (1993), the subscale and total scores are calculated using a specific weighting system, with only the Nausea, Oculomotor, and Disorientation subscales contributing to the final score. The scores for each subscale are derived from the sum of the ratings for the items in that category, and then multiplied by a specific constant. The final Total Score (TS) is the sum of the weighted Nausea, Oculomotor, and Disorientation scores. The SSQ items and their corresponding symptom categories are presented in Table 4.2.

Table 4.2: SSQ items groups by symptoms category.

SSQ Symptom	Nausea (N)	Oculomotor (O)	Disorientation (D)
General discomfort	1	1	
Fatigue		1	
Headache		1	
Eyestrain		1	
Difficulty focusing		1	1
Increased salivation	1		
Sweating	1		
Nausea	1		1
Difficulty concentrating	1	1	
Fullness of head			1
Blurred vision		1	1
Dizzy (eyes open)			1
Dizzy (eyes closed)			1
Vertigo			1
Stomach awareness	1		
Burping	1		
Total[1]	[1]	[2]	[3]
TS = ([1]+[2]+[3])*3.74	N =[1]*9.54	O=[2]*7.58	D=[3]*13.92
Total is the sum obtained by adding the symptoms scores. Omitted scores are zero. TS. Total score			

4.1.10 Study 1: Results

The findings of Study 1 are presented across three dimensions: performance metrics, explicit metrics and implicit metrics. This segmentation reflects the multidimensional assessment model adopted in the study, which aimed to capture user experience holistically during the wheelchair simulator task. As display type (desktop vs. immersive headset) was the independent variable, all comparisons presented here focus on differences between these two groups.

4.1.11 Demographics

A total of 41 participants were included in this analysis, comprising Group 1 (Desktop; N = 24) and Group 2 (Headset-1; N = 17). The mean age was 26.17 years (SD = 5.05) for the Desktop group and 30.29 years (SD = 7.33) for the Headset-1 group. The Desktop group included an equal number of male and female participants (12 males, 12 females), while the Headset-1 group included slightly more males than females (10 males, 7 females). Participants were recruited through convenience sampling and selected based on eligibility criteria, as outlined in Section 3.8.2 Pre-Test Methods.

4.1.12 Normality Test Analysis

The Shapiro-Wilk test was conducted to assess the normality of each variable for the Desktop and Headset-1 groups $\alpha = 0.05$). Variables with p-values > 0.05 were considered to follow a normal distribution; those with $p \leq 0.05$ were considered non-normal and analysed with non-parametric methods. Normality test results can be found at Table 7.3 Appendix E summarizes the results by indicating whether the assumption of normality was met for each variable.

4.1.13 Simulator-Based Performance Results

Simulator performance was compared between the Desktop and Headset-1 groups using the Mann–Whitney U test, due to non-normal distribution of the data. As shown in Table 4.3, no statistically significant differences were found across the three-performance metrics: task completion time, total joystick commands, and number of collisions.

Although the Headset-1 group showed slightly lower means across all three metrics, these differences were accompanied by small effect sizes ($r < 0.3$) and non-significant p-values, indicating that the observed variations may be attributable to individual variability rather than consistent performance differences between display conditions.

Table 4.3: Study 1 descriptive statistics and Mann–Whitney U test results for performance metrics

Metric	Group	Mean (SD)	Median [IQR]	U	Z	p-value	Effect size (r)
Task completion time (s)	Desktop	186.71 (45.25)	176.00 [162.00 – 203.00]	174.0	-0.794	0.427	0.12
	Headset-1	179.18 (38.41)	169.00 [150.25 – 199.25]				
Total commands	Desktop	78.21 (21.60)	73.50 [67.00 – 82.50]	141.5	-1.655	0.098	0.26
	Headset-1	69.12 (16.77)	67.00 [57.00 – 74.00]				
Total collisions	Desktop	5.96 (3.83)	5.50 [3.00 – 8.00]	165.5	-1.025	0.305	0.16
	Headset-1	4.71 (3.64)	3.00 [2.00 – 7.25]				

4.1.14 Explicit metrics - Subjective Responses

This section presents user-reported outcomes from standardized questionnaires, including SUS, IPQ, NASA-TLX, and SAM. Results compare perceived usability, workload, presence, and emotional responses between the Desktop and Headset-1 conditions using appropriate statistical tests and effect size estimates. SSQ is addressed in the following study.

4.1.14.1 Usability (SUS)

The System Usability Scale (SUS) item scores were compared between the Desktop and Headset-1 conditions using the Mann–Whitney U test. Descriptive statistics and test outcomes are presented in Table 4.4. Median scores across all items and the composite score were generally high in both groups, indicating good perceived usability (see Figure 4.6).

A statistically significant difference was found for Item Q2 (“Easy to use”), with higher ratings in the Headset-1 group ($p = 0.0098$), which remained significant after Bonferroni correction ($\alpha = 0.01$). The effect size, calculated as $r=-0.403$, indicates a moderate effect, suggesting a meaningful difference in the perceived ease of use, with the VR headset being rated as more intuitive than the desktop setup.

In contrast, all other individual SUS items, as well as the total SUS score, showed no statistically significant differences between the conditions. The effect sizes for these comparisons were uniformly low ($|r|<0.21$). Furthermore, the composite SUS score passed normality checks, allowing for a parametric comparison (see Table 4.5). An independent samples t-test confirmed the overall non-significant result ($t(39)=-1.07, p=0.293, d=-0.16$), reinforcing that overall usability perceptions were not meaningfully different between the two conditions.

In summary, while overall usability perceptions were high across both conditions and the composite SUS scores were not significantly different, the specific item concerning the ease of use was rated

significantly higher for the Headset-1 condition. This single, significant finding suggests that while both interfaces were considered usable, the VR headset was perceived as more intuitive or effortless to operate than the desktop counterpart.

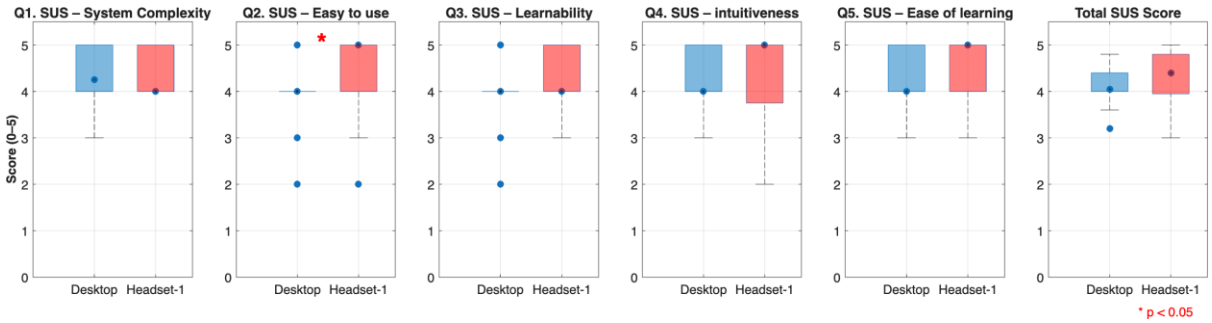


Figure 4.6: Study 1 boxplots SUS item scores by Group. Desktop(blue) and Headset-1(red).

Table 4.4: Study 1 descriptive statistics and Mann–Whitney U test results for SUS items

Metric	Group	Mean (SD)	Median [IQR]	U	Z	p-value	Effect Size (r)
Q1. SUS – System Complexity	Desktop	4.40 (0.64)	4.25 [4.00 – 5.00]	188	-0.47	0.635	0.073
	Headset-1	4.35 (0.49)	4.00 [4.00 – 5.00]				
Q2. SUS – Easy to use	Desktop	3.83 (0.64)	4.00 [4.00 – 4.00]	116.5	-2.58	0.0098**	0.403
	Headset-1	4.35 (0.86)	5.00 [4.00 – 5.00]				
Q3. SUS – Learnability	Desktop	3.96 (0.75)	4.00 [4.00 – 4.00]	166	-1.14	0.255	0.178
	Headset-1	4.24 (0.75)	4.00 [4.00 – 5.00]				
Q4. SUS – Navigation/ intuitiveness	Desktop	4.25 (0.61)	4.00 [4.00 – 5.00]	189	-0.43	0.665	0.067
	Headset-1	4.24 (0.97)	5.00 [3.75 – 5.00]				
Q5. SUS – Ease of learning	Desktop	4.29 (0.62)	4.00 [4.00 – 5.00]	187	-0.49	0.621	0.077
	Headset-1	4.35 (0.79)	5.00 [4.00 – 5.00]				
Total SUS Score	Desktop	4.15 (0.37)	4.05 [4.00 – 4.40]	154	-1.34	0.181	0.209
	Headset-1	4.31 (0.59)	4.40 [3.95 – 4.80]				
* p < 0.05. ** p < 0.01 after Bonferroni correction (α = 0.01)							

Table 4.5: Study 1 independent Samples t-test for total SUS and IPQ scores

Metric	Desktop M (SD)	Headset-1 M (SD)	t(df)	p-value	Levene's F (p)	Mean Diff.	95% CI
SUS Score	4.15 (0.37)	4.31 (0.59)	-1.07 (39)	0.293	5.80 (0.021)	-0.16	[-0.46, 0.14]
IPQ Score	3.20 (0.68)	3.84 (0.59)	-3.11 (39)	0.003*	0.03 (0.868)	-0.64	[-1.05, -0.22]

4.1.14.2 Presence (IPQ)

Table 4.6 presents descriptive statistics and the results of the Mann–Whitney U test for each IPQ item and the composite score. Participants using the Headset-1 condition consistently reported higher scores across all IPQ items compared to those in the Desktop group. These findings are visually summarized in the radar plot and boxplots (Figure 4.7 and Figure 4.8), clearly illustrating the enhanced sense of presence afforded by the VR headset.

The analysis revealed statistically significant differences for several key aspects of presence. Most notably, the VR headset significantly improved both General Presence (Q6) and the user's ability to Act in the Virtual Environment (Q7). The significant results for these items persisted even after applying a stringent Bonferroni correction for multiple comparisons ($p<0.01$ for both), and their moderate-to-large effect sizes ($r=-0.55$ and $r=-0.41$) indicate a meaningful and substantial difference in user experience. These results mean that users not only felt a much stronger overall sense of "being there" but also felt more capable of interacting and moving within the virtual space when using the VR headset.

The analysis of the composite IPQ score further confirmed these findings. As seen in Table 4.5, an independent samples t-test showed a statistically significant difference between conditions ($t(39)=-3.11, p=0.003$), with the Headset-1 group reporting a higher overall sense of presence. This supports the item-level analysis, confirming that the VR headset provides a significantly more immersive experience.

Interestingly, while the General and Spatial Presence dimensions were significantly enhanced, no significant differences were found for Involvement (Q9) or Realism (Q10). This suggests that while a VR display is highly effective at creating the fundamental feeling of "being there" and enabling spatial interaction, other factors like the virtual environment's interactivity, content, and the task itself may play a more dominant role in fostering a deeper sense of immersion and realism.

These findings align with prior literature, which also observed a strong connection between display modality and presence, For instance, research by (Hernandez-Ossa et al., 2017) reported high general and spatial presence levels in SimCadRom simulator. Similarly, (Zorzi et al., 2023a) observed strong general ($M = 4.2$) and spatial presence ($M = 4.1$), while involvement and realism were lower, attributed to limited interactivity and gamification. This mirrors the current study's pattern, where involvement and realism (Q9 and Q10) did not differ significantly across groups. (Vailland et al., 2020, 2021) also noted lower involvement scores in an immersive simulator setting, potentially due to external distractions and limited environmental immersion.

Together, the results indicate that immersive VR displays improve users' perception of presence, especially the general sense of "being there" and spatial interaction, but further enhancements to the virtual environment might be needed to strengthen involvement and realism dimensions of user experience.

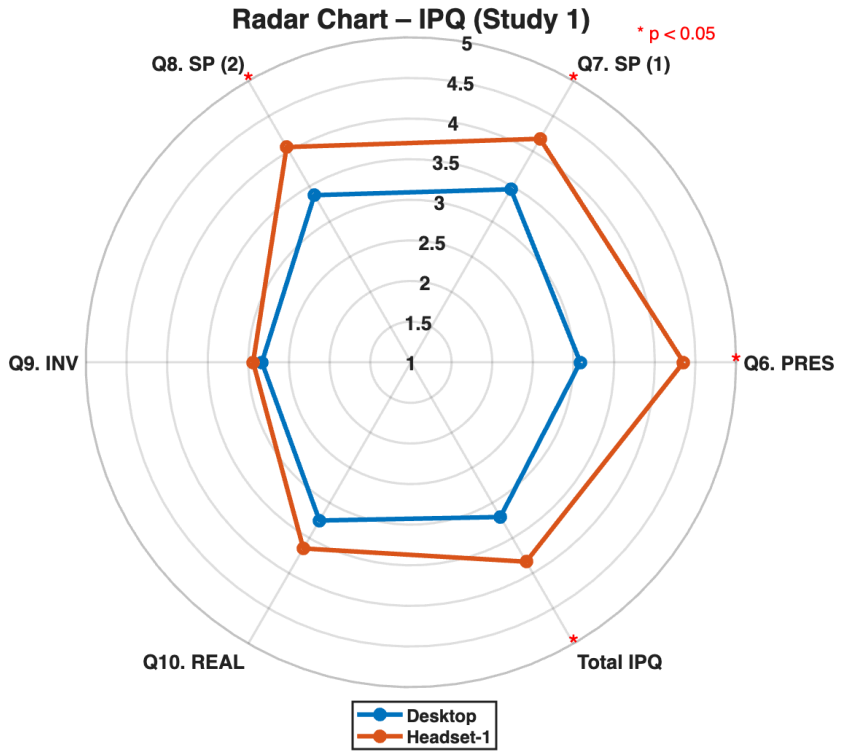


Figure 4.7: Study 1 radar plot for IPQ items and composite.

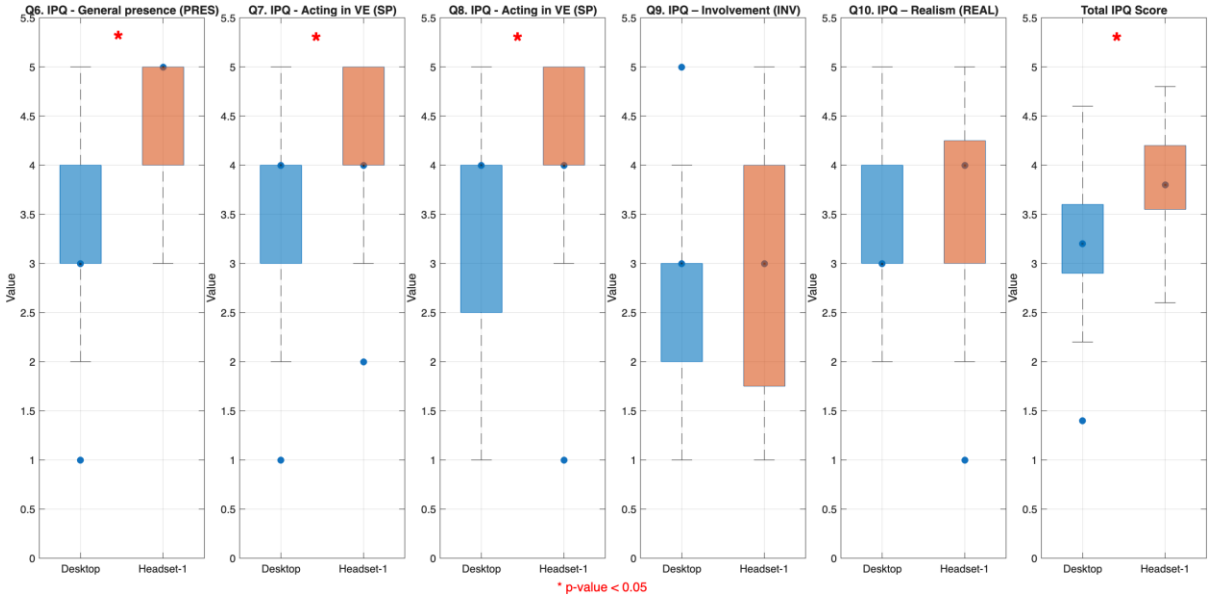


Figure 4.8: Study 1 boxplots of IPQ items.

Table 4.6: Study 1 descriptive statistics and Mann–Whitney U test results for IPQ item scores

Metric	Group	Mean (SD)	Median [IQR]	U	Z	p-value	Effect Size (r)
Q6. IPQ – General presence (PRES)	Desktop	3.08 (1.14)	3.00 [3.00–4.00]	77.5	-3.5	0.000**	0.55
	Headset-1	4.35 (0.79)	5.00 [4.00–5.00]				
Q7. IPQ – Acting in VE (SP1)	Desktop	3.46 (0.93)	4.00 [3.00–4.00]	112.5	-2.6	0.009**	0.41
	Headset-1	4.18 (0.81)	4.00 [4.00–5.00]				
Q8. IPQ – Sense in being in VE (SP2)	Desktop	3.38 (1.17)	4.00 [2.50–4.00]	132	-2.02	0.043*	0.32
	Headset-1	4.06 (1.03)	4.00 [4.00–5.00]				
Q9. IPQ – Involvement (INV)	Desktop	2.83 (1.09)	3.00 [2.00–3.00]	188.5	-0.42	0.671	0.07
	Headset-1	2.94 (1.39)	3.00 [1.75–4.00]				
Q10. IPQ– Realism (REAL)	Desktop	3.25 (0.68)	3.00 [3.00–4.00]	145.5	-1.67	0.095	0.26
	Headset-1	3.65 (1.11)	4.00 [3.00–4.25]				
Total IPQ Score	Desktop	3.20 (0.68)	3.20 [2.90–3.60]	94.5	-2.91	0.004**	0.45
	Headset-1	3.84 (0.59)	3.80 [3.55–4.20]				
* $p < 0.05$. ** $p < 0.01$ after Bonferroni correction ($\alpha = 0.01$)							

4.1.14.3 Emotional Response

Emotional responses were measured using the Self-Assessment Manikin (SAM) across three dimensions: valence (pleasure), arousal, and dominance(control) (VAD). As shown in Table 4.7 and Figure 4.9Figure 4.9: Study1 three-dimensional Valence–Arousal–Dominance (VAD) plot., no statistically significant differences were found between the Desktop and Headset-1 groups across these domains.

However, a notable trend emerged in the dominance ratings, where the difference between groups approached statistical significance (p=0.081). The observed effect size (r=0.27) suggests a minor to moderate trend, with participants in the Headset-1 condition reporting a slightly higher sense of control or influence within the virtual environment..

While these subjective findings did not reach the threshold for statistical significance on their own, they provide valuable context for a multimodal analysis (Magalhães et al., 2024). For example, similar studies have used SAM alongside physiological data, such as electrodermal activity (EDA) and heart rate (HR), to identify specific emotional states like negative arousal (Liao et al., 2020). Although the correlation between subjective SAM scores and physiological measures was not analysed in this specific study, Study 2 will extend this investigation. By combining the SAM scores with the concurrently collected physiological signals, Study 2 aims to provide a more comprehensive understanding of the emotional and affective responses to the immersive wheelchair simulation.

Table 4.7: Study 1 descriptive statistics and Mann–Whitney U test results for SAM item scores

Metric	Group	Mean (SD)	Median [IQR]	U	Z	p-value	Effect size(r)
Valence	Desktop	6.67 (1.69)	7 [5.00 - 8.00]	189.00	-0.42	0.676	0.07
	Headset-1	6.71 (1.76)	7 [7.00 - 7.00]				
Arousal	Desktop	6.67 (1.61)	7 [6.50 - 7.00]	200.00	-0.11	0.909	0.02
	Headset-1	6.82 (1.24)	7 [6.00 - 7.25]				
Dominance	Desktop	6.38 (1.81)	7 [5.00 - 7.50]	140.50	-1.74	0.081	0.27
	Headset-1	7.29 (1.53)	7 [7.00 - 8.25]				

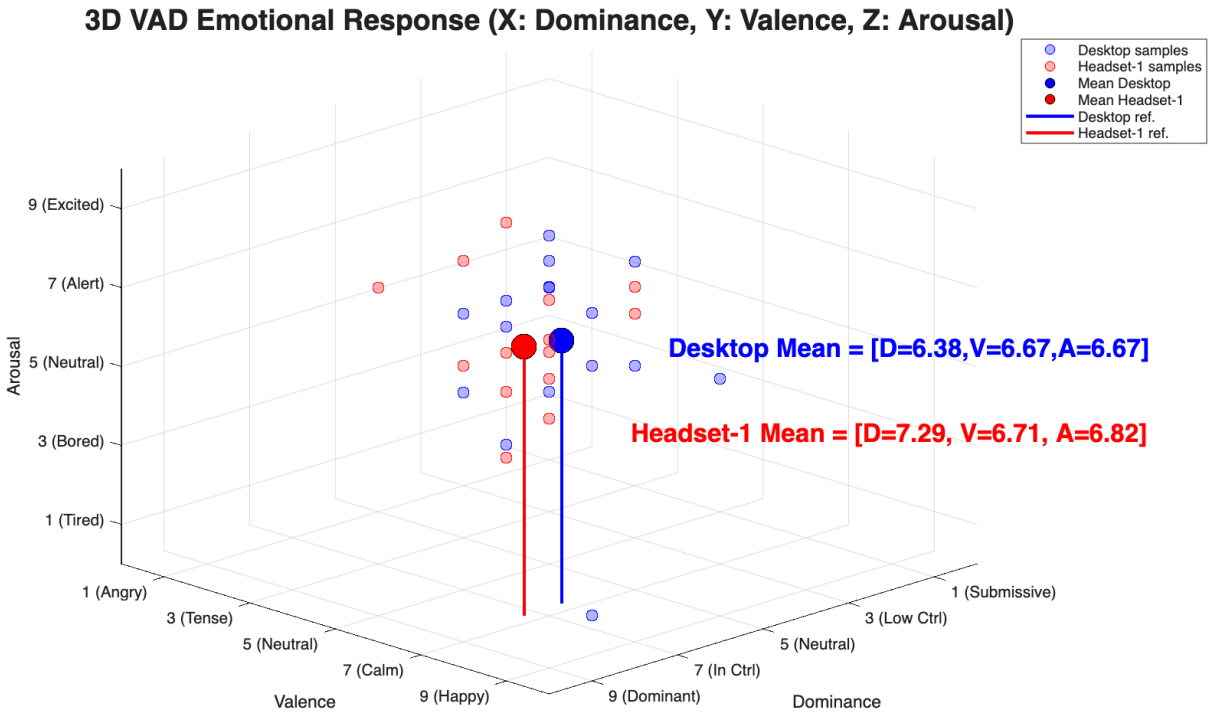


Figure 4.9: Study1 three-dimensional Valence–Arousal–Dominance (VAD) plot.

4.1.14.4 Cognitive Task Workload (mental load)

Cognitive workload was evaluated using the NASA-TLX subscales. Mann–Whitney U tests revealed no significant group differences in Mental, Physical, Temporal, Effort, or Frustration demands after Bonferroni correction (adjusted α = .0083), indicating that for these specific factors, both conditions were perceived as equally demanding. The key statistical comparisons are presented in Table 4.8, while the complete TLX item-level distributions are described in Table 7.4 and Table 7.5 at Appendix.

A key finding emerged in the Performance dimension, where the Headset-1 group reported significantly worse self-perceived performance compared to the Desktop group in both raw

(p<.001,r=0.70) and weighted subscales (p<.001,r=0.66). This means that although the participants were able to successfully complete the tasks in both conditions, those in the VR environment felt they were performing less effectively. This discrepancy suggests that the immersive environment imposed additional challenges, causing users to feel less successful in their task execution, despite the objective task completion.

These patterns are visually depicted in the radar plots and boxplots (Figure 4.10, Figure 4.11 and Figure 4.12), which show the largest gap in the Performance dimension. The composite NASA-TLX scores further reinforced this trend: participants in the Headset-1 condition reported a significantly higher overall workload than Desktop users in both raw (p=.007) and weighted (p=.001) scores. This finding indicates a broader cognitive burden in the immersive setup, which is not confined to a single subscale but reflects a cumulative effect across the task.

This increase in perceived cognitive workload is a well-documented phenomenon in VR wheelchair simulator research. The presented findings align with previous studies, such as (Kamaraj, Dicianno, et al., 2016b), who found that users of a virtual wheelchair simulator experienced a higher workload than those in real-world scenarios. Similarly, (Fraudet et al., 2024) reported elevated NASA-TLX scores in VR across all difficulty levels. These consistent findings suggest that while VR can provide enhanced realism for training, it often introduces an additional cognitive processing burden. This extra mental effort may stem from navigating a less-familiar interface, processing sensory information from an unfamiliar display, or managing the physical discomfort that can sometimes arise from immersive technology.

In contrast, (L. Devigne et al., 2017) observed relatively low cognitive workload during simulator use, which they attributed to controlled task demands and prior familiarization. Moreover, (Vailland et al., 2020) demonstrated that vestibular feedback could mitigate VR-induced frustration and stress, suggesting that the absence of relevant sensory cues may have contributed to the increased perceived workload. This leads to a critical question for designers: how can we use HMDs to improve training while mitigating this cognitive burden?

In summary, the presented results support the conclusion that immersive VR systems, while valuable for training, require careful attention to their design to avoid overwhelming users. This study did not find any significant differences in Mental, Physical, Temporal, Effort, or Frustration demands between the two conditions, but the significantly lower perceived performance and higher overall workload in the immersive environment are key findings that must be addressed. It's plausible that physical discomfort or disorientation, often referred to as cybersickness, could be a contributing factor, as

these symptoms have been shown to impair concentration and elevate cognitive demands (Sepich et al., 2022).

Although correlations between cybersickness and NASA-TLX were not tested in this phase, the analysis of cybersickness data in Study 2 will explicitly examine this relationship to clarify how physical discomfort interacts with cognitive workload in VR-based assessment contexts. Future research should explicitly examine this relationship to clarify how physical discomfort interacts with cognitive workload in VR-based assessment contexts.

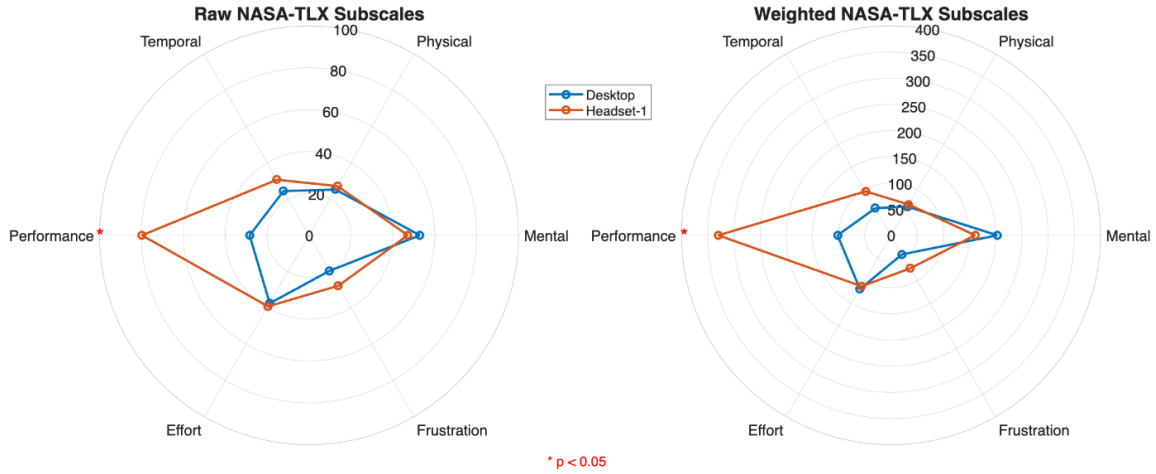


Figure 4.10: Study 1 radar plots of mean NASA-TLX subscale scores for Desktop and Headset-1 conditions.

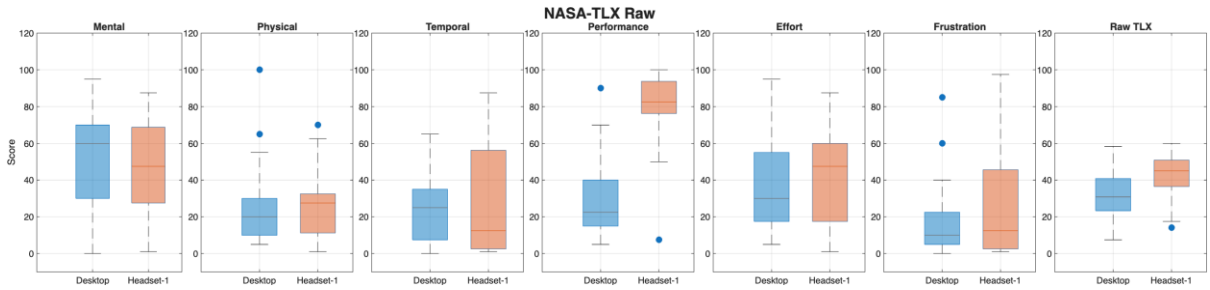


Figure 4.11: Study 1 boxplots of NASA-TLX items (raw).

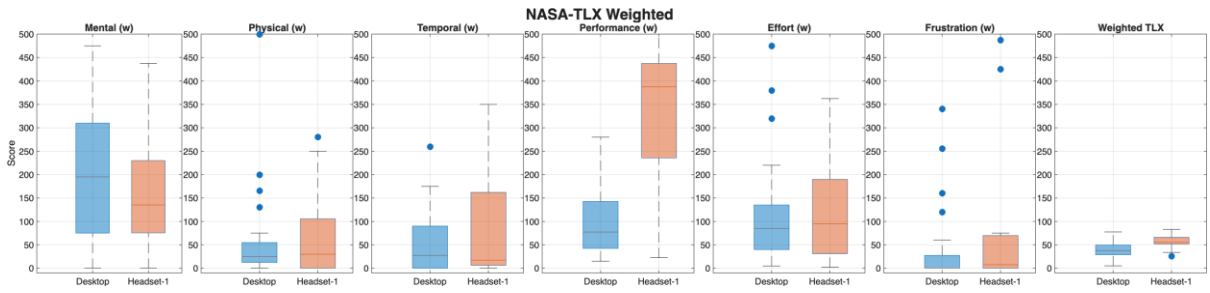


Figure 4.12: Study 1 boxplots of NASA-TLX items (weighted).

Table 4.8: Study 1 summary of test results for NASA-TLX item scores.

Metric	Group	Mean (SD)	Median [IQR]	U	Z	p-value	Effect size (r)
Performance	Desktop	28.33 (20.41)	22.5 [15.00- 40.00]	34.00	-4.51	0.000*	0.70
	Headset-1	79.71 (22.34)	82.5 [76.25- 93.75]				
Performance weighted	Desktop	101.88 (80.47)	77.5 [42.50- 142.50]	44.50	-4.22	0.000*	0.66
	Headset-1	329.56 (144.17)	387.5 [235.62- 437.50]				
Raw TLX	Desktop	31.32 (12.13)	30.835 [23.33- 40.83]	102.50	-2.69	0.007**	0.42
	Headset-1	41.94 (12.72)	45.0 [36.46- 50.94]				
Weighted TLX	Desktop	39.24 (16.88)	37.5 [29.17- 49.50]	82.50	-3.22	0.001*	0.50
	Headset-1	56.03 (13.39)	55.5 [51.71- 65.92]				
* $p < 0.05$. ** $p < 0.0083$ after Bonferroni correction ($\alpha = .0083$)							

4.1.15 Implicit Metrics - Physiological Responses

Physiological signals were analysed to assess implicit indicators of cognitive workload and emotional arousal during simulator use. Electrodermal activity (EDA), heart rate (HR), and inter-beat interval (IBI) were continuously recorded and segmented into: I. baseline, II. task execution, and III. first collision phases. Derived metrics included baseline-corrected values, percentage changes, and heart rate variability (HRV) indices (SDNN and RMSSD) in the time domain. Shapiro–Wilk tests indicated that many physiological variables, particularly EDA-derived and change-based metrics, violated normality assumptions. Consequently, group comparisons were conducted using the non-parametric Mann–Whitney U test.

Table 4.9 summarises the significant group differences observed in HR and IBI metrics. Participants in the Headset-1 group exhibited significantly higher HR during the task phase ($p = .013$) and the first collision phase ($p = .037$) compared to those in the Desktop group. Baseline relative changes measures also showed significant increases in the immersive condition: both the absolute HR difference from baseline to task ($p = .019$) and the corresponding percentage change ($p = .016$) were greater in the Headset-1 group. Although the HR change from baseline to first collision did not reach significance ($p = .121$, $r=.25$), it showed a similar upward trend. These group differences were associated with effect sizes ranging from $r = 0.33$ to 0.39, indicating meaningful physiological differences despite the modest sample size.

Heart rate variability analyses further supported these findings. While IBI values during the task phase did not significantly differ between groups, a significant difference was observed in the change from baseline to first collision ($p = .005$, $r =.43$), with the Headset-1 group exhibiting a more pronounced

reduction in IBI, suggestive of increased autonomic arousal under immersive conditions. This contrast was accompanied by a medium effect size (r = 0.43).

Time-domain HRV indices showed similar trends. The change in SDNN from baseline to task nearly reached statistical significance (p = .065, r = 0.29). In contrast, RMSSD values remained comparable across conditions. The complete list of results is provided in Table 7.10 at Appendix.

EDA metrics based on absolute values at baseline, task, and first collision phases did not yield statistically significant differences. However, change-based metrics revealed a trend toward greater sympathetic activation in the Headset-1 group. In particular, the percentage change in EDA from baseline to the first collision phase (p = .120, r=.25), suggesting a modest increase in physiological arousal under immersive condition. While these findings do not reach statistical significance, they could indicate that relative EDA changes may be more sensitive to context-dependent autonomic fluctuations than absolute values.

These findings are further supported by the findings of (Zorzi et al., 2023b, 2023a), who examined heart rate responses during powered wheelchair training in a virtual reality environment using a chest-worn Polar H10 sensor. Their study reported the highest heart rate values during the backward slalom task, which was attributed to increased complexity and user engagement. Participants were observed performing behaviours consistent with real-world navigation, such as looking backward to avoid obstacles, suggesting greater involvement in the immersive condition.

Although their study did not include subjective assessments of workload, such as the NASA-TLX, it incorporated user's quality of experience measures through the IPQ and the SSQ, focusing on presence and cybersickness. These dimensions are further addressed in Study 2 of the presented thesis, where subjective assessments of workload, presence, and tolerance are explicitly integrated with physiological and performance-based measures.

In summary, the immersive condition elicited stronger autonomic responses, particularly through elevated HR and reduced IBI, with intermediate effect sizes across significant metrics. While EDA differences did not reach statistical significance, their trends aligned with heightened sympathetic activation. These results support the hypothesis that immersive environments can evoke stronger physiological responses compared to non-immersive, desktop-based interaction.

Table 4.9: Study 1 summary of test results comparing HR, EDA, and HRV (IBI).

Metric	Group	Mean (SD)	Median [IQR]	U	Z	p-value	Effect size (r)
Mean HR - Test	Desktop	74.43 (9.30)	74.81 [68.19 – 81.21]	105.0	-2.48	0.013*	0.39
	Headset-1	83.75 (13.41)	86.27 [73.62 – 91.18]				
Mean HR - 1st Collision	Desktop	76.84 (10.21)	77.79 [69.07 – 82.56]	111.0	-2.08	0.037*	0.33
	Headset-1	87.43 (16.76)	82.03 [79.53 – 91.60]				
Mean HR - Difference (Test - Baseline)	Desktop	-0.71 (5.36)	-1.47 [-3.52 – 3.17]	110.0	-2.34	0.019*	0.37
	Headset-1	5.53 (10.32)	3.50 [-1.36 – 7.72]				
Mean HR - Difference (1st Collision - Baseline)	Desktop	1.70 (7.11)	0.18 [-3.10 – 7.21]	129.0	-1.57	0.121	0.25
	Headset-1	8.14(15.40)	3.56[-1.62 – 12.80]				
Mean HR % Change (Test vs. Baseline)	Desktop	-0.75 (7.13)	-2.07 [-5.39 – 4.50]	108.0	-2.39	0.016*	0.37
	Headset-1	7.26 (13.19)	4.09 [-1.81 – 11.04]				
Mean IBI - Difference (1st Collision - Baseline)	Desktop	-0.06 (0.21)	-0.01 [-0.03 – 0.01]	69.0	-2.75	0.005**	0.43
	Headset-1	-0.50 (0.40)	-0.70 [-0.81 – -0.03]				
* $p < 0.05$. ** $p < 0.0125$ after Bonferroni correction ($\alpha = 0.05/4$)							

4.1.16 Study 1: Discussion and Contribution to SRQ1

Study 1 examined the influence of system configuration, desktop versus immersive VR (Headset-1), on user experience and physiological responses during a wheelchair simulator session. This first lab-based evaluation focused on four core dimensions of Quality of Experience (QoE): usability, emotional response, sense of presence/immersion, and cognitive workload. Additionally, physiological metrics (EDA, HR, IBI) were recorded to capture implicit indicators of workload and arousal.

Revisiting the Hypotheses

H1: Participants using the immersive simulator (Headset 1 group) will report higher QoE scores—covering usability (SUS), emotional response (SAM), sense of presence/immersion (IPQ), and workload (NASA-TLX)—than those using the non-immersive (Desktop group) version of Wheelchair Simulator.

H1 was partially supported. Participants in the immersive condition reported significantly higher presence scores (IPQ) and cognitive workload (NASA-TLX) scores, suggesting enhanced sensory engagement and mental demand. These findings are consistent with prior studies reporting higher perceived presence in immersive wheelchair simulators using HMDs compared to non-immersive setups (Zorzi et al., 2023a, 2023b); (Hernandez-Ossa et al., 2020). Higher workload composites and sub-scores in immersive VR also align with previous research showing increased mental effort and frustration during simulator-based mobility tasks (Kamaraj, Dicianno, et al., 2016b); (Fraudet et al., 2024).

However, no significant differences were observed in overall usability (SUS) or emotional responses (SAM). This aligns with findings (Arlati et al., 2020), who suggested that increased immersion may not always translate to higher usability, especially when display fidelity or motion cues are mismatched. Moreover, the lack of significant emotional differences may reflect the controlled and non-threatening nature of the tasks.

H2: Physiological responses (EDA, HR, IBI) will significantly differ between the immersive and non-immersive groups, indicating variation in arousal and engagement.

H2 received stronger empirical support. Heart rate and inter-beat interval showed significant group differences during both the task and first collision phases, indicating increased autonomic arousal in the Headset-1 condition. This supports previous findings by (Zorzi et al., 2023a, 2023b) , who reported elevated HR during complex wheelchair tasks in VR, especially those involving higher psychophysical load. Although EDA values did not significantly differ, change-based EDA metrics showed consistent trends of greater sympathetic activation in the immersive group. These findings reinforce the value of

including both absolute and baseline-relative physiological metrics in simulator evaluation, to better understand user responses in immersive contexts.

However, these physiological changes were not mirrored in participants' self-reported emotional states, as measured by the SAM, where no significant group differences were found for valence, arousal, or dominance. This suggests that the heightened autonomic arousal observed in the immersive condition may be related to factors other than an explicitly perceived emotional response, such as cognitive burden or discomfort.

4.1.17 Key findings of Study 1 include:

Methodological contribution

Study 1 demonstrated the feasibility of applying a multidimensional QoE evaluation framework to powered mobility simulation. By integrating validated subjective instruments (e.g., SUS, IPQ, NASA-TLX, SAM) with physiological data, the study captured complementary insights into users' emotional, cognitive, and sensory responses, advancing prior frameworks such as those proposed by (Arlati et al., 2020).

Physiological sensitivity to immersion

The study confirmed that physiological metrics, particularly HR and IBI, are sensitive to immersive system configurations, with significantly higher autonomic activation observed in the Headset-1 condition. This aligns with results reported in immersive mobility training by (Zorzi et al., 2023a, 2023b) , as well as general VR research (Liao et al., 2020; Magalhães et al., 2024). While EDA did not show significant differences, derived change scores pointed toward a pattern of increased sympathetic activity.

Immersion/Presence workload trade-offs

Participants in the immersive condition reported both higher presence and cognitive workload, highlighting the balance between realism and mental demand in VR systems. This trade-off was also noted by (Fraudet et al., 2024; Vailland et al., 2020, 2021), who emphasized that high presence could coincide with elevated cognitive and sensory load if the system lacks motion realism or induces discomfort.

These contributions align with Objective 1.1 and underpin Contribution 1 of this thesis: the development of a QoE-based framework that integrates subjective, behavioural (performance), and physiological metrics to inform the design and evaluation of simulator systems for powered mobility.

4.1.18 Limitations and Rationale for Study 2

While Study 1 demonstrated the feasibility and value of a multidimensional approach to evaluating simulator-based QoE, several limitations were identified that warrant further investigation in subsequent work.

Firstly, the comparison was limited to two system configurations: a desktop-based condition and a single immersive setup (Headset-1). Although this contrast revealed meaningful differences across multiple QoE dimensions, it did not allow for the isolation of specific immersive design features, such as motion intensity, display parameters, or sensory feedback, that may contribute to changes in workload, presence, or physiological arousal. As such, the impact of immersive fidelity and realism remains underexplored.

Second, although physiological markers revealed meaningful group differences, emotional responses (SAM) did not show significant variation between groups. SAM provided limited resolution for capturing subtle affective shifts during task performance. Furthermore, no correlation analysis was conducted between physiological and subjective emotional indicators. This was a deliberate scope limitation of Study 1, which focused on establishing a foundational framework and demonstrating group differences. A correlation analysis would have required a more complex methodological approach to synchronize and analyse two distinct data streams, which was deemed more appropriate for a follow-up investigation.

Third, and notably, cybersickness was not directly investigated. This was an oversight in the initial design of Study 1, as the primary focus was on establishing a baseline comparison of system configurations. However, prior research has shown that immersive environments, particularly those lacking vestibular feedback or motion congruence, can induce cybersickness symptoms (Vailland et al., 2020). These symptoms may elevate cognitive load and negatively impact user experience, confounding interpretations of workload, presence, and performance. Without a validated tool such as the Simulator Sickness Questionnaire (SSQ), it is difficult to determine whether physiological arousal observed in Study 1 was due solely to task complexity or also reflected a negative tolerance response to the immersive setup.

To address these limitations, Study 2 introduces a third experimental group (Headset-2) featuring a modified motion profile designed to investigate the role of visual-vestibular congruence in user experience. This study builds on the framework established in Study 1 and enhances it by examining how changes in simulator motion dynamics influence users' QoE in terms of perceived workload,

usability, sense of presence (SoP), emotional (arousal, valence and dominance) and tolerance (cybersickness) and physiological arousal.

In particular, the addition of the SSQ in Study 2 allows for the direct assessment of cybersickness and its potential interaction with cognitive workload and emotional experience. This integration will help clarify whether elevated physiological responses observed in immersive conditions reflect increased engagement, physical discomfort, or a combination of both.

In summary, Study 1 established a foundational methodological framework for simulator-based QoE assessment and identified key patterns in how immersive systems influence user experience. These insights directly informed the refinements implemented in Study 2, which aims to disentangle the effects of immersive fidelity and motion realism on user tolerance, workload, and psychophysiological engagement in wheelchair simulation.

Study 2 – Comparative QoE and Cybersickness Assessment

4.2 Study 2: Introduction

Building on the findings and limitations of Study 1, Study 2 expanded the evaluation scope by examining how motion profile and display modality affect Quality of Experience (QoE) and simulator-induced discomfort (cybersickness).

While Study 1 compared desktop and immersive conditions with the same motion (acceleration/deceleration) profile, Study 2 introduced a third configuration: an immersive headset using a low-jerk (smooth) motion profile. This specific condition was prompted by qualitative feedback from pilot tests, where users commented on the harsh, "too real" motion felt during collisions in the immersive high-jerk setup. This design enabled a controlled comparison of both motion dynamics and visual immersion across three configurations: I) Desktop with high-jerk motion, II) Headset-1 (immersive HMD) with high-jerk motion and III) Headset-2 (immersive HMD) with low-jerk motion.

Jerk is the derivative of acceleration and represents the rate at which acceleration changes. The human body naturally perceives both acceleration and jerk in the real world (Eager et al., 2016; Grant & Haycock, 2008). However, in VR systems, exposure to high jerk levels can exacerbate visual-vestibular conflict, the mismatch between visual motion cues and the lack of corresponding physical movement, potentially resulting in discomfort or cybersickness (Vlahovic et al., 2022).

The presented study objective was to investigate how motion smoothness in an immersive display affects usability, emotional response, cognitive workload, physiological arousal, and simulator sickness. This study directly supports Objectives 1.2 and 1.3, which focus on understanding the impact of immersive simulator design on user experience and tolerance (cybersickness).

To maintain consistency with Study 1 and strengthen the comparative analysis, Study 2 employed the same subjective measures (SUS, IPQ, SAM, and NASA-TLX) and physiological signals captured via the Empatica E4 wristband. Additionally, the Simulator Sickness Questionnaire (SSQ) was introduced to formally assess symptoms of cybersickness. The SSQ was administered only to participants in the immersive conditions (Headset-1 and Headset-2), as simulator sickness is typically associated with immersive VR exposure, and pilot testing confirmed that jerk differences were not perceptible in the non-immersive (Desktop) condition.

Head movements were also monitored to explore behavioural indicators of user response. For the Desktop condition, head movement was estimated using facial landmarks via OpenFace, while for the immersive headset groups, built-in head tracking sensors were used. These data were collected during

the simulation phase and processed through the framework described in the methodology section 3.3.2.

4.2.1 Study 2: Aim and Hypotheses

The aim of Study 2 was to examine how simulator configuration, defined by the combination of display modality (desktop vs. immersive) and motion profile (high-jerk vs. low-jerk), affects users' QoE and simulator-induced discomfort.

QoE was assessed across four experiential dimensions: usability, emotional response, sense of presence, and cognitive workload, using the same validated instruments as in Study 1. Simulator sickness was evaluated using the SSQ, and physiological signals (EDA, HR, IBI) were recorded to capture implicit indicators of arousal and cognitive effort.

This study builds upon Study 1 by adding a third configuration featuring a smoothed, low-jerk motion profile within the immersive condition. This allows for examination of how motion smoothness interacts with display modality in shaping users' physiological and experiential responses. The following hypotheses were tested:

- H1: Participants in the Headset 2 condition (immersive, low-jerk motion) will report higher QoE scores-covering usability (SUS), emotional response (SAM), sense of presence (IPQ), and lower cognitive workload (NASA-TLX) as well as lower simulator sickness (SSQ) compared to Headset 1 (immersive, high-jerk) condition.
- H2: Physiological responses (EDA, HR, IBI) will significantly differ across the simulator configurations, with expected differences between the Headset 2 condition and both the Desktop and Headset 1 groups. These variations will reflect how display modality and motion profile influence users' autonomic responses during simulator use.

4.2.2 Study 2: Material and Methods

Study 2 employed a between-groups experimental design to compare user responses across three simulator configurations that varied by motion profile and display modality. The three conditions were:

1. **Desktop Group**: A conventional 2D display setup using a trapezoidal (high jerk) motion profile.
2. **Headset 1 Group**: An immersive condition using a head-mounted display (HMD) with the same high jerk motion.

3. **Headset 2 Group**: An immersive condition using the HMD with a smoothed sinusoidal (low jerk) motion profile.

Each participant experienced only one configuration, allowing comparisons to be drawn between independent groups while minimizing carryover or adaptation effects.

4.2.2.1 Study 2 participants and sampling strategy

A total of 57 participants were included in the final analysis. Participants were recruited through convenience sampling and provided informed consent before participation. They were randomly assigned to one of three groups: 24 in the Desktop group, 17 in the Headset 1 group, and 16 in the Headset 2 group. A power analysis using *G*Power 3.1 indicated that this sample size is sufficient to detect large effect sizes (Cohen's $d > 0.80$) in pairwise comparisons between groups with 80% power at $\alpha = 0.05$. However, the study may be underpowered for detecting large effects in three-group omnibus comparisons. To support interpretation, effect size estimates are reported alongside p-values throughout the analysis.

4.2.2.2 Study 2 experimental procedure

The experimental protocol followed the same five-phase structure established in Study 1, including, baseline recording, a joystick training session, a standardized navigation task (see , and post-experience assessments. The virtual task environment and joystick interface remained unchanged, ensuring comparability. One notable addition was the inclusion of SSQ evaluation for participants in the HMD (Headset-1 and Headset-2) conditions. As in the previous study, the assessment protocol incorporated SUS, IPQ, SAM, and NASA-TLX. Additionally, physiological signals were recorded using the Empatica E4 wristband during the simulation phase.

Head movements were also recorded as part of an exploratory analysis of behavioural indicators. For the Desktop group, head motion was estimated using OpenFace facial landmark tracking, while the Headset groups used the built-in head tracking sensors of the Oculus Rift DK2. These data were processed through the framework described in Section 3.3.2 of the methodology.

Figure 4.13: Ramp route first-person camera view.

4.2.2.3 System and Interface Configuration

Simulator configurations remained consistent across groups in terms of virtual environment, joystick control, and task layout. All participants navigated at first-person view (see Figure 4.13) the same obstacle course using the same joystick interface. The primary variation across groups related to the motion behaviour of the virtual wheelchair, which influenced the acceleration and deceleration dynamics during simulation. These motion design differences are detailed in the following section.

4.2.2.4 Motion Design and Display Conditions

To investigate the effects of motion dynamics and visual immersion on user experience, Study 2 compared three simulator configurations that varied by motion profile and display modality. Each condition was carefully designed to isolate the influence of these factors on performance, cognitive workload, physiological arousal, and simulator tolerance.

4.2.2.4.1 Motion Profiles

The motion behaviour of the virtual wheelchair was defined by the shape of its acceleration and deceleration curves, representing different levels of “jerk”, the rate of change of acceleration over time (see Figure 4.14).

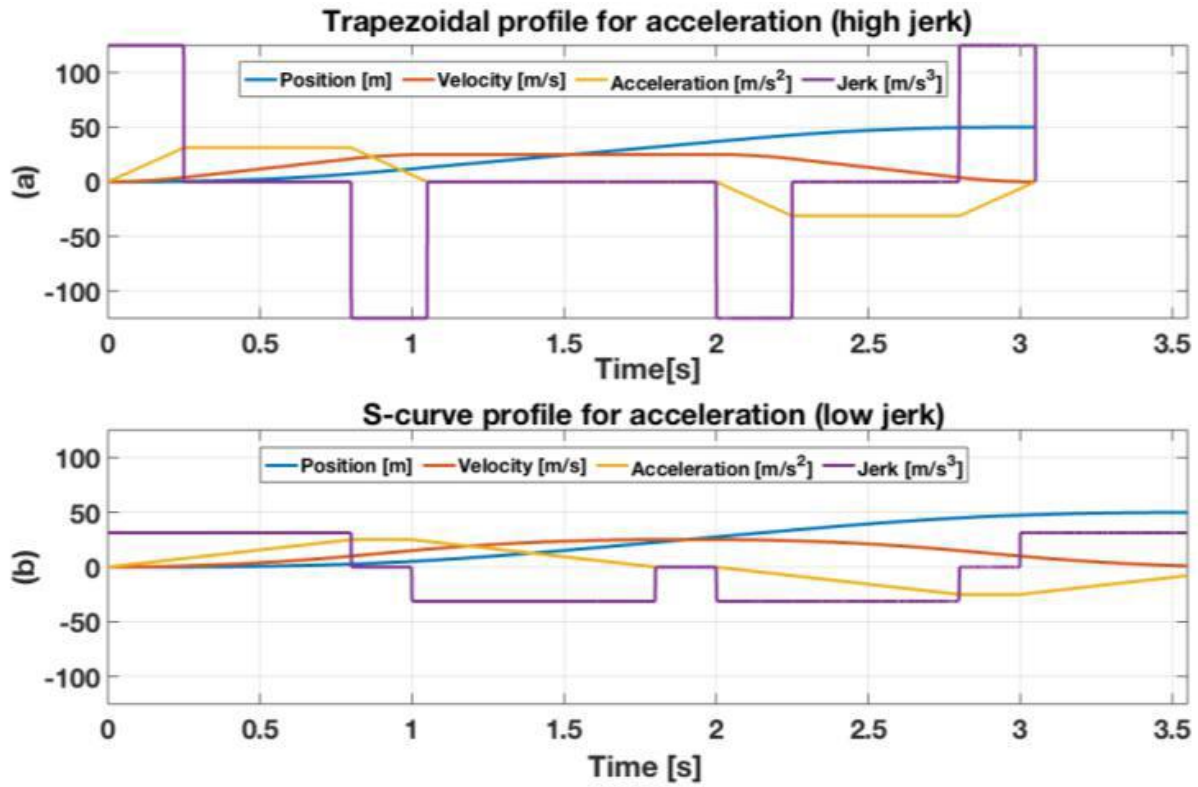


Figure 4.14: a) High jerk (acceleration trapezoidal) profile and (b) low jerk sinusoidal (s-curve) profile.

Two distinct motion profiles were implemented using Unity3D Engine's physics-based control system.

I. Trapezoidal Profile (High Jerk): Used in the Desktop and Headset 1 groups, this profile introduced abrupt transitions in velocity, simulating fast acceleration and deceleration without smoothing. These transitions correspond to high jerk levels and were intended to mimic the sudden changes users often experience in real-world powered wheelchair operation.

II. Sinusoidal (S-curve) Profile (Low Jerk): Used in the Headset 2 group, this profile applied smooth transitions in velocity using an s-curve pattern for acceleration and deceleration. This motion reduced the rate of change in acceleration, thereby lowering the visual-vestibular mismatch often associated with simulator sickness. It was expected to improve user comfort and physiological tolerance without compromising task realism.

The jerk profiles were calibrated based on prior pilot testing, and parameters such as peak speed and joystick sensitivity were kept constant across groups to ensure comparability. The only variable altered was the timing and slope of acceleration ramps. In proportional terms, the low-jerk condition was approximately 75% lower in acceleration/deceleration magnitude.

4.2.2.4.2 Display Conditions

Two types of visual output were used to manipulate the level of immersion:

I. Conventional Monitor (Desktop Group): Participants viewed the simulator on a 22-inch monitor (resolution: 1600×900). This setup offered a first-person camera view of the virtual environment and no head-tracking capability. Participants interacted using the joystick alone.

II. Head-Mounted Display (HMD Groups): Participants in the Headset 1 and Headset 2 groups used the Oculus Rift DK2, which provided stereoscopic 3D visuals and head tracking. This immersive setup allowed for a more dynamic first-person perspective, enhancing presence and embodiment within the simulator. Both HMD groups viewed the same virtual scenes and completed the same navigation course but differed in the motion profile of the wheelchair avatar.

4.2.3 Study 2: Results

To provide a comprehensive evaluation of simulator configurations, the results of Study 2 are organized across four data modalities: simulator-based performance metrics, explicit measures obtained through subjective self-report questionnaires, implicit physiological indicators, and behavioural markers derived from head movement data.

Each modality underwent appropriate pre-processing and statistical analysis to detect between-group differences and explore relationships between variables. To assess associations between subjective and implicit metrics responses, correlation analyses were conducted using Spearman's rank correlation coefficient, given the non-parametric nature of several measures. Behavioural data on head movements were included as an exploratory metric to supplement the interpretation of user affective engagement and tolerance across conditions.

4.2.3.1 Demographics

Fifty-seven participants were distributed across three experimental groups: Desktop (n = 24), Headset-1 (n = 17), and Headset-2 (n = 16). Mean age was 26.17 years (SD = 5.05) in the Desktop group, 30.29 years (SD = 7.33) in Headset-1, and 30.13 years (SD = 12.70) in Headset-2. Gender distribution was balanced in the Desktop (12 males, 12 females) and Headset-2 groups (8 males, 8 females), while Headset-1 included more males (10 males, 7 females).

4.2.3.2 Statistics Analysis Results

The normality of continuous variables was assessed using the Shapiro–Wilk test within each experimental group. Normality assumptions were met for a few key variables in the Headset-2 group

and in at least one of the comparison groups (Headset-1 or Desktop), permitting the use of parametric tests in those specific cases. For variables that did not meet the assumption of normality, non-parametric methods were applied.

Group differences across the three simulator configurations were first analysed using the Kruskal–Wallis test. Where significant effects were found, pairwise comparisons were conducted using the Mann–Whitney U test with Bonferroni correction applied to adjust for multiple testing. A detailed summary of Shapiro–Wilk test results is provided in Appendix.

4.2.3.3 Simulator-based Performance Metrics

A Kruskal–Wallis H test revealed significant differences in task completion time and number of collisions across the groups (Desktop, Headset-1, and Headset-2 groups). No significant difference was found for joystick command count. As seen in the mean rank comparisons in Table 4.10, Headset-2 users took longer to complete the task but had fewer collisions.

Follow-up Mann–Whitney U tests, with a Bonferroni-adjusted significance level ($p<.0167$), confirmed that Headset-2 participants had significantly longer task times than both the Desktop and Headset-1 groups (Table 4.11). This is directly attributable to the s-curve acceleration profile used in this condition, which, by design, smooths motion transitions to reduce jarring movements and cybersickness. While this approach improves user comfort, it inherently extends the time required to complete the task.

In terms of collision counts, Headset-2 users experienced significantly fewer collisions than the Desktop group ($p=0.005$), suggesting improved obstacle avoidance under smoother motion. However, no significant difference was found between Headset-2 and Headset-1. Joystick command frequency did not differ significantly between any groups. Despite a slight numerical decrease in the Headset-2 group, the results suggest that participants maintained similar control strategies across all simulator configurations. The frequency of joystick command usage did not differ significantly between any of the groups. Despite a slight numerical decrease in the Headset-2 group, the non-significant result suggests that participants maintained similar control strategies regardless of the display modality or motion dynamics. As illustrated in the boxplots in Figure 4.15, this consistency in control behaviour across all conditions is a notable finding.

In summary, the smoother motion profile in the Headset-2 configuration led to a trade-off: an increase in task duration in exchange for a significant reduction in collisions. The stability of joystick usage across all three groups indicates that while the motion profile and display modality affected performance outcomes, they did not fundamentally alter the participants' control behaviour.

Table 4.10: Study 2 Kruskal–Wallis test results for performance metrics

Metric	Group	Mean [SD]	Median [IQR]	H	p-value	Effect Size (η^2)	Post hoc
Total commands	Desktop	78.21 (21.60)	73.50 [67.00 – 82.50]	5.35	0.069	0.06	ns
	Headset-1	69.12 (16.77)	67.00 [57.00 – 74.00]	5.35	0.069	0.06	ns
	Headset-2	64.94 (24.76)	59.00 [44.50 – 79.00]	5.35	0.069	0.06	ns
Total collisions	Desktop	5.96 (3.83)	5.50 [3.00 – 8.00]	7.24	0.027	0.10	*
	Headset-1	4.71 (3.64)	3.00 [2.00 – 7.25]	7.24	0.027	0.10	ns
	Headset-2	3.50 (4.82)	2.00 [1.50 – 3.50]	7.24	0.027	0.10	*
Task completion time (s)	Desktop	186.71 (45.25)	176.00 [162.00 – 203.00]	12.29	0.002	0.19	*
	Headset-1	179.18 (38.41)	169.00 [150.25 – 199.25]	12.29	0.002	0.19	**
	Headset-2	231.31 (52.11)	218.00 [186.00 – 263.00]	12.29	0.002	0.19	*

ns= not significant, *= significant difference vs. Desktop ($p < .0167$), **= significant difference vs. Headset-1 ($p < .0167$).

Table 4.11: Study 2 Mann-Whitney U pairwise comparisons for performance metrics

		Headset-2	Compared Group	Test Statistics		
	Group Comparison	Mean (SD)	Mean (SD)	U	Z	p-value
Number of Joystick Commands	Headset-2 vs. Desktop	64.94 (24.76)	78.21 (21.60)	118.50	-2.030	.041
	Headset-2 vs. Headset-1		69.12 (16.77)	109.50	-0.955	.345
Number of Collisions	Headset-2 vs. Desktop	3.50 (4.82)	5.96 (3.83)	91.00	-2.806	.005
	Headset-2 vs. Headset-1		4.71 (3.64)	101.00	-1.278	.217
Task Completion Time (s)	Headset-2 vs. Desktop	231.31 (52.11)	186.71 (45.25)	82.000	-3.037	.002*
	Headset-2 vs. Headset-1		179.18 (38.41)	52.000	-3.027	.002*

*. $p < .0167$, Bonferroni-corrected threshold for multiple comparisons. Values are reported as Mean.

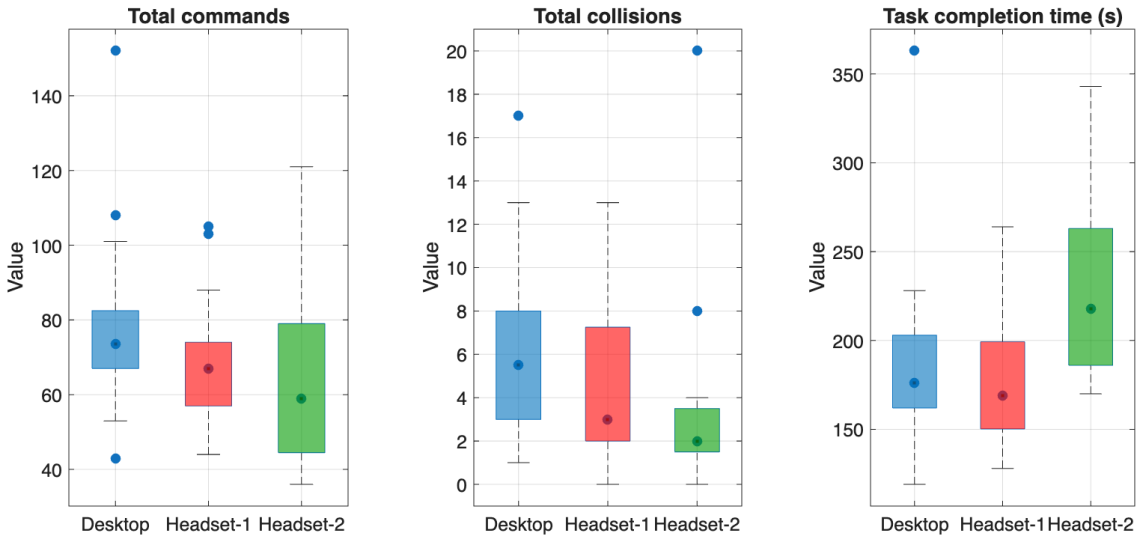


Figure 4.15: Study 2 boxplots showing distribution of task time, joystick commands and collisions.

4.2.3.4 Explicit Measures (subjective self-report questionnaires)

This section presents user-reported outcomes from standardized questionnaires used to assess QoE across different simulator configurations. The instruments include the SUS, IPQ, NASA-TLX, SAM, and SSQ. Results are compared across the three conditions (Desktop, Headset-1, and Headset-2) focusing on perceived usability, cognitive workload, sense of presence, emotional response, and simulator-induced discomfort. Appropriate non-parametric statistical tests were employed (Kruskal–Wallis with Bonferroni-corrected Mann–Whitney U post hoc tests), and effect sizes were calculated to aid interpretation.

4.2.3.4.1 Usability (SUS)

User perceptions of usability were assessed using the System Usability Scale (SUS). While Study 1 found a significant difference only for Item Q2 ("Easy to use") between the Desktop and Headset-1 conditions, Study 2 expanded this comparison to include the Headset-2 group and revealed significant effects for Q2 ("Easy to use"), Q3 ("Learnability"), and the overall SUS score.

As shown in the Kruskal-Wallis test results in Table 4.12, Headset-2 consistently outperformed the Desktop group, with statistically significant differences for both Q2 and Q3 ($p < .0167$) and large effect sizes ($\eta^2 = 0.23$ and $\eta^2 = 0.12$, respectively). The overall SUS score was also significantly higher in Headset-2 than Desktop ($p = .030$, $\eta^2 = 0.09$).

In contrast, no statistically significant differences were found between Headset-2 and Headset-1 on these items, nor were any differences observed for Items Q1 (System Complexity), Q4 (Navigation/Intuitiveness), or Q5 (Ease of Learning).

As shown in Figure 4.16 and Mann–Whitney U tests (Table 4.13) which presents mean SUS item scores by group, participants rated Headset-2 higher than Desktop on several items, specifically "Easy to use" (Q2) and "Learnability" (Q3). This significant increase in learnability scores for Headset-2 suggests that the smoother motion and immersive setup may have made the simulator feel more intuitive or approachable for new users.

These findings build upon Study 1 by confirming that the immersive Headset-2 setup not only maintains the high usability levels observed in Headset-1 but also improves user-reported ease of use (Q2) and learnability (Q3) compared to the Desktop. The significant difference in Q2 ("I thought the system was easy to use") indicates that participants perceived the interface and controls in Headset-2 as more accessible during the session.

The lack of significant differences for Q4 ("Navigation/Intuitiveness") suggests that while the smooth motion and immersion may have made the system feel easier to learn and use, they did not fundamentally change participants' perceptions of how intuitive the navigation was.

Regarding Q5 ("Ease of Learning"), the data shows no significant difference between the groups (Kruskal-Wallis p=.842), and the post-hoc pairwise comparisons confirm this. While the question asks about "ease of learning," the significant result for Q3 ("Learnability") is particularly noteworthy because it assesses the perceived learnability from a user's perspective. The higher ratings for Q3 in the Headset-2 condition, despite Q5 showing no significant change, suggest that users found the system's design itself to be more inherently learnable.

The lack of significant differences between Headset-2 and Headset-1 for any of these items may reflect that both immersive configurations provided comparable levels of intuitiveness and user support, despite their differing motion profiles.

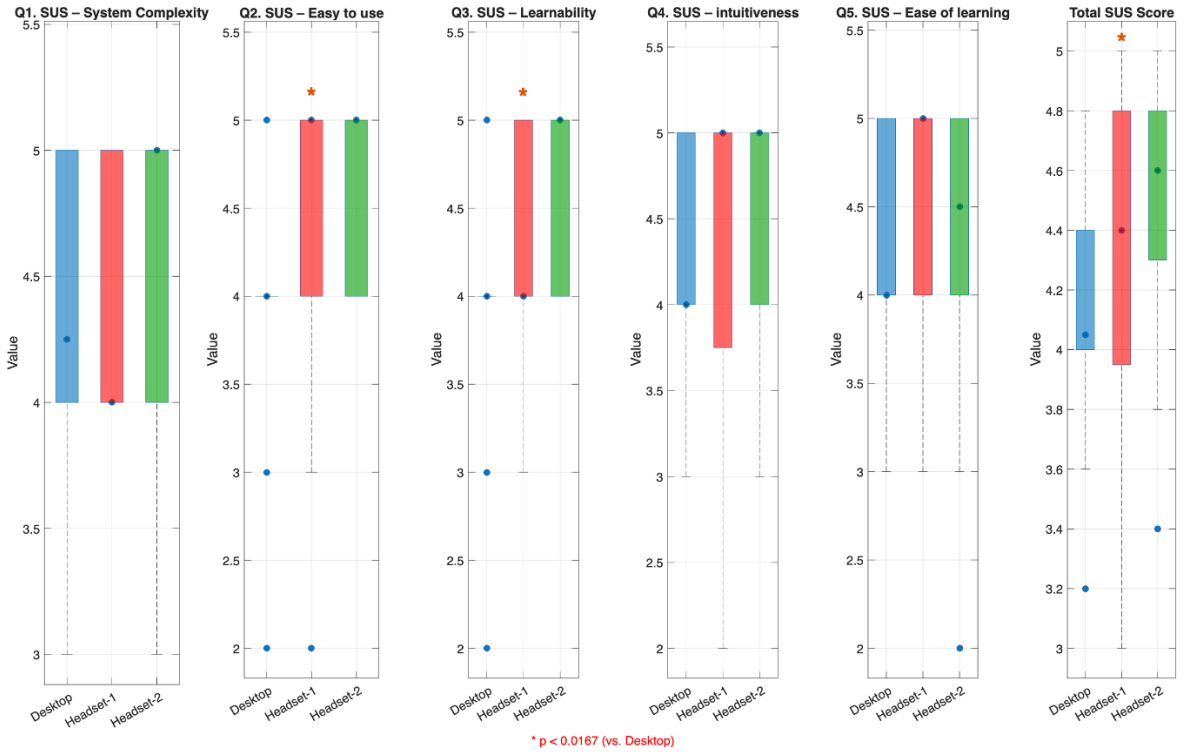


Figure 4.16: Study 2 boxplot for SUS item scores by Group. Desktop(blue), headset-1(red) and headset-2 (green).

Table 4.12: Study 2 Kruskal–Wallis test results for SUS items

Metric	Group	Mean (SD)	Median [IQR]	H	p-value	Effect Size (η^2)	Post hoc
Q1. SUS – System Complexity	Desktop	4.40 (0.64)	4.25 [4.00 – 5.00]	0.31	0.857	0.03	ns
	Headset-1	4.35 (0.49)	4.00 [4.00 – 5.00]	0.31	0.857	0.03	ns
	Headset-2	4.38 (0.81)	5.00 [4.00 – 5.00]	0.31	0.857	0.03	ns
Q2. SUS – Easy to use	Desktop	3.83 (0.64)	4.00 [4.00 – 4.00]	14.39	0.001	0.23	*
	Headset-1	4.35 (0.86)	5.00 [4.00 – 5.00]	14.39	0.001	0.23	ns
	Headset-2	4.62 (0.50)	5.00 [4.00 – 5.00]	14.39	0.001	0.23	*
Q3. SUS – Learnability	Desktop	3.96 (0.75)	4.00 [4.00 – 4.00]	8.57	0.014	0.12	*
	Headset-1	4.24 (0.75)	4.00 [4.00 – 5.00]	8.57	0.014	0.12	ns
	Headset-2	4.62 (0.50)	5.00 [4.00 – 5.00]	8.57	0.014	0.12	*
Q4. SUS – Navigation/ intuitiveness	Desktop	4.25 (0.61)	4.00 [4.00 – 5.00]	3.60	0.165	0.03	ns
	Headset-1	4.24 (0.97)	5.00 [3.75 – 5.00]	3.60	0.165	0.03	ns
	Headset-2	4.62 (0.62)	5.00 [4.00 – 5.00]	3.60	0.165	0.03	ns
Q5. SUS – Ease of learning	Desktop	4.29 (0.62)	4.00 [4.00 – 5.00]	0.34	0.842	0.03	ns
	Headset-1	4.35 (0.79)	5.00 [4.00 – 5.00]	0.34	0.842	0.03	ns
	Headset-2	4.31 (0.87)	4.50 [4.00 – 5.00]	0.34	0.842	0.03	ns
Total SUS Score	Desktop	4.15 (0.37)	4.05 [4.00 – 4.40]	7.03	0.030	0.09	*
	Headset-1	4.31 (0.59)	4.40 [3.95 – 4.80]	7.03	0.030	0.09	ns
	Headset-2	4.51 (0.46)	4.60 [4.30 – 4.80]	7.03	0.030	0.09	*

ns= not significant, *= significant difference vs. Desktop ($p < .0167$), **= significant difference vs. Headset-1 ($p < .0167$).

Table 4.13: Study 2 Mann-Whitney U Test Pairwise comparisons for SUS and IPQ items.

	Group Comparison	Headset-2	Compared Group	Test Statistics		
		Mean (SD)	Mean (SD)	U	Z	p-value
Q1. SUS	Headset-2 vs. Desktop	4.375 (0.81)	4.40 (0.64)	186.50	-0.17	.881
	Headset-2 vs. Headset-1		4.35 (0.49)	124.00	-0.48	.683
Q2. SUS	**Headset-2 vs. Desktop**	**4.625 (0.50)**	**3.83 (0.64)**	**73.00**	**-3.71**	**.001***
	Headset-2 vs. Headset-1		4.35 (0.86)	117.00	-0.78	.510
Q3. SUS	Headset-2 vs. Desktop	4.625 (0.50)	3.96 (0.75)	95.00	-3.06	.007
	Headset-2 vs. Headset-1		4.24 (0.75)	98.00	-1.53	.179
Q4. SUS	Headset-2 vs. Desktop	4.625 (0.62)	4.25 (0.61)	127.00	-2.00	.075
	Headset-2 vs. Headset-1		4.24 (0.97)	108.00	-1.16	.326
Q5. SUS	Headset-2 vs. Desktop	4.3125 (0.87)	4.29 (0.62)	176.00	-0.49	.672
	Headset-2 vs. Headset-1		4.35 (0.79)	134.50	-0.06	.958
Q6. IPQ	**Headset-2 vs. Desktop**	**4.375 (0.72)**	**3.08 (1.14)**	**68.00**	**-3.57**	**.000***
	Headset-2 vs. Headset-1		4.35 (0.79)	136.00	0.00	1.000
Q7. IPQ	Headset-2 vs. Desktop	4.125 (0.81)	3.46 (0.93)	111.50	-2.40	.025
	Headset-2 vs. Headset-1		4.18 (0.81)	130.50	-0.22	.845
Q8. IPQ	Headset-2 vs. Desktop	4.375 (0.50)	3.38 (1.17)	94.00	-2.95	.006
	Headset-2 vs. Headset-1		4.06 (1.03)	118.00	-0.73	.533
Q9. IPQ	Headset-2 vs. Desktop	3.3125 (1.08)	2.83 (1.09)	142.00	-1.45	.174
	Headset-2 vs. Headset-1		2.94 (1.39)	117.50	-0.69	.510
Q10. IPQ	Headset-2 vs. Desktop	3.4375 (0.89)	3.25 (0.68)	168.50	-0.72	.521
	Headset-2 vs. Headset-1		3.65 (1.11)	114.50	-0.81	.444
Total SUS Score	**Headset-2 vs. Desktop**	**4.5125 (0.46)**	**4.15 (0.37)**	**94.50**	**-2.72**	**.006***
	Headset-2 vs. Headset-1		4.31 (0.59)	109.50	-0.97	.345
Total IPQ Score	**Headset-2 vs. Desktop**	**3.925 (0.50)**	**3.20 (0.68)**	**70.00**	**-3.39**	**.000***
	Headset-2 vs. Headset-1		3.84 (0.59)	129.50	-0.24	.817

***= significant difference ($p < .0167$).

4.2.3.4.2 Presence (IPQ)

Table 4.14 summarizes the results of the Kruskal–Wallis tests for IPQ items. Table 4.13 presents the post-hoc pairwise comparisons. Statistically significant group differences were found for Q6 (General Presence), Q8 (Spatial Presence – "Sense of being in VE"), and the Total IPQ Score. Headset-2 participants reported significantly higher scores than Desktop in these three measures (all $p < .0167$), with large effect sizes ($\eta^2 = 0.30$, 0.14, and 0.23, respectively). No statistically significant differences were found between Headset-2 and Headset-1 across any item.

These results expand upon the findings from Study 1, where Headset-1 significantly outperformed Desktop in Q6 and Q7, as well as in the total score, indicating that immersive display improves general and spatial presence perceptions. In Study 2, Headset-2 maintained these gains while slightly increasing spatial presence scores further (notably Q8), suggesting a positive trend in presence under the enhanced motion profile and immersive visuals.

Despite the higher mean values for Headset-2, no significant differences emerged between the two immersive conditions (Headset-1 and Headset-2). This may suggest that after the introduction of a headset-based setup, further enhancements in immersive technology result in limited gains in perceived presence. Notably, Q9 (Involvement) and Q10 (Realism) remained statistically unchanged across all groups in both studies, reinforcing earlier literature findings that these dimensions are less sensitive to differences in display simulator configuration without vestibular platform (Vailland et al., 2020).

Figure 4.17 presents a radar plot of IPQ item scores by group and Figure 4.18 presents boxplots, visually highlighting the increased ratings in both immersive configurations relative to Desktop, particularly for general and spatial presence. The consistently non-significant outcomes for involvement and realism suggest that additional interactivity, contextual detail, or narrative elements may be required to meaningfully influence these dimensions of user experience.

Overall, Study 2 supports and extends Study 1's findings by demonstrating that immersive configurations, especially Headset-2, enhance users' perceived presence environments. The most pronounced improvements were observed the general feeling of being "inside" the simulation and interacting spatially with it, while further design improvements may be required to influence deeper engagement or realism, pointing to opportunities for further system development.

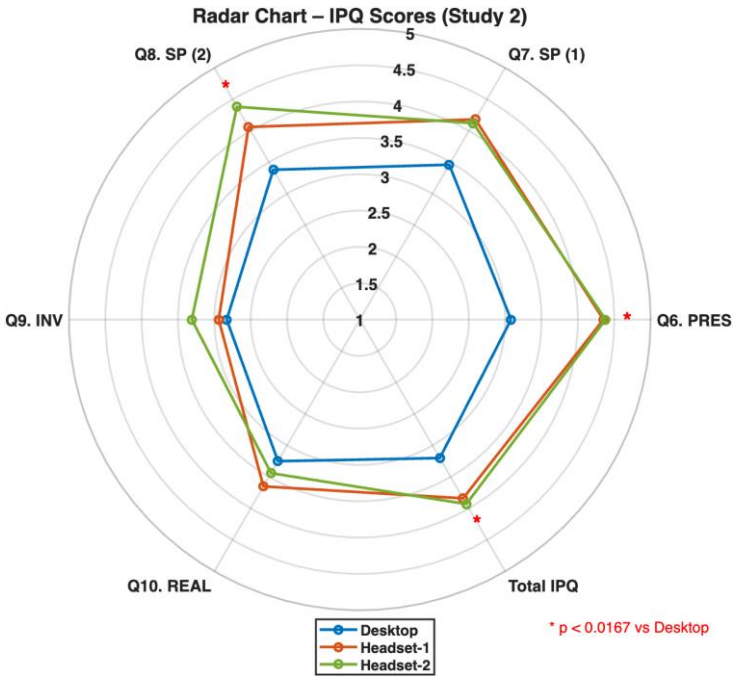


Figure 4.17: Study 2 Radar plot for IPQ items.

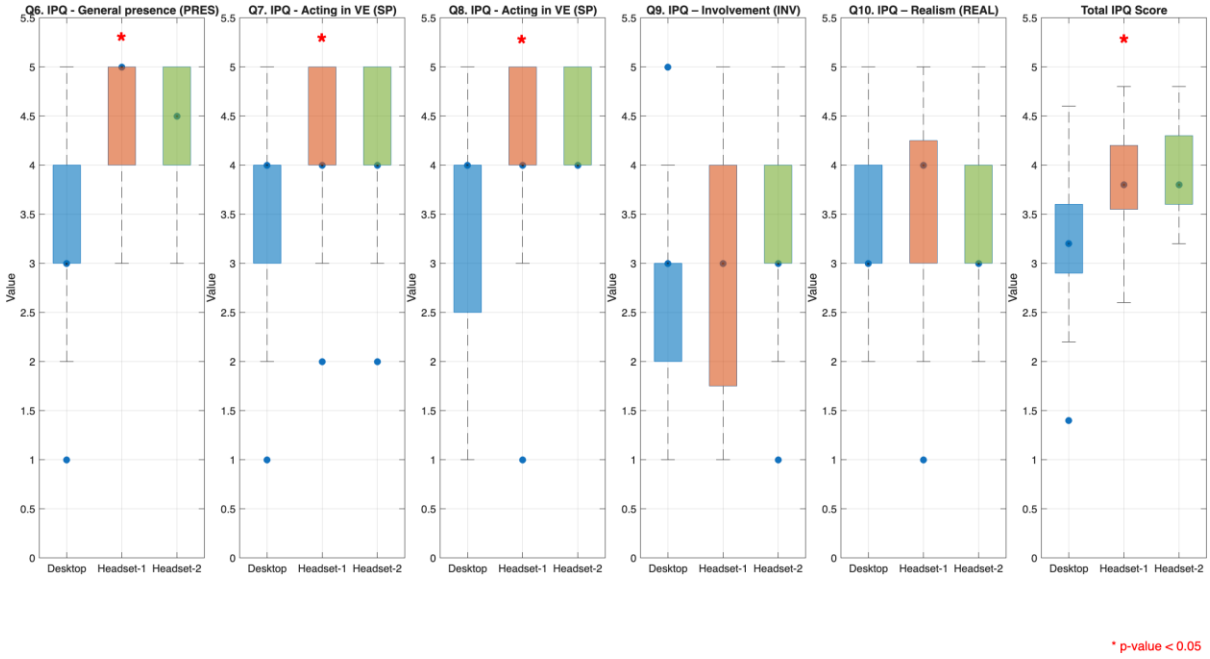


Figure 4.18: Study 2 boxplots of IPQ items.

Table 4.14: Study 2 Kruskal–Wallis test results for IPQ items

Metric	Group	Mean (SD)	Median [IQR]	H	p-value	Effect Size (η^2)	Posthoc
Q6. IPQ - General presence (PRES)	Desktop	3.08 (1.14)	3.00 [3.00 – 4.00]	17.97	0.000	0.30	*
	Headset-1	4.35 (0.79)	5.00 [4.00 – 5.00]	17.97	0.000	0.30	ns
	Headset-2	4.38 (0.72)	4.50 [4.00 – 5.00]	17.97	0.000	0.30	*
Q7. IPQ - Acting in VE (SP)	Desktop	3.46 (0.93)	4.00 [3.00 – 4.00]	9.12	0.010	0.13	ns
	Headset-1	4.18 (0.81)	4.00 [4.00 – 5.00]	9.12	0.010	0.13	ns
	Headset-2	4.12 (0.81)	4.00 [4.00 – 5.00]	9.12	0.010	0.13	ns
Q8. IPQ - Acting in VE (SP)	Desktop	3.38 (1.17)	4.00 [2.50 – 4.00]	9.48	0.009	0.14	*
	Headset-1	4.06 (1.03)	4.00 [4.00 – 5.00]	9.48	0.009	0.14	ns
	Headset-2	4.38 (0.50)	4.00 [4.00 – 5.00]	9.48	0.009	0.14	*
Q9. IPQ – Involvement (INV)	Desktop	2.83 (1.09)	3.00 [2.00 – 3.00]	1.84	0.398	0.00	ns
	Headset-1	2.94 (1.39)	3.00 [1.75 – 4.00]	1.84	0.398	0.00	ns
	Headset-2	3.31 (1.08)	3.00 [3.00 – 4.00]	1.84	0.398	0.00	ns
Q10. IPQ – Realism (REAL)	Desktop	3.25 (0.68)	3.00 [3.00 – 4.00]	2.76	0.252	0.01	ns
	Headset-1	3.65 (1.11)	4.00 [3.00 – 4.25]	2.76	0.252	0.01	ns
	Headset-2	3.44 (0.89)	3.00 [3.00 – 4.00]	2.76	0.252	0.01	ns
Total IPQ Score	Desktop	3.20 (0.68)	3.20 [2.90 – 3.60]	14.27	0.001	0.23	*
	Headset-1	3.84 (0.59)	3.80 [3.55 – 4.20]	14.27	0.001	0.23	ns
	Headset-2	3.93 (0.50)	3.80 [3.60 – 4.30]	14.27	0.001	0.23	*

ns= not significant, *= significant difference vs. Desktop ($p < .0167$), **= significant difference vs. Headset-1 ($p < .0167$).

4.2.3.4.3 Emotional Response (SAM)

Emotional responses were assessed using SAM, covering valence, arousal, and dominance. A Kruskal–Wallis H test revealed a significant group effect for arousal ($p = .004$, $\eta^2=0.16$), with no significant differences for valence or dominance (Table 4.15). These results are illustrated in Figure 4.19, which depicts a three-dimensional scatterplot of group-level Valence–Arousal–Dominance (VAD) responses with group-level centroids.

Post hoc Mann–Whitney U tests revealed that participants in the Headset-2 condition reported significantly higher arousal levels compared to those in the Desktop and Headset-1 conditions. On average, Headset-2 users showed the highest average arousal, followed by Headset-1 and Desktop (see Table 4.16). In contrast, valence ratings were consistently high across all groups, with no statistically significant differences. Although Headset-1 presented slightly higher mean dominance scores, these differences were not statistically meaningful.

The elevated arousal levels in the Headset-2 condition, relative to both the Desktop and Headset-1 conditions, may reflect a heightened emotional response to the combined effect of the immersive visual display and the smooth motion profile. This suggests that the smoother, more predictable motion of the s-curve profile may have been a more engaging or psychologically activating experience than the harsher, high-jerk motion of the Headset-1 group.

However, it is important to interpret this heightened arousal with caution. While it may indicate greater excitement or alertness, it could also signal physiological discomfort, particularly given that cybersickness symptoms were more frequently reported in both immersive groups. This suggests that the arousal dimension in immersive virtual environments may reflect a complex interplay of engagement and sensory strain. The relationship between arousal and simulator-induced discomfort is discussed further in the Simulator Sickness Questionnaire (SSQ) results section.

In summary, Headset-2 configuration elicited stronger affective activation in terms of arousal, while valence and dominance remained unaffected. This observation also was found with a broader literature suggesting arousal may be responsive to immersive system characteristics, particularly those involving multisensory input (Magalhães et al., 2024). Nonetheless, changes in valence and dominance have also been reported in other contexts, depending on factors such as content, interactivity, and user traits (Liao et al., 2020; Tian et al., 2021). Thus, while arousal appears more sensitive to immersive system features in this study, its interpretation should account for both engagement-related and discomfort-related factors commonly associated with immersive virtual environments.

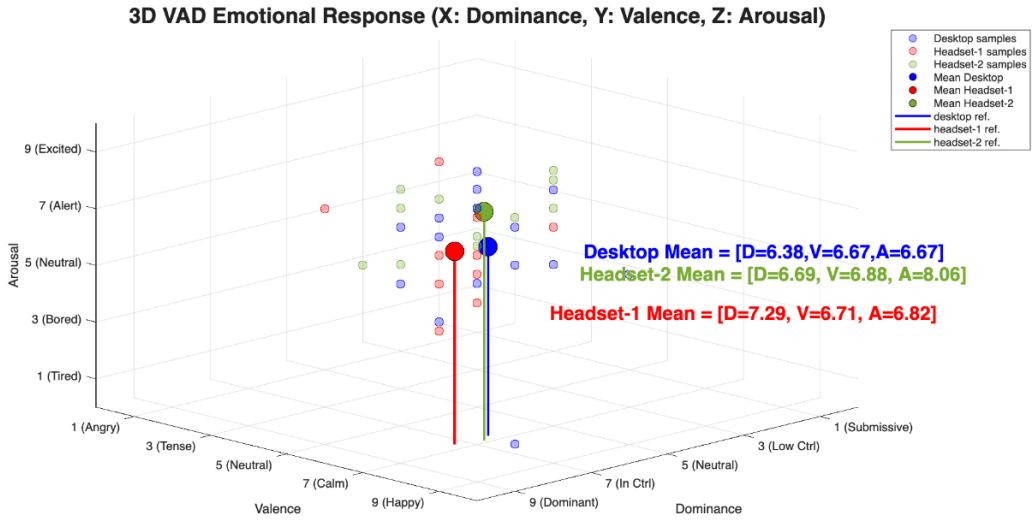


Figure 4.19: Study 2 VAD (Valence–Arousal–Dominance) emotional responses.

Table 4.15: Study 2 Kruskal–Wallis test results for SAM items

Metric	Group	Mean (SD)	Median [IQR]	H	p-value	Effect Size (η^2)	Post hoc
Valence	Desktop	6.67 (1.69)	7.00 [5.00 – 8.00]	0.23	0.890	0.03	ns
	Headset-1	6.71 (1.76)	7.00 [7.00 – 7.00]	0.23	0.890	0.03	ns
	Headset-2	6.88 (1.45)	7.00 [6.00 – 7.50]	0.23	0.890	0.03	ns
Arousal	Desktop	6.67 (1.61)	7.00 [6.50 – 7.00]	10.89	0.004	0.16	*
	Headset-1	6.82 (1.24)	7.00 [6.00 – 7.25]	10.89	0.004	0.16	**
	Headset-2	8.06 (1.06)	8.50 [7.00 – 9.00]	10.89	0.004	0.16	*
Dominance	Desktop	6.38 (1.81)	7.00 [5.00 – 7.50]	2.96	0.228	0.02	ns
	Headset-1	7.29 (1.53)	7.00 [7.00 – 8.25]	2.96	0.228	0.02	ns
	Headset-2	6.69 (1.62)	7.00 [5.00 – 8.00]	2.96	0.228	0.02	ns

ns= not significant,*= significant difference vs. Desktop (p< .0167), **= significant difference vs. Headset-1 (p< .0167).

Table 4.16: Study 2 Post Hoc Mann-Whitney U Test Pairwise comparisons for SAM items.

		Headset-2	Compared Group	Test Statistics		
SAM Scales	Group Comparison	Mean (SD)	Mean (SD)	U	Z	p-value
Valence (SAM)	Headset-2 vs. Desktop	6.88 (1.45)	6.67 (1.69)	179.00	-0.371	.733
	Headset-2 vs. Headset-1		6.71 (1.76)	133.00	-0.118	.929
Arousal (SAM)	Headset-2 vs. Desktop	8.06 (1.06)	6.67 (1.61)	88.50	-3.053	.003*
	Headset-2 vs. Headset-1		6.82 (1.24)	64.00	-2.703	.009*
Dominance (SAM)	Headset-2 vs. Desktop	6.69 (1.62)	6.38 (1.81)	174.50	-0.496	.633
	Headset-2 vs. Headset-1		7.29 (1.53)	108.50	-1.027	.326

*= significant difference (p< .0167).

4.2.3.4.4 Cognitive Workload (NASA-TLX)

Kruskal–Wallis tests indicated significant group effects for performance demand, performance-weighted scores, and the overall weighted TLX score (see Table 4.17). Participants in both immersive conditions (Headset-1 and Headset-2) reported higher perceived workload than those in the Desktop condition, particularly in how well they believed they performed the task. However, no significant differences emerged between the two headset groups, despite slightly lower workload ratings in Headset-2 (post hoc table summarised in Table 4.18).

Post hoc tests confirmed that performance-related workload was significantly higher in Headset-2 compared to Desktop ($p < .0167$), but no significant pairwise differences were observed for the overall TLX scores, even though effect sizes for raw ($\eta^2 = 0.11$) and weighted TLX scores ($\eta^2 = 0.18$) were moderate. This suggests meaningful group-level variance, though not strong enough to yield significance between immersive conditions after correction. Figure 4.20 shows radar plots of the subscale distributions, while Figure 4.21 and Figure 4.23 summarizes TLX scores across groups.

The non-significant difference in overall workload scores between the Headset-1 (high-jerk) and Headset-2 (low-jerk) conditions is a key finding. This suggests that while a smoother motion profile may offer a slight subjective improvement in workload, it is not a large enough factor on its own to significantly reduce the overall cognitive burden imposed by the immersive display. The data reinforce the conclusion from Study 1 that the immersive display itself is the primary contributor to elevated perceived workload. The most pronounced effect remained on the Performance subscale, where participants felt they performed worse in the immersive conditions, regardless of the motion profile.

These findings align with previous simulator studies that have demonstrated how perceived cognitive workload can vary depending on the type of system feedback and the level of immersion in simulator design. For example, (Fraudet et al., 2024) reported increased mental workload in immersive VR settings compared to equivalent non-immersive (real-world) tasks. Interestingly, variations in task difficulty within the immersive setting did not significantly affect perceived cognitive load, suggesting that the immersive setup itself contributed to the increased demand. Similarly, (Kamaraj, Dicianno, et al., 2016b) found elevated frustration and mental workload in four-screen VR conditions relative to non-immersive driving. Collectively, these studies support the view that simulator-induced cognitive workload is influenced more by core design features such as immersive display characteristics than by subtle changes in motion dynamics.

In conclusion, both Study 1 and Study 2 reinforce the finding that immersive virtual environments can increase perceived cognitive workload, particularly affecting users' self-perception of their

performance. While smoother motion may offer slight relief, it was insufficient to produce statistically significant differences in cognitive workload between the headset conditions. Future work should therefore focus on how refinements to sensory feedback, interaction, and task design can reduce this cognitive burden without compromising the immersive experience.

Table 4.17: Study 2 summary test results for NASA-TLX items

Metric	Group	Mean (SD)	Median [IQR]	H	p-value	Effect size (η^2)	Post hoc
performance	Desktop	28.33 (20.41)	22.50 [15.00 – 40.00]	28.29	0.000	0.49	*
	Headset-1	79.71 (22.34)	82.50 [76.25 – 93.75]	28.29	0.000	0.49	ns
	Headset-2	77.19 (22.98)	82.50 [67.50 – 95.00]	28.29	0.000	0.49	*
Performance weighted	Desktop	101.88 (80.47)	77.50 [42.50 – 142.50]	25.52	0.000	0.44	*
	Headset-1	329.56 (144.17)	387.50 [235.62 – 437.50]	25.52	0.000	0.44	ns
	Headset-2	276.09 (111.25)	283.75 [228.75 – 343.75]	25.52	0.000	0.44	*
raw TLX	Desktop	31.32 (12.13)	30.84 [23.33 – 40.83]	8.18	0.017	0.11	ns
	Headset-1	41.94 (12.72)	45.00 [36.46 – 50.94]	8.18	0.017	0.11	ns
	Headset-2	39.53 (12.12)	40.84 [32.08 – 45.41]	8.18	0.017	0.11	ns
weighted TLX	Desktop	39.24 (16.88)	37.50 [29.17 – 49.50]	11.78	0.003	0.18	ns
	Headset-1	56.03 (13.39)	55.50 [51.71 – 65.92]	11.78	0.003	0.18	ns
	Headset-2	51.31 (13.13)	54.83 [42.17 – 60.84]	11.78	0.003	0.18	ns

ns= not significant, *= significant difference vs. Desktop (p< .0167), **= significant difference vs. Headset-1 (p< .0167).

Table 4.18: Study 2 Pairwise Comparisons Using Mann–Whitney U Tests for NASA-TLX Subscales

		Headset-2	Compared Group	Test Statistics		
Metric	Group Comparison	Mean (SD)	Mean (SD)	U	Z	p-value
Performance Demand	Headset-2 vs. Desktop	77.19 (22.98)	28.33 (20.41)	34.00	-4.38	0.000*
	Headset-2 vs. Headset-1		79.71 (22.34)	127.50	-0.31	0.763
Weighted Performance	Headset-2 vs. Desktop	276.09 (111.25)	101.88 (80.47)	43.50	-4.10	0.000*
	Headset-2 vs. Headset-1		329.56 (144.17)	96.50	-1.42	0.157
Raw TLX Score	Headset-2 vs. Desktop	39.53 (12.12)	31.32 (12.13)	122.50	-1.92	0.054
	Headset-2 vs. Headset-1		41.94 (12.72)	114.50	-0.78	0.444
Weighted TLX Score	Headset-2 vs. Desktop	51.31 (13.13)	39.24 (16.88)	107.00	-2.35	0.018*
	Headset-2 vs. Headset-1		56.03 (13.39)	113.00	-0.83	0.423

*. The difference is significant at the 0.05 level.

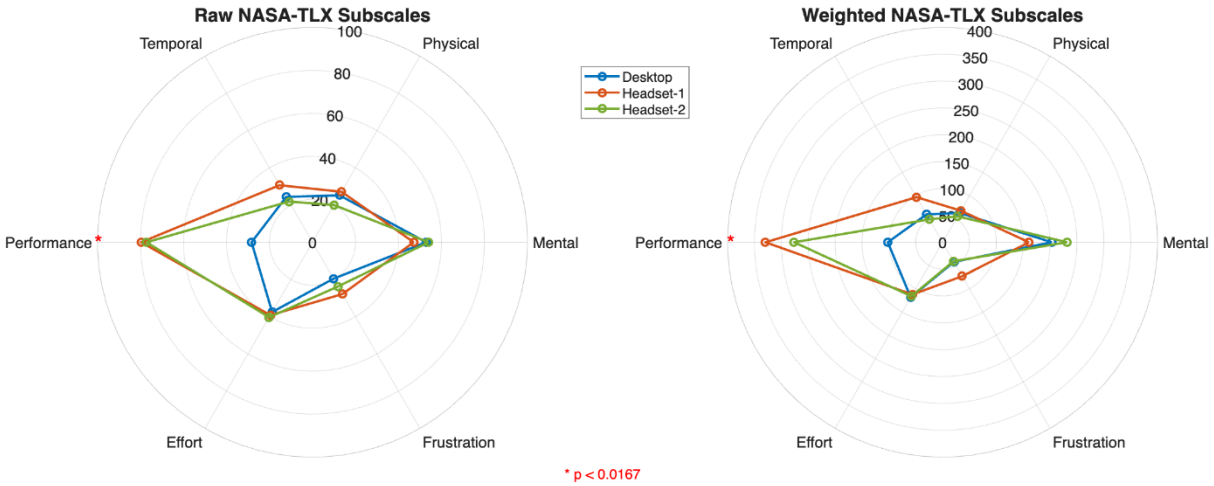


Figure 4.20: Study 2 radar plots of mean NASA-TLX items.

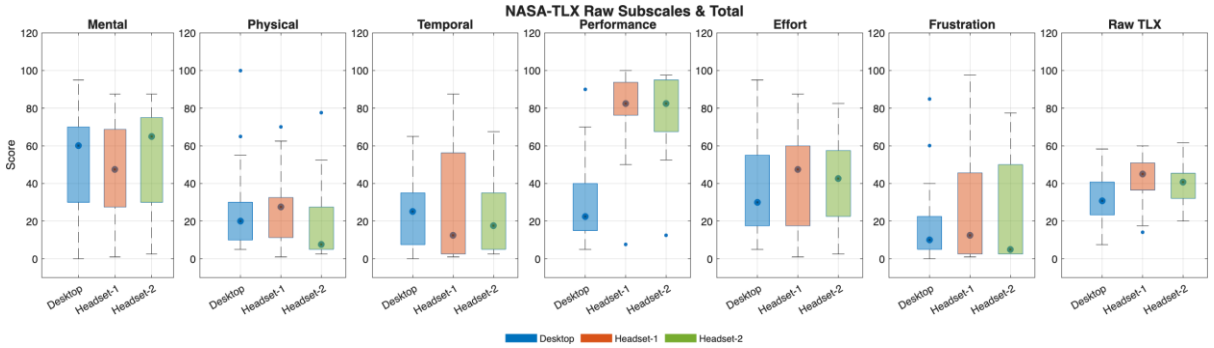


Figure 4.21: Study 2 boxplots of NASA-TLX items (raw).

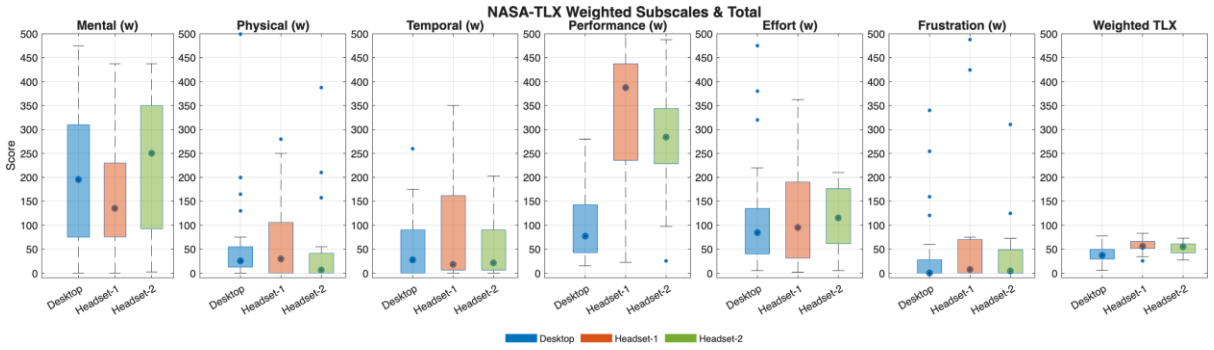


Figure 4.22: Study 2 boxplots of NASA-TLX items (weighted).

4.2.3.4.5 Cybersickness (SSQ)

Cybersickness was assessed using the Simulator Sickness Questionnaire (SSQ), comparing the two immersive conditions: Headset-1 (high-jerk motion) and Headset-2 (low-jerk motion). Previous studies by (Vailland et al., 2021; Zorzi et al., 2023a) interpreted SSQ severity levels using reference thresholds reported by (John et al., 2018), reproduced in Table 4.19. These thresholds offer classification bands for each subscale and the total score, ranging from "None" to "Severe." Descriptive statistics for the pre- and post-evaluation SSQ scores are presented Table 4.20 and Table 4.21.

Difference scores (Post–Pre) and inferential test results are provided in Table 4.22. These tables report the mean, standard deviation, range, and interquartile range for each SSQ subscale and the total score. Across both Headset conditions, mean SSQ scores remained below the "Slight" severity classification for all subscales and overall total scores. Although symptom levels generally increased after simulator use, they did not reach thresholds indicative of moderate or severe cybersickness. These results suggest that both immersive configurations were relatively well-tolerated.

Table 4.19: Study 2 SSQ reference scores from (John et al., 2018)

Severity	Nausea	Oculomotor	Disorientation	Total Score
None	0	0	0	0
Slight	66.8	53.1	97.4	78.5
Moderate	133.6	106.1	194.9	157.1
Severity	200.3	159.2	292.3	235.6

Table 4.20: Study 2 pre-evaluation findings of SSQ items

Metric	Group	Mean(SD)	Range [min-max]	Median [IQR]
General Discomfort	Headset-1	0.71 (1.36)	[0.00 - 5.00]	0.00 [0.00 – 1.00]
	Headset-2	1.00 (1.15)	[0.00 - 4.00]	1.00 [0.00 – 2.00]
Nausea	Headset-1	8.98 (13.69)	[0.00 - 47.70]	0.00 [0.00 – 19.08]
	Headset-2	11.33 (17.84)	[0.00 - 66.78]	4.77 [0.00 – 14.31]
Oculomotor	Headset-1	12.48 (14.42)	[0.00 - 45.48]	7.58 [0.00 – 17.05]
	Headset-2	15.63 (18.66)	[0.00 - 68.22]	7.58 [3.79 – 18.95]
Disorientation	Headset-1	6.73 (9.99)	[0.00 - 38.16]	0.00 [0.00 – 9.54]
	Headset-2	8.94 (13.70)	[0.00 - 47.70]	0.00 [0.00 – 14.31]
Total Score	Headset-1	12.32 (14.10)	[0.00 - 41.14]	7.48 [0.00 – 16.83]
	Headset-2	15.66 (20.40)	[0.00 - 78.54]	7.48 [3.74 – 20.57]

Table 4.21: Study 2 pos-evaluation findings of SSQ items

Metric	Group	Mean(SD)	Range [min-max]	Median [IQR]
General Discomfort	Headset-1	1.06 (2.19)	[0.00 - 9.00]	0.00 [0.00 – 1.00]
	Headset-2	0.94 (1.18)	[0.00 - 4.00]	1.00 [0.00 – 1.00]
Nausea	Headset-1	22.45 (25.67)	[0.00 - 85.86]	19.08 [0.00 – 38.16]
	Headset-2	19.08 (18.43)	[0.00 - 57.24]	9.54 [9.54 – 38.16]
Oculomotor	Headset-1	22.29 (24.48)	[0.00 - 75.80]	15.16 [0.00 – 32.22]
	Headset-2	13.27 (14.51)	[0.00 - 45.48]	7.58 [0.00 – 22.74]
Disorientation	Headset-1	22.45 (22.36)	[0.00 - 66.78]	19.08 [0.00 – 47.70]
	Headset-2	11.33 (13.59)	[0.00 - 47.70]	9.54 [0.00 – 14.31]
Total Score	Headset-1	28.60 (29.12)	[0.00 - 89.76]	26.18 [0.00 – 37.40]
	Headset-2	18.47 (17.35)	[0.00 - 56.10]	13.09 [3.74 – 35.53]

Table 4.22: Study 2 descriptive statistics and Mann–Whitney U test results for SSQ items (Post – Pre scores)

Metric	Group	Mean (SD)	Median [IQR]	U	Z	p-value	Effect size (r)
General Discomfort	Headset-1	0.35 (2.42)	0.00 [0.00 – 1.00]	114.50	-0.82	0.444	0.14
	Headset-2	-0.06 (1.00)	0.00 [-1.00 – 0.50]				
Nausea	Headset-1	13.47 (23.62)	0.00 [0.00 – 23.85]	133.00	-0.11	0.929	0.02
	Headset-2	7.75 (16.05)	9.54 [0.00 – 9.54]				
Oculomotor	Headset-1	9.81 (24.38)	7.58 [0.00 – 22.74]	75.50	-2.25	0.028*	0.39
	Headset-2	-2.37 (11.98)	0.00 [-7.58 – 0.00]				
Disorientation	Headset-1	15.71 (21.05)	9.54 [0.00 – 38.16]	82.00	-2.10	0.053	0.37
	Headset-2	2.38 (9.54)	0.00 [0.00 – 4.77]				
Total	Headset-1	16.28 (27.42)	7.48 [0.00 – 29.92]	89.50	-1.69	0.094	0.29
	Headset-2	2.81 (11.22)	0.00 [-3.74 – 5.61]				
*. The difference is significant at the 0.05 level.							

Inferential Statistics Analysis

Inferential statistical analysis was conducted using Mann–Whitney U tests on the difference scores (Post–Pre). As shown in Table 4.22, higher symptom levels were observed in the Headset-1 condition across all subscales. However, only the Oculomotor subscale reached nominal statistical significance ($p = .028$), while Disorientation approached significance ($p = .053$). These differences did not survive Bonferroni correction for multiple comparisons ($\alpha = .0125$). Despite this, moderate effect sizes were observed for Oculomotor ($r = .39$) and Disorientation ($r = .37$), suggesting that the high-jerk motion profile may meaningfully contribute to visual strain and spatial disorientation.

Notably, Headset-2 produced negative mean scores on the Oculomotor subscale, likely due to post–pre difference scoring procedures, and indicative of very low symptom levels under smoother motion. As illustrated in Figure 4.23, boxplots visually reinforce these trends, with Headset-1 exhibiting a broader and higher distribution of symptom scores, especially for Oculomotor and Disorientation. Colour coding highlights the contrast between motion profiles (red = Headset-1; green = Headset-2).

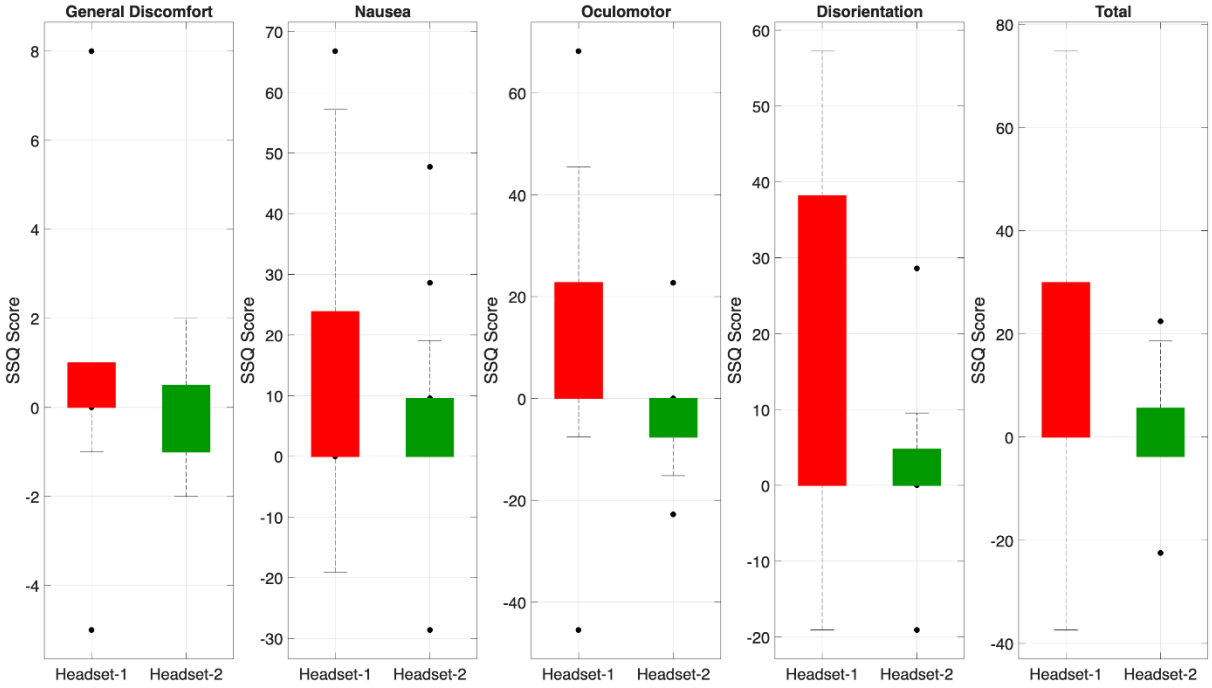


Figure 4.23: Study 2 boxplots of Simulator Sickness Questionnaire (SSQ) items.

Exploratory correlation analysis with Emotional state (SAM)

To further explore the emotional dimensions of cybersickness, a Spearman correlation analysis was conducted between SSQ subscales and SAM affective ratings (Figure 4.24). The correlation analysis revealed moderate, statistically significant negative correlations between Dominance and SSQ scores for Nausea ($\rho = -.517$, $p = .002$), Disorientation ($\rho = -.442$, $p = .010$), and Total SSQ ($\rho = -.451$, $p = .008$). These findings indicate that increased symptom severity was associated with a reduced sense of control. No significant associations were found with Valence or Arousal.

The lack of a significant correlation between SAM Arousal and any of the SSQ subscales is a crucial finding. This suggests that the heightened arousal reported in the Headset-2 condition was not a direct emotional response to the aversive symptoms of cybersickness. Instead, it likely reflects increased engagement, mental effort, or a general physiological response to the immersive environment itself, independent of feeling unwell. This distinguishes between the emotional excitement of the experience and the physical discomfort it might cause.

The most telling finding is the moderate negative correlation between cybersickness and dominance. This suggests that as users' symptoms worsened, they felt a loss of control or agency within the virtual environment.

Notably, the matrix shows no significant correlation between any SSQ or SAM metric and the perceived NASA-TLX Performance scores. This suggests that even when participants experienced discomfort or a reduced sense of control, it didn't significantly impact their self-reported feeling of how well they completed the task. This could imply that participants were able to push through the negative symptoms to focus on the task, or that the task itself was not demanding enough to be affected by the symptoms.

Although SSQ and SAM, has both been applied in immersive VR research studies (Kaufeld et al., 2022), their combined use in wheelchair simulator studies remains limited. To our knowledge, this is the first study in the wheelchair simulator context to employ both instruments. Including both affective and discomfort indicators enables a more comprehensive understanding of users' subjective experiences in immersive mobility environments. However, (Vailland et al., 2021) analysed SSQ with behavioural indicators such as simulator-based performance metrics (velocity dynamics and task time completion) and head movements kinematics (acceleration and velocities). This analysis is explored further in the head movement analysis section.

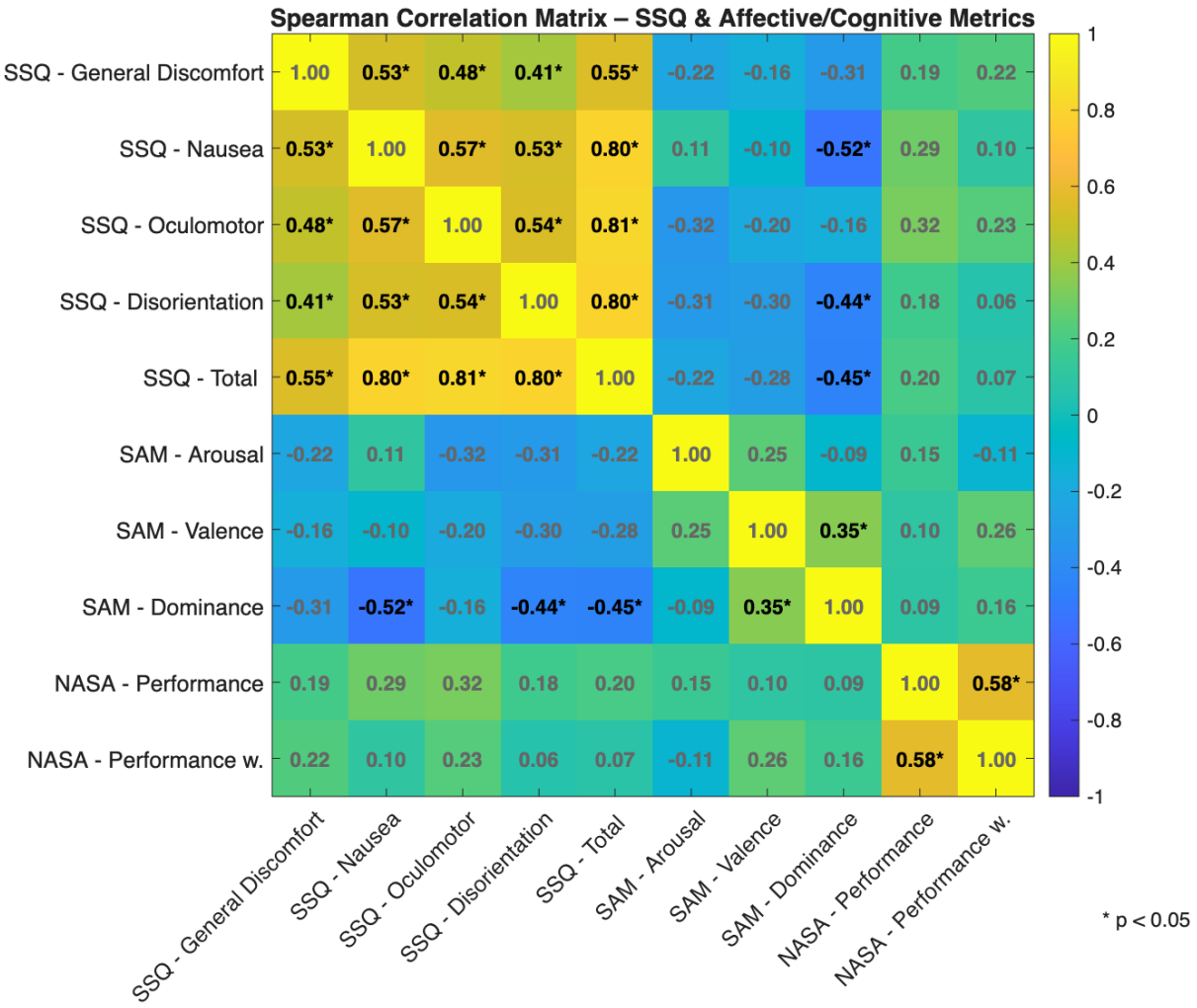


Figure 4.24: Study 2 correlation matrix between SSQ, SAM and NASA-TLX measures.

Summary

As noted, previous studies by (John et al., 2018; Vailland et al., 2021; Zorzi et al., 2023a) have employed the SSQ in immersive wheelchair simulator contexts. (Vailland et al., 2021), used the SSQ to assess the effects of vestibular feedback, reporting scores below the "Slight" threshold and improvements in presence measures (IPQ). (Zorzi et al., 2023a) reported subscale scores within the "Slight" range and a total score in the "Moderate" classification. (John et al., 2018) also reported average symptom levels falling in the "Slight" category. In comparison, Study 2 results presented here indicate that both immersive configurations produced SSQ scores consistently below "Slight" levels across all subscales and total scores.

Nonetheless, caution is warranted when interpreting SSQ results. There is ongoing debate in the literature about how to define baseline versus symptomatic levels and what constitutes an acceptable level of simulator-induced discomfort (Brown et al., 2022). Given this uncertainty and considering the importance of providing a comfortable user experience in clinical and assistive contexts, it is recommended that SSQ assessments be complemented with alternative methods. Such techniques could include physiological measuring, such as heart-rate variability and electrodermal activity (Brown et al., 2022).

Overall, the results provide preliminary evidence that high-jerk motion profiles in immersive environments may elevate cybersickness symptoms, particularly in domains related to visual fatigue and spatial disorientation. In contrast, the low-jerk immersive configuration, while offering comparable levels of immersion, was associated with lower symptom reports and greater comfort. Although some subscales reached nominal significance, these differences did not remain statistically significant after applying Bonferroni correction for multiple comparisons. This limitation, combined with a small sample size, may have reduced the sensitivity of the analysis and increased the likelihood of Type II errors. Nevertheless, moderate effect sizes and clear distribution trends reinforce the potential importance of motion smoothness in immersive simulator design. Future studies with larger sample sizes and more targeted analysis of symptom and sensor dynamics are warranted to validate these results and support motion parameter optimization in virtual training and assessment platforms.

4.2.3.5 Implicit Measures

4.2.3.5.1 Physiological Markers

Physiological responses, including HR , IBI, and EDA, were examined across three simulator configurations: Desktop, Headset-1 (high jerk) and Headset-2 (low jerk). A Kruskal–Wallis H test revealed significant group differences for HR and EDA change scores (Table 4.23 and Table 4.24).

Table 4.23: Study 2 Kruskal–Wallis test results for physiological metrics.

Metric	Group	Mean (SD)	Median [IQR]	H	p-value	Effect Size (η^2)	Post hoc
Mean HR - Difference (Test - Baseline)	Desktop	-0.71 (5.36)	-1.47 [-3.52 – 3.17]	8.63	0.013	0.12	ns
	Headset-1	5.53 (10.32)	3.50 [-1.36 – 7.72]	8.63	0.013	0.12	**
	Headset-2	-2.81 (7.92)	-3.07 [-6.98 – 1.88]	8.63	0.013	0.12	**
Mean HR - Difference (1st Collision - Baseline)	Desktop	1.70 (7.11)	0.18 [-3.10 – 7.21]	6.07	0.048	0.08	ns
	Headset-1	8.14 (15.40)	3.56 [-1.62 – 12.80]	6.07	0.048	0.08	ns
	Headset-2	-4.19 (9.52)	-1.88 [-12.50 – 2.09]	6.07	0.048	0.08	ns
Mean HR % Change (Test vs. Baseline)	Desktop	-0.75 (7.13)	-2.07 [-5.39 – 4.50]	8.53	0.014	0.12	ns
	Headset-1	7.26 (13.19)	4.09 [-1.81 – 11.04]	8.53	0.014	0.12	**
	Headset-2	-3.03 (8.65)	-3.52 [-8.79 – 2.54]	8.53	0.014	0.12	**
Mean HR % Change (1st Collision vs. Baseline)	Desktop	2.56 (10.11)	0.22 [-3.96 – 9.22]	6.39	0.041	0.08	ns
	Headset-1	10.62 (19.37)	4.55 [-2.02 – 16.84]	6.39	0.041	0.08	ns
	Headset-2	-4.32 (10.59)	-2.41 [-15.77 – 2.56]	6.39	0.041	0.08	ns
Mean IBI - Difference (1st Collision - Baseline)	Desktop	-0.06 (0.21)	-0.01 [-0.03 – 0.01]	7.11	0.029	0.09	ns
	Headset-1	-0.50 (0.40)	-0.70 [-0.81 – -0.03]	7.11	0.029	0.09	ns
	Headset-2	-0.29 (0.39)	-0.02 [-0.75 – -0.00]	7.11	0.029	0.09	ns
Mean EDA - Test	Desktop	0.85 (1.24)	0.29 [0.16 – 1.28]	8.24	0.016	0.12	*
	Headset-1	0.90 (1.27)	0.49 [0.17 – 0.76]	8.24	0.016	0.12	**
	Headset-2	5.00 (6.94)	1.53 [0.49 – 5.75]	8.24	0.016	0.12	*
Mean EDA - 1st Collision	Desktop	0.70 (0.86)	0.30 [0.17 – 0.89]	6.02	0.049	0.07	ns
	Headset-1	0.92 (1.36)	0.47 [0.16 – 0.61]	6.02	0.049	0.07	ns
	Headset-2	5.13 (7.75)	1.22 [0.32 – 6.32]	6.02	0.049	0.07	ns
Mean EDA - Difference (Test - Baseline)	Desktop	0.20 (1.01)	0.03 [-0.01 – 0.08]	10.63	0.005	0.16	*
	Headset-1	0.45 (0.97)	0.07 [-0.04 – 0.62]	10.63	0.005	0.16	ns
	Headset-2	2.71 (4.75)	0.65 [0.11 – 3.33]	10.63	0.005	0.16	*
Mean EDA - Difference (1st Collision - Baseline)	Desktop	0.04 (0.57)	0.02 [-0.02 – 0.05]	6.25	0.044	0.08	ns
	Headset-1	0.45 (1.10)	0.06 [-0.02 – 0.46]	6.25	0.044	0.08	ns
	Headset-2	2.49 (5.40)	0.34 [0.02 – 2.94]	6.25	0.044	0.08	ns

ns= not significant, *= significant difference vs. Desktop ($p < .0167$), **= significant difference vs. Headset-1 ($p < .0167$).

Table 4.24: Study 2 Mann–Whitney U test results comparing HR, EDA, and HRV (IBI) metrics

		Headset-2	Compared Group	Test Statistics		
	Group Comparison	Mean (SD)	Mean (SD)	U	Z	p-value
Mean HR - Difference (Test - Baseline)	Headset-2 vs. Desktop	-2.81 (7.92)	-0.71 (5.36)	148.00	-1.03	0.315
	Headset-2 vs. Headset-1		5.53 (10.32)	64.00	-2.59	0.009*
Mean HR - Difference (1st Collision - Baseline)	Headset-2 vs. Desktop	-4.19 (9.52)	1.70 (7.11)	106.00	-1.43	0.159
	Headset-2 vs. Headset-1		8.14 (15.40)	53.00	-2.24	0.025*
Mean HR % Change (Test vs. Baseline)	Headset-2 vs. Desktop	-3.03 (8.65)	-0.75 (7.13)	149.00	-1.00	0.329
	Headset-2 vs. Headset-1		7.26 (13.19)	66.00	-2.52	0.011*
Mean HR % Change (1st Collision vs. Baseline)	Headset-2 vs. Desktop	-4.32 (10.59)	2.56 (10.11)	106.00	-1.43	0.159
	Headset-2 vs. Headset-1		10.62 (19.37)	52.00	-2.28	0.022*
Mean IBI - Baseline	Headset-2 vs. Desktop	0.80 (0.13)	0.83 (0.12)	169.00	-0.43	0.682
	Headset-2 vs. Headset-1		0.77 (0.11)	112.00	-0.86	0.402
Mean EDA - Baseline	Headset-2 vs. Desktop	2.30 (4.22)	0.65 (0.93)	126.00	-1.66	0.101
	Headset-2 vs. Headset-1		0.45 (0.53)	79.00	-2.05	0.041*
Mean EDA - Test	Headset-2 vs. Desktop	5.00 (6.94)	0.85 (1.24)	94.00	-2.57	0.009*
	Headset-2 vs. Headset-1		0.90 (1.27)	68.00	-2.45	0.014*
Mean EDA - 1st Collision	Headset-2 vs. Desktop	5.13 (7.75)	0.70 (0.86)	80.00	-2.29	0.022*
	Headset-2 vs. Headset-1		0.92 (1.36)	58.00	-2.02	0.045*
Mean EDA - Difference (Test - Baseline)	Headset-2 vs. Desktop	2.71 (4.75)	0.20 (1.01)	72.00	-3.20	0.001*
	Headset-2 vs. Headset-1		0.45 (0.97)	81.00	-1.98	0.049*
Mean EDA - Difference (1st Collision - Baseline)	Headset-2 vs. Desktop	2.49 (5.40)	0.04 (0.57)	78.00	-2.35	0.018*
	Headset-2 vs. Headset-1		0.45 (1.10)	77.00	-1.18	0.249
Mean EDA % Change (Test vs. Baseline)	Headset-2 vs. Desktop	238.14 (357.75)	53.68 (150.09)	101.00	-2.37	0.017*
	Headset-2 vs. Headset-1		138.01 (172.04)	123.00	-0.47	0.657
Mean EDA % Change (1st Collision vs. Baseline)	Headset-2 vs. Desktop	220.55 (363.59)	31.44 (81.51)	82.00	-2.22	0.026*
	Headset-2 vs. Headset-1		122.70 (167.14)	98.00	-0.26	0.812

*= significant difference ($p < .05$).

Heart Rate

Significant differences were observed in HR change scores from test to baseline ($H = 8.63$, $p = .013$, $\eta^2 = 0.12$), see Figure 4.25. Post hoc comparisons indicated a significantly greater HR increase in Headset-1 compared to Headset-2, suggesting enhanced cardiovascular activation under high-jerk motion. This result aligns with higher SSQ scores previously reported in Headset-1 and may reflect increased sensory conflict and vestibular stimulation. No significant differences were found between Headset-2 and Desktop. Similar trends were observed in Study 1, reinforcing that Headset-1 consistently elicits stronger sympathetic activation in high-motion conditions.

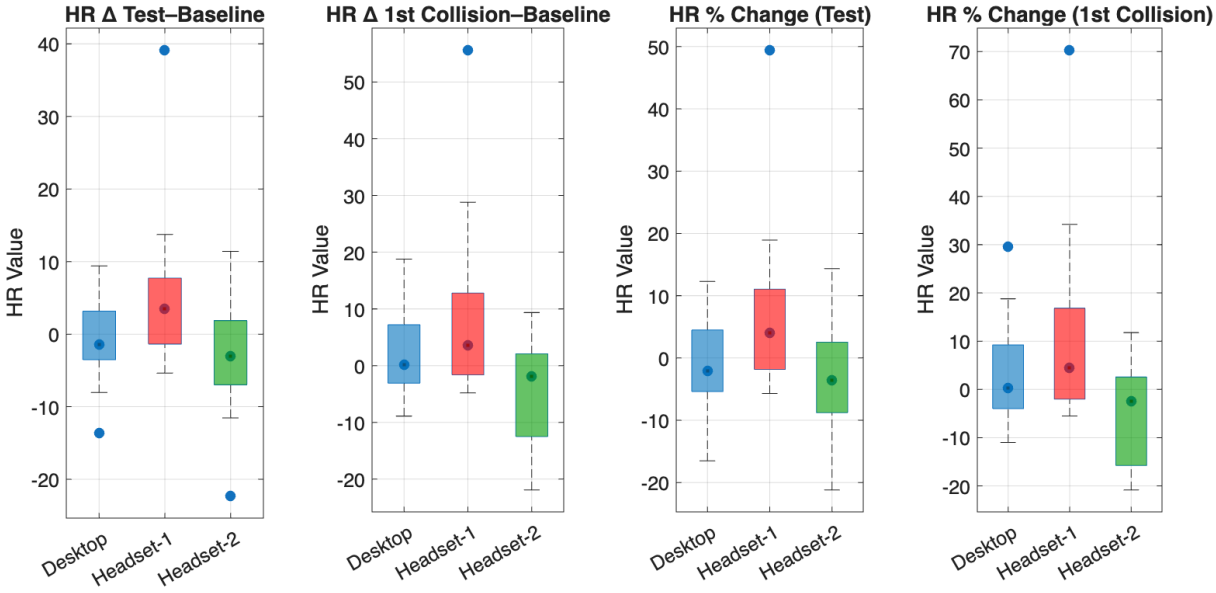


Figure 4.25: Study 2 heart rate (HR) differences across simulator conditions. Headset-1 elicited significantly higher cardiovascular activation than headset-2.

Inter-beat Interval and Heart Rate Variability

Although no significant group differences in IBI were found, Headset-1 showed a trend toward lower IBI, consistent with elevated HR and sympathetic arousal. The coherence between HR and IBI findings supports the interpretation that Headset-1 induced higher physiological demand. The non-significant p-values may reflect variability across participants or methodological factors, such as short task duration or signal artefacts. Further analysis with larger samples or additional IBI-derived metrics (e.g., HRV) could offer deeper insights into cardiac autonomic responses in immersive simulations.

Electrodermal Activity

Significant group differences were observed in multiple EDA metrics, particularly during the test condition and in change scores from baseline (Figure 4.26). The Mean EDA during the test condition

differed significantly between groups (H = 8.24, p = .016, η^2 = 0.12). Post hoc comparisons showed that Headset-2 elicited significantly greater EDA responses than both Desktop and Headset-1 (p < .017, Bonferroni-corrected). No significant differences were observed between Desktop and Headset-1.

These findings indicate that the low-jerk immersive configuration (Headset-2) was associated with increased sympathetic activation, possibly reflecting heightened arousal during the virtual navigation tasks. It suggests that a more comfortable and predictable immersive experience (Headset-2) leads to a more pronounced physiological response, likely related to engagement and alertness, without the confounding effects of discomfort or sensory strain.

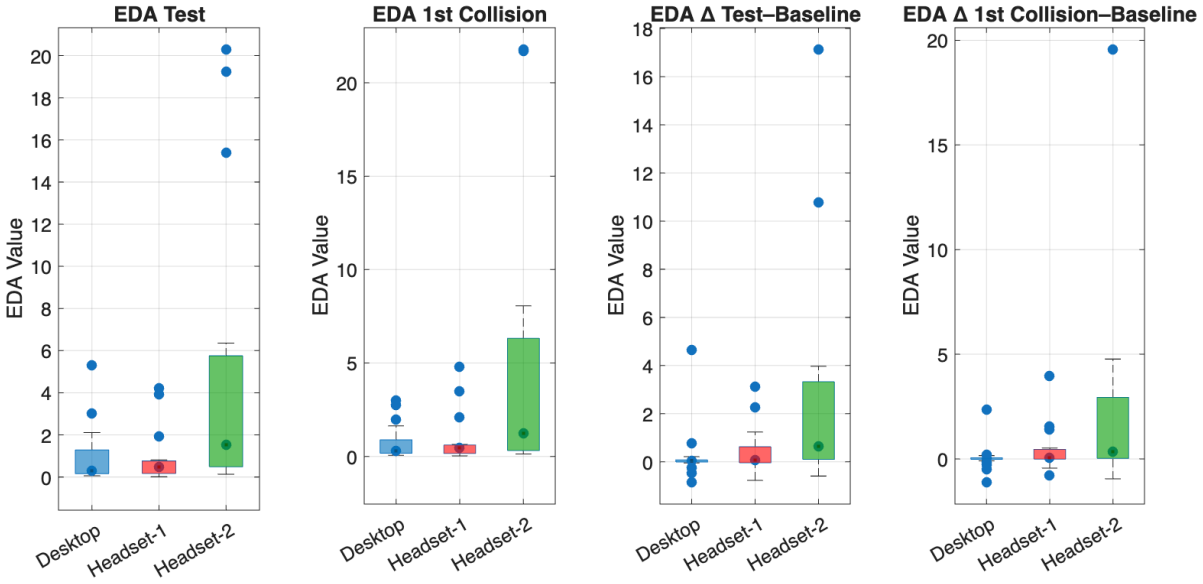


Figure 4.26: Study 2 electrodermal activity differences across simulator conditions. Headset-2 showed significantly higher sympathetic activation than both Desktop and Headset-1.

In terms of EDA change scores from baseline, a significant group effect was also found (H = 10.63, p = .005, η^2 = 0.16). Post hoc comparisons indicated that Headset-2 participants exhibited significantly higher EDA increases compared to Desktop group but not compared to Headset-1. These trends reinforce the interpretation that smooth immersive motion (Headset-2) produces elevated autonomic responses, while Headset-1 responses may be dampened due to competing effects of cybersickness or discomfort is supported by the presented data.

For example, The SSQ results showed that the Headset-1 group had significantly higher Oculomotor symptoms (p=.028) and a trend toward higher Disorientation symptoms (p=.053) compared to Headset-2. The EDA data, therefore, can be interpreted in light of these symptoms. In Headset-1, the high-jerk motion caused sensory overload and conflict, which manifested as higher heart rate and

reported cybersickness but possibly dampened the general sympathetic response measured by EDA, as the body was dealing with competing physiological demands.

EDA at the first collision event showed a marginal overall effect ($H = 6.02$, $p = .049$), but no significant post hoc differences emerged. Similarly, EDA change scores at the collision moment were not statistically different between conditions, although Headset-2 displayed descriptively higher values in both conditions.

These findings suggest that Headset-2 (low jerk, low cybersickness) was associated with the highest electrodermal activation, indicative of elevated physiological arousal. However, as EDA is a non-specific marker of sympathetic activity, this increase could reflect either heightened emotional engagement or stress-related arousal (Ahmad & Khan, 2022; Giannakakis et al., 2022; Kleiman et al., 2021; Wilhelm et al., 2006). In this case, the elevated EDA in Headset-2 occurred alongside lower SSQ scores and higher SAM arousal ratings, supporting the interpretation that participants experienced greater engagement (heightened emotional arousal) with less discomfort.

In contrast, Headset-1 did not elicit significantly elevated EDA despite higher heart rate and SSQ scores, possibly indicating a reduced emotional response due to discomfort or sensory overload. These results underscore the importance of interpreting EDA within a broader multidimensional context. Overall, the findings suggest that smoother motion profiles, as implemented in Headset-2, may help optimize both user comfort and positive arousal, making them preferable in immersive simulator design.

Summary

Overall, these physiological analysis results suggests that Headset-1 elicited stronger cardiovascular responses likely due to higher sensory conflict and cybersickness, while Headset-2 induced greater electrodermal activation, consistent with higher reported (positive) arousal in SAM ratings and lower discomfort (SSQ).

However, interpreting physiological signals in isolation remains challenging, as physiological arousal can arise from both stress and positive arousal engagement, and signal quality may be affected by movement artifacts, short recording windows, or individual variability ((Ahmad & Khan, 2022; Giannakakis et al., 2022; Kleiman et al., 2021; Wilhelm et al., 2006).The absence of significant group differences in IBI likely reflects such methodological limitations and further highlights the need for robust HRV-derived metrics and multimodal validation.

Nonetheless, the triangulation with self-report data strengthens the interpretation. Participants in the Headset-1 group consistently reported higher cognitive load (NASA-TLX), lower arousal (SAM), and

elevated cybersickness (SSQ), a pattern that aligns with HR findings. In contrast, Headset-2 was associated with higher (SAM) arousal and significantly stronger EDA responses, suggesting a more immersive and positively arousing experience. Despite limitations, the alignment between self-reported and physiological responses provides insight into how different motion designs shape perceived workload, emotional engagement, and physical comfort in each simulator condition.

Importantly, the use of physiological data to assess affective engagement and demand remains underexplored in power wheelchair simulator research. An exception is the work of (Zorzi et al., 2023a), which analysed heart rate changes across simulator tasks and identified a statistically significant HR increase during the backward slalom task. This task not only elicited the highest cardiovascular demand but also corresponded with higher involvement and skill improvement, suggesting a potential link between physiological activation, task engagement, and learning outcomes. Such findings reinforce the relevance of physiological measures for capturing psychophysical load in virtual training tasks.

In summary, these results highlight the importance of motion profile design in shaping user experience. Smoother motion appears to support more comfortable yet immersive simulator interactions, whereas high-jerk motion may undermine comfort through physiological and perceptual overload. While further studies with larger samples and enhanced signal processing are needed, the current findings demonstrate the value of combining subjective and physiological data to assess user response in VR-based simulator environments.

4.2.3.5.2 Behavioural Markers - Head Movements

To examine behavioural differences across simulator configurations and their relationship with subjective experience, head orientation data were analysed across three rotational axes: pitch (X), yaw (Y), and roll (Z). The analysis included two key metrics for each axis: range of motion and mean angular velocity, computed over the entire simulation task. These indicators help interpret how visual engagement, immersion, and physical behaviours manifest during simulator use.

In immersive simulations, the three axes represent distinct movement types: pitch refers to vertical nodding (e.g., looking up or down), yaw represents horizontal head turns (e.g., looking left and right, or scanning), and roll describes side-to-side tilting (e.g., ear to shoulder). Understanding these movements can offer insights into user attention, environmental interaction (Somarathna et al., 2023), and even physical discomfort (Palmisano et al., 2024).

A Kruskal–Wallis H test revealed significant group effects for yaw range and mean yaw angular velocity ($p < .001$), with immersive conditions (Headset-1 and Headset-2) showing broader and faster

horizontal head movements than Desktop (This finding suggests that the immersive nature of the headset prompted a more active physical engagement with the virtual environment, a behaviour consistent with naturalistic environmental scanning (yaw movements) and a key differentiator from the constrained interaction typical of a desktop display.

Table 4.25). Post hoc Mann–Whitney U tests confirmed that Headset-2 had significantly greater yaw range ($p < .001$) and yaw velocity ($p = .003$) compared to Desktop (Table 4.26), suggesting greater head engagement and environmental scanning during immersive use. This finding suggests that the immersive nature of the headset prompted a more active physical engagement with the virtual environment, a behaviour consistent with naturalistic environmental scanning (yaw movements) and a key differentiator from the constrained interaction typical of a desktop display.

Table 4.25: Study 2 Kruskal–Wallis Test results for head movements metrics.

Metric	Group	Mean (SD)	Median [IQR]	H	p-value	Effect Size (η^2)	Post hoc
Pitch Range (rad)	Desktop	0.65 (0.42)	0.56 [0.30 – 0.84]	0.117	0.943	-0.03	ns
	Headset-1	0.62 (0.34)	0.63 [0.32 – 0.87]	0.117	0.943	-0.03	ns
	Headset-2	0.58 (0.33)	0.68 [0.28 – 0.86]	0.117	0.943	-0.03	ns
Yaw Range (rad)	Desktop	0.91 (0.37)	0.98 [0.62 – 1.21]	40.131	0.000	0.71	
	Headset-1	1.62 (0.05)	1.60 [1.58 – 1.64]	40.131	0.000	0.71	***
	Headset-2	1.63 (0.06)	1.61 [1.58 – 1.69]	40.131	0.000	0.71	*
Roll Range (rad)	Desktop	0.49 (0.41)	0.32 [0.19 – 0.70]	3.936	0.140	0.04	ns
	Headset-1	0.27 (0.16)	0.20 [0.13 – 0.39]	3.936	0.140	0.04	ns
	Headset-2	0.26 (0.17)	0.27 [0.10 – 0.33]	3.936	0.140	0.04	ns
Mean Angular Velocity Pitch (X) (rad/s)	Desktop	0.10 (0.06)	0.08 [0.05 – 0.12]	3.379	0.185	0.03	ns
	Headset-1	0.08 (0.06)	0.05 [0.04 – 0.09]	3.379	0.185	0.03	ns
	Headset-2	0.08 (0.06)	0.06 [0.03 – 0.12]	3.379	0.185	0.03	ns
Mean Angular Velocity Yall (Y) (rad/s)	Desktop	0.10 (0.06)	0.08 [0.05 – 0.12]	16.894	0.000	0.28	
	Headset-1	0.18 (0.08)	0.15 [0.12 – 0.24]	16.894	0.000	0.28	***
	Headset-2	0.18 (0.10)	0.15 [0.10 – 0.22]	16.894	0.000	0.28	*
Mean Angular Velocity Roll (Z) (rad/s)	Desktop	0.09 (0.08)	0.06 [0.04 – 0.11]	1.609	0.447	-0.01	ns
	Headset-1	0.07 (0.05)	0.05 [0.03 – 0.11]	1.609	0.447	-0.01	ns
	Headset-2	0.06 (0.05)	0.05 [0.02 – 0.10]	1.609	0.447	-0.01	ns

ns= not significant,

*= significant difference Headset-2 vs. Desktop ($p < .0167$),

**= significant difference Headset-2 vs. Headset-1 ($p < .0167$).

***=significant difference (headset-1 vs. Desktop ($p < .0167$).

Table 4.26: Study 2 Mann–Whitney U Comparisons for Head Movements

		Headset-2	Compared Group	Test Statistics			
Metric	Group Comparison	Mean (SD)	Mean (SD)	U	Z	p-value	Effect size (r)
Head Pitch Range (rad)	Headset-2 vs. Desktop	0.58 (0.33)	0.65 (0.42)	178.00	-0.17	0.877	0.027
	Headset-2 vs. Headset-1		0.62 (0.34)	126.50	-0.34	0.736	0.059
Head Yaw Range (rad)	Headset-2 vs. Desktop	1.63 (0.06)	0.91 (0.37)	0.00	-5.26	0.000*	0.842
	Headset-2 vs. Headset-1		1.62 (0.05)	124.00	-0.44	0.683	0.077
Head Roll Range (rad)	Headset-2 vs. Desktop	0.26 (0.17)	0.49 (0.41)	124.50	-1.70	0.090	0.272
	Headset-2 vs. Headset-1		0.27 (0.16)	123.50	-0.45	0.657	0.078
Mean Angular Velocity X (rad/s)	Headset-2 vs. Desktop	0.08 (0.06)	0.10 (0.06)	147.00	-1.06	0.301	0.17
	Headset-2 vs. Headset-1		0.08 (0.06)	134.00	-0.07	0.958	0.012
Mean Angular Velocity Y (rad/s)	Headset-2 vs. Desktop	0.18 (0.10)	0.10 (0.06)	74.50	-3.13	0.001*	0.501
	Headset-2 vs. Headset-1		0.18 (0.08)	122.00	-0.51	0.631	-0.089
Mean Angular Velocity Z (rad/s)	Headset-2 vs. Desktop	0.06 (0.05)	0.09 (0.08)	144.00	-1.15	0.263	-0.184
	Headset-2 vs. Headset-1		0.07 (0.05)	123.50	-0.45	0.657	-0.078

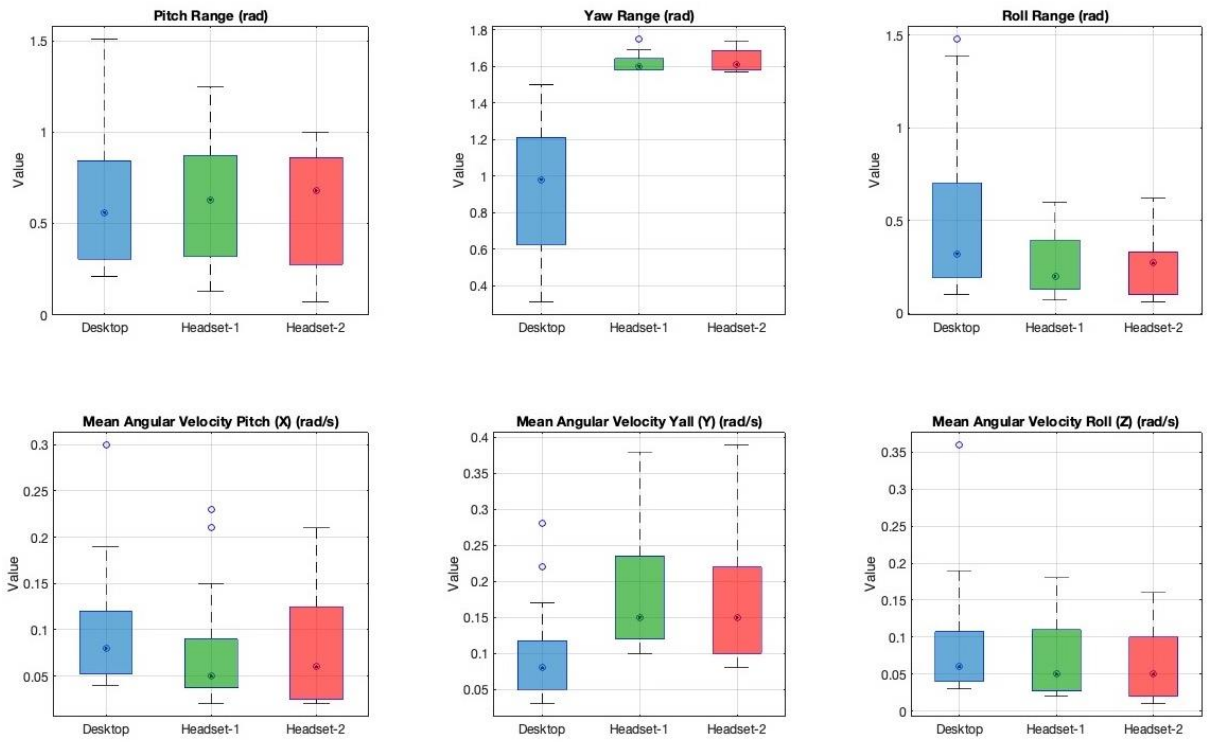


Figure 4.27: Study 2 boxplots of head movement metrics by group. Top: range of motion (rad) for pitch, yaw, and roll. Bottom: mean angular velocity (rad/s). Immersive conditions, especially Headset-2, showed greater yaw movement compared to Desktop.

These findings are visually summarized in Figure 4.27, where boxplots highlight the elevated yaw motion in immersive groups, particularly Headset-2. While pitch and roll movements showed less variation across conditions, yaw emerged as the most sensitive axis to changes in display modality. The broader and more dynamic yaw behaviour likely reflects increased environmental scanning and visual engagement in immersive settings.

Exploratory Correlation Analysis between Head Movements and SSQ and NASA-TLX

In addition to group comparisons, Spearman's rank-order correlations were calculated to examine associations between head movement metrics and subjective measures SSQ and NASA-TLX workload scores. No significant correlations were found between head movement behaviours and any SSQ subscale (nausea, oculomotor, disorientation) or total SSQ score ($p > .05$), suggesting that head movement was not a reliable indicator of cybersickness in this sample. However, this contrasts with prior findings from (Vailland et al., 2021), who reported positive correlations between head acceleration/velocity and SSQ scores, indicating a need for further investigation across different simulator designs and populations.

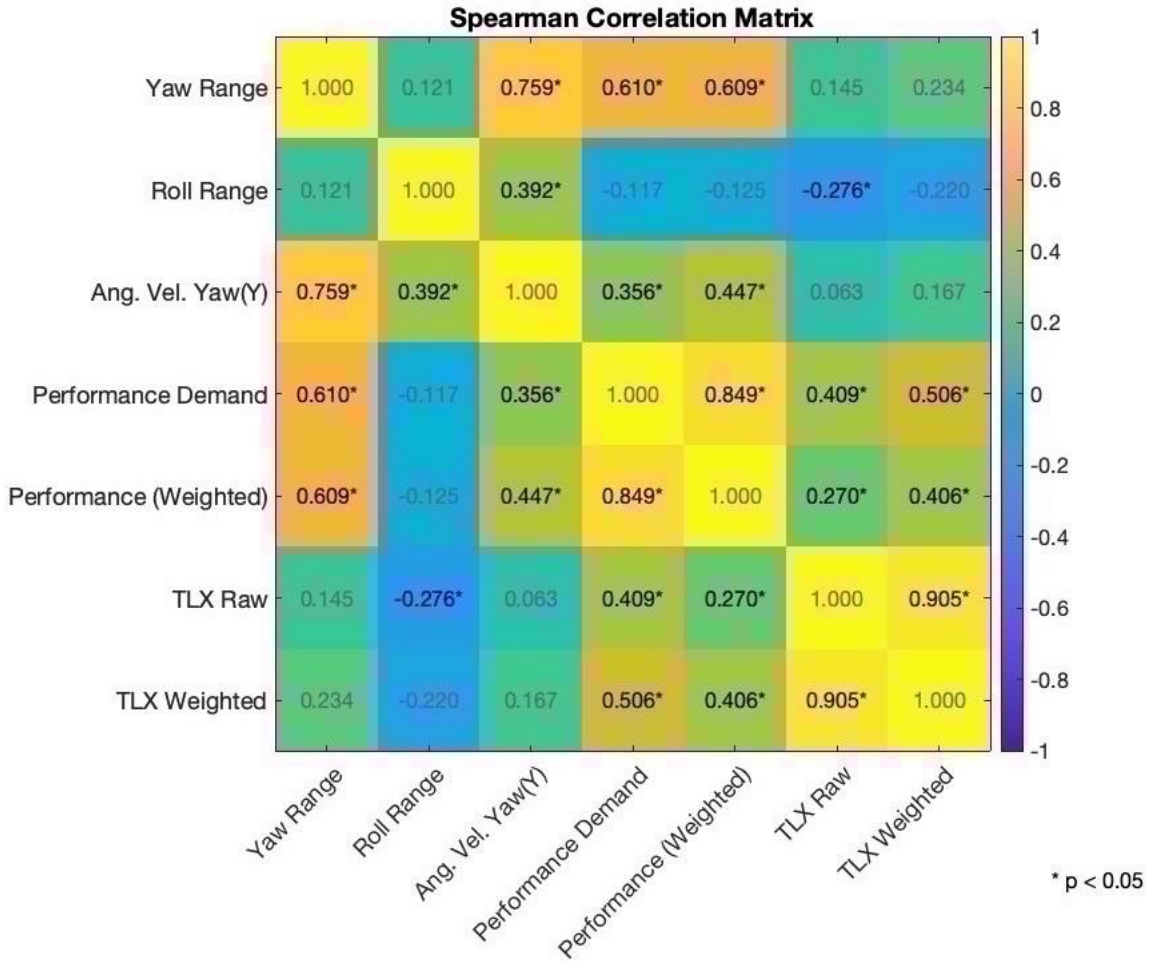


Figure 4.28: Study 2 correlation matrix between head movement metrics and subjective cognitive workload measures.

In the present study, significant associations were found between head movement metrics and perceived cognitive workload. Head yaw range was positively correlated with both performance demand (ρ=.609,p<.001) and weighted performance score (ρ=.610,p<.001). Similarly, mean angular velocity Yaw showed a weak positive correlation with performance demand (ρ=.356,p<.007) but a moderate positive correlation with weighted performance score (ρ=.447,p<.001). These results indicate that participants exhibiting broader or faster horizontal head movements also perceived the task as more demanding in terms of performance, potentially reflecting increased attentional engagement or effort.

Additionally, a statistically significant negative correlation was found between head roll range and the raw TLX score (ρ=–.276,p=.039). This means that the correlation is not due to chance, and there is an inverse relationship between the two variables: as lateral tilting of the head increased, overall workload ratings tended to decrease. These findings suggest that increased lateral tilting of the head

was modestly associated with lower overall workload ratings. In other words, this finding is particularly interesting because it hints at a potential compensatory behaviour. It's possible that this type of head movement, though not directly related to environmental scanning (which is linked to yaw), serves a different purpose. One hypothesis is that it could be a subtle sign of relaxation or comfort in the virtual environment, or it might be a subconscious action taken to reduce strain and, consequently, perceived workload. This is in contrast to the increased yaw movements, which were associated with higher perceived performance demands. The negative correlation with head roll, therefore, offers a unique insight into how different head movements may relate to different aspects of the user's cognitive state and experience.

A visual summary of these associations is presented in Figure 4.28, which shows the Spearman correlation matrix between head movement metrics and subjective cognitive workload measures. Statistically significant correlations are marked with an asterisk (*), indicating $p < .05$.

Summary

Overall, the findings indicate that immersive simulator conditions elicited greater head movement activity, particularly along the yaw axis, which likely reflects enhanced visual scanning and spatial interaction. These yaw movements were meaningfully associated with perceived cognitive workload, especially in relation to performance demand but showed no link to simulator-induced discomfort (cybersickness). This contrasts with some prior findings, such as (Vailland et al., 2021) highlighting the need to consider context-specific factors. The results suggest that head yaw behaviour may serve as a non-intrusive behavioural marker of cognitive workload in wheelchair simulation environments, offering potential value for integration into multidimensional evaluation frameworks. However, further research is required to validate its reliability and interpretability across different tasks and user groups.

4.2.4 Study 2: Discussion

Study 2 evaluated how display modality and motion dynamics influence user experience within a virtual wheelchair simulator. By introducing a third condition (Headset-2) with a low-jerk motion profile, this study extended the findings of Study 1 and addressed Sub-Research Question 1 (SRQ1) and Objectives 1.2 and 1.3, with a focus on usability, cognitive workload, emotional response, simulator tolerance, and performance.

Revisiting the Hypotheses

H1: Participants in the Headset 2 condition (immersive, low-jerk motion) will report higher QoE scores-covering usability (SUS), emotional response (SAM), sense of presence (IPQ), and lower cognitive workload (NASA-TLX) as well as lower simulator sickness (SSQ) compared to Headset 1 (immersive, high-jerk) condition.

This hypothesis was partially supported. Participants using Headset-2 rated the simulator significantly more usable and immersive than the Desktop condition, but differences between Headset-2 and Headset-1 in presence and usability scores did not reach statistical significance after Bonferroni correction. Thus, the benefits of immersion were largely attributed to the transition from non-immersive to immersive display, rather than motion dynamics alone.

Headset-2 participants reported significantly higher arousal (SAM) than both Headset-1 and Desktop, consistent with elevated electrodermal activity (EDA), suggesting positive arousal without additional discomfort. In contrast, Headset-1 triggered higher simulator sickness symptoms (SSQ) and stronger cardiovascular responses (HR), particularly in oculomotor and disorientation domains. Although overall workload (NASA-TLX) differences between Headset-2 and Headset-1 were not statistically significant, trends indicated reduced cognitive effort in the low-jerk condition. These findings support the conclusion that smoother motion might enhance comfort and affective experience, but improvements across all QoE dimensions were not uniform.

H2: Physiological responses (EDA, HR, IBI) will significantly differ across the simulator configurations, with expected differences between the Headset 2 condition and both the Desktop and Headset 1 groups. These variations will reflect how display modality and motion profile influence users' autonomic responses during simulator use.

This hypothesis was supported. Headset-2 participants showed significantly higher EDA compared to other conditions, consistent with arousal and immersion. Headset-1, by contrast, showed elevated HR, indicating greater physical or sensory strain. IBI metrics did not yield significant results, likely due to signal quality limitations.

These results align with prior findings in immersive research. Notably, (Zorzi et al., 2023a) reported heart rate variations across virtual wheelchair tasks, showing higher psychophysical load during more demanding navigational tasks. In that context, increased HR were associated with higher task engagement, as reflected in the backward slalom task performance. The present study adds to this growing body of work by demonstrating that changing the motion profile itself can alter physiological responses during immersive wheelchair simulation.

While physiological data offered valuable insights into autonomic responses, their interpretation in isolation remains challenging due to the complexity of overlapping emotional, cognitive, and physical states. For instance, elevated heart rate may reflect either heightened arousal or discomfort, and electrodermal activity can signal both engagement and stress.

In this study, triangulating physiological findings with self-reported assessments (e.g., SSQ, SAM, and NASA-TLX) provided essential contextual grounding. For example, the elevated EDA observed in Headset-2 participants was accompanied by higher SAM arousal scores and lower SSQ symptoms, suggesting that the sympathetic activation reflected positive engagement rather than distress.

In contrast, the increased HR in Headset-1 aligned with higher SSQ scores, indicating discomfort rather than beneficial stimulation. This multi-source interpretation was critical for discerning the nature of physiological responses, distinguishing between positive arousal and simulator-induced strain, and reinforces the importance of using combined subjective and objective measures in simulator evaluation.

4.2.5 Key Findings and Methodological Lessons from Study 2

Virtual Motion Dynamics Should be Tailored to User Needs

Headset-2, with low-jerk motion and immersive display, improved usability and positive emotional arousal compared to the Desktop condition. Comparisons between Headset-1 and Headset-2, differing only in motion, indicated that reduced jerk supported better comfort and engagement. These effects underscore the importance of configurable system parameters in simulator design, particularly in clinical contexts.

Increasing Sense of Presence Alone is Not Universally Beneficial

While immersive displays enhanced presence and engagement, they also introduced greater variability in user tolerance. Headset-1 led to significantly higher simulator sickness symptoms despite offering immersive visuals. To reduce discomfort and support broader usability, the non-immersive desktop configuration was selected as preferred option for the field study.

Multimodal Assessment Strengthens User's Quality of Experience Evaluations

The use of multiple data sources, including subjective questionnaires (NASA-TLX, SUS, IPQ), physiological signals (EDA, HR), and behavioural metrics (head movements), enabled a more nuanced interpretation of user experience in Study 2. While not all patterns were consistent across measures, alignment between self-reported arousal and physiological indicators (e.g., EDA, HR) provided converging evidence in key comparisons. This approach helped to contextualize physiological responses and reduce the risk of misinterpretation when relying on a single measure. Elements of this mixed-methods strategy were carried forward into the field study, where implicit data offered additional insight into user variability during simulator use.

Head Movement Metrics Offered Preliminary Behavioural Insights

Exploratory analyses revealed that Yaw range and velocity correlated with self-reported performance demand, while roll range demonstrated a weak negative correlation with perceived mental workload. These findings suggest that head movement patterns may serve as behavioural markers of cognitive effort. Consequently, head tracking was retained in the field protocol, and wrist-worn sensors were added to capture additional behavioural interactions.

Informing Field Study Design and Technical Adjustments

Findings from Study 2 informed several refinements to the simulator system in preparation for its deployment in community-based settings. These refinements include:

- Expanded Camera Views: To allow user preferences, it was added multiple auxiliary camera views (e.g., third-person, top-down) to supplement the primary first-person perspective. This approach provides users with flexible navigation options and helps increase the surroundings awareness.
- User-Controlled Settings: Adding an user interface (UI) to allow for real-time adjustments of speed, acceleration, and deceleration. This feature allows the participants to personalize the experience to their specific comfort and preferences.
- Customizable Haptics and Audio: We also added the ability to turn haptic feedback and background sound on or off via the UI. This provides critical sensory customization, making the system more inclusive for individuals who may find these stimuli distracting or overstimulating.
- Physical Hardware Adaptations: The joystick hardware was also refined to include different customizable knobs to accommodate varying hand sizes and dexterity levels, further improving the system's physical accessibility.

Participant feedback and comparative outcomes across different display modalities and motion profiles emphasised the importance of adaptability to support user comfort and reduce simulator-induced discomfort. These insights led to the implementation of configurable motion settings and the decision to adopt a non-immersive (desktop) display for the field study, prioritizing accessibility and broad usability over maximal immersion.

Additionally, the mixed-methods approach applied in Study 2, combining subjective, physiological, and behavioural measures, provided practical guidance for data collection strategies in more variable, community-based contexts. Table 4.27 below summarises the key findings and their methodological impact on the field deployment protocol.

Table 4.27: Key Findings from Study 2 and Their Methodological Impact

Area	Key Finding	Methodological Impact
Display and Motion Interaction	Combined effects of immersive display and low-jerk motion influenced usability, arousal, and cybersickness	Informed inclusion of adjustable motion settings; desktop configuration is preferred option for field use.
Immersion and Tolerance	Immersive displays increased presence and arousal but also contributed to higher simulator sickness in some users	Supported decision to prioritize broad usability by selecting a non-immersive display
Multimodal Assessment Approach	Partial convergence across self-report (SUS, IPQ, NASA-TLX) and physiological data (EDA, HR)	Justified continued use of an integrated QoE framework to capture diverse user responses in the field
Behavioural Cues (Head Movements)	Yaw range and speed correlated with perceived performance demand; roll modestly linked to workload ratings	Head tracking retained to explore behavioural indicators of cognitive effort within a fixed display setup
System Design Refinement	Findings highlighted the need for varied task types and strategies to support spatial awareness in desktop-based setups	Informed inclusion of multiple task scenarios and optional software-based viewpoint adjustments to improve spatial context

4.2.6 Limitations

Despite its contributions, Study 2 has several limitations. First, the participant sample consisted exclusively of healthy adults, which limits the generalizability of findings to clinical populations, such as individuals with mobility impairments or cognitive challenges. The use of convenience sampling and the exclusion of end-users from rehabilitation contexts mean that results should be interpreted with caution when considering broader applications.

Second, although physiological signals were successfully collected and processed, the analysis focused on time-domain features only. The absence of frequency-domain analyses (e.g., HRV spectral

components or EDA phasic response rates) may have limited the depth of insight into autonomic regulation during simulator use.

Third, the between-subjects design may have introduced uncontrolled variability due to differences in participants' prior experience with virtual reality technology, despite random assignment to conditions.

Finally, while the implementation of a low-jerk motion profile was a key methodological advancement, other simulator features, such as sound design, avatar embodiment, or training aids, were held constant. These elements may also influence user experience and should be considered in future research to enhance immersion, personalization, and user-centred design.

4.2.7 Study 2: Contribution to SRQ1

The findings informed Contribution 2, revealing that smoother acceleration profiles (low jerk) improved arousal and reduced simulator sickness symptoms, with significant differences user responses between groups. Also, Study 2 provided substantial evidence to support and extend the response to Sub-Research Question 1 (SRQ1):

> ***"How can a virtual wheelchair simulator be designed and tested in a controlled environment to establish a clinically relevant proof of concept that supports multidimensional assessment, incorporating immersive technologies, physiological signals, subjective feedback, and Quality of Experience (QoE) evaluation?"***

This question was investigated through two controlled studies. Study 1 primarily addressed Objective 1.1 by demonstrating the feasibility of combining subjective Quality of Experience (QoE) ratings with physiological signals to assess user interaction. Study 2 extended this foundation by addressing Objectives 1.2 and 1.3, examining how immersive display and motion dynamics influence usability, workload, emotional response, and simulator tolerance.

Findings from Study 2 showed that smoother motion (low-jerk) improved comfort and positive arousal while reducing physiological stress compared to high-jerk conditions. Immersive displays enhanced presence and usability, though their benefits depended on motion settings optimized for user tolerance. The use of integrated assessment combining subjective, physiological, and behavioural data enabled a more comprehensive understanding of user experience across conditions.

Together, the results of both studies (1 and 2) fulfilled Objective 1.4 by informing the design of an initial evaluation framework for clinical use. This framework incorporates tailored simulator configurations, QoE-based assessment methods, and design considerations necessary for

transitioning to real-world application, as implemented in the field study described in the next chapter.

4.3 Summary

This chapter presented two laboratory-based studies that collectively addressed Sub-Research Question 1 (SRQ1). Study 1 focused on foundational system evaluation, demonstrating the feasibility of integrating subjective, physiological and behavioural measures to assess user experience across different display types. It revealed initial differences in user tolerance and affective response, which informed refinements in motion design and justified the addition of cybersickness assessment.

Building on these insights, Study 2 introduced a third experimental condition to investigate the interaction between display modality and motion dynamics in. headset conditions. The results indicated that both factors significantly influenced usability, emotional response, simulator tolerance, and physiological arousal. Smoother acceleration profiles reduced discomfort and stress, while immersive displays enhanced presence and usability when paired with motion settings tuned for user comfort.

Together, these studies established a structured approach for evaluating wheelchair simulators in controlled environments. They provided a testing protocol, a multidimensional set of user experience metrics, and evidence-based design guidance for simulator configuration. These contributions define the methodological foundation for transitioning from lab-based investigation to real-world implementation. The next chapter presents how this evaluation framework was applied and further examined in a field study involving end users.

Chapter 5 Field-based Study: Powered Wheelchair Simulator Pilot Feasibility Study

5.1.1 Introduction

This chapter presents the field-based evaluation of the virtual wheelchair simulator, assessing its feasibility, usability, and preliminary clinical relevance within real-world assistive settings. This phase of the research bridges the gap between the controlled laboratory experiments (Chapter 4) and the practical realities of clinical deployment.

The lab studies revealed a critical challenge: while the high-jerk immersive configuration (Headset-1) significantly increased physiological arousal and cybersickness, the low-jerk profile (Headset-2) still presented usability challenges. This highlighted a key decision for clinical application, the simulator design must prioritize user comfort and ease of use to ensure feasibility and viability.

To address this, It was important to collaborated with the Irish Wheelchair Association (IWA) and a panel of four healthcare professionals, including a wheelchair user. The feedback from these domain experts was instrumental in guiding the transition from a research-focused lab environment to a practical field setting. Their recommendations included the use of a non-immersive setup over the potentially discomforting VR headsets. Furthermore, they advised against a single, continuous route (e.g. circuit with a mix of ramps, turns around cones, etc.), suggesting instead that the PMRT/WST-inspired tasks be broken down into discrete, more manageable components to facilitate targeted skill evaluation. They also emphasized the importance of giving users the choice to select their preferred motion profile (e.g., acceleration/deceleration and target speed), acknowledging that perceptions of high- and low-jerk motion could vary and impact comfort even in a non-immersive environment.

These expert-informed adjustments created a clear developmental narrative, where the scientific data from the lab studies on immersive discomfort and motion dynamics directly informed the design of a more clinically feasible system. Consequently, this investigation was designed as a pilot feasibility study to inform future intervention studies.

The study, conducted in partnership with the IWA, utilized two of its community centres as implementation sites. By embedding the evaluation within a familiar and service-integrated environment, this research aimed to enhance ecological validity by reflecting authentic usage scenarios. The study employed a multidimensional framework that integrated established clinical tools (WST and MoCA) with self-report, physiological, and behavioural metrics.

The study's sample was drawn from participants who regularly visited two specific community-based locations for power electric wheelchair users. This sampling method introduces potential self-selection and location biases, as the participants may be more active or have different demographic profiles than those who do not attend such centres. Therefore, the presented findings must be interpreted with caution and may not be generalizable to all powered wheelchair users.

This stage of the research addresses Sub-Research Question 2 (SRQ2):

> ***"How can the proof-of-concept simulator be transferred into clinical settings, using Irish Wheelchair Association (IWA) centres as a use case, and how can protocols and evaluation methods be developed to test the feasibility its components, reflect the perspectives of wheelchair users, and support standardised implementation?"***

To answer this question, the study was guided by the following objectives:

- **Objective 2.1**: Conduct a field pilot study to assess the feasibility and acceptability of the simulator within IWA centres.
- **Objective 2.2**: Define and analyse simulator-based metrics in relation to standard assessments (WST and MoCA) to explore alignment and potential complementarity.
- **Objective 2.3**: Develop preliminary guidelines for simulator use in future pilot and validation studies, informed by feasibility findings, comparative analysis, and user feedback.

Together, these objectives supported the transition from controlled laboratory testing to real-world field implementation, emphasizing clinical feasibility, exploratory metric alignment, and framework refinement. In particular, simulator sessions were embedded within existing IWA activities, including physiotherapy, mobility training, and leisure programs. This approach aimed to ensure ecological validity and maintain user engagement while minimizing disruption to centre routines.

This investigation contributes to the broader research goal of supporting safer and more effective wheelchair prescription and training practices using virtual reality-based simulation. Traditional clinical assessments for mobility and cognitive readiness often lack structured, repeatable mechanisms capable of integrating user experience with performance. Simulators offer a promising alternative, but their successful integration into practice requires evidence of feasibility, tolerability, and preliminary clinical relevance across diverse user populations and assessment tools.

This study supports two major contributions of the thesis:

- **Contribution 3:** The development of a mixed-methods pilot feasibility protocol that integrates simulator-based metrics with feedback from wheelchair users and clinicians, facilitating a structured application in clinical workflows.
- **Contribution 4:** The refinement of the QoE-based evaluation framework, a set of practical guidelines and methodological recommendations for integrating tools such as WST and MoCA within simulator-based evaluations.

This chapter is structured as follows. Section 5.2 describes the study aim and motivation. Section 5.3 presents the study setting and participant recruitment strategy. Section 5.4 outlines the simulator system used in the field implementation. Section 5.5 details participant recruitment and eligibility criteria. Section 5.6 presents the full study procedure, followed by Section 5.7 which outlines the simulator-based tasks. Section 5.8 describes the assessment tools employed, and Section 5.9 defines the primary and secondary outcomes. Section 5.10 explains the statistical analysis methods. Section 5.11 presents the results, including demographics, performance metrics, physiological data, and exploratory correlations with clinical assessments. Section 5.12 introduces the EMPOWER-SIM as a set of practical guidelines and methodological recommendations. Section 5.13 provides an integrated discussion of findings, limitations, and system-level design considerations. Finally, Section 5.14 summarises the key insights and implications from this field study.

5.1.2 Study 3: Aim and Motivation

This study aimed to evaluate the feasibility, usability, and preliminary clinical relevance of the virtual wheelchair simulator in real-world assistive settings. Building on the controlled lab experiments from Chapter 4, the simulator was deployed within the routine activities of wheelchair users at two IWA centres to assess its performance under naturalistic conditions and its potential preliminary alignment with established clinical tools.

The motivation for this field-based evaluation was to test the practicality and feasibility of the simulator in the context where it is intended to be used-embedded in user-centred rehabilitation services and adapted to diverse user needs. Unlike the lab environment, community settings introduced practical constraints and variability that are critical for testing ecological validity and workflow fit.

A secondary objective was to determine whether simulator-derived metrics (e.g., task completion time, collisions) show exploratory associations with standardised clinical assessments, WST and MoCA. These tools, while widely used, often rely on subjective interpretation; the simulator offers an

opportunity for more objective and repeatable assessment, though this study can only provide preliminary findings rather than definitive validation

Finally, this study contributed to refining the QoE-based evaluation framework, by examining its adaptability across users with varying cognitive and mobility profiles and gathering feedback on protocols, questionnaires, and sensor integration to inform future pilot and validation studies.

5.1.3 Study 3: Design and Setting

This field-based study followed a mixed-methods design, integrating both quantitative and qualitative approaches to evaluate the feasibility, usability, and preliminary clinical relevance of the virtual wheelchair simulator within assistive community settings.

The evaluation framework applied in this study was based on QoE assessment studies described in Chapter 4. This framework incorporates subjective assessments, objective performance data, and physiological metrics. It was further aligned with established clinical tools for assessing power mobility and cognitive function, specifically the WST and MoCA.

Primary feasibility endpoints included recruitment and eligibility rates, session and task completion rates, data completeness for questionnaires, simulator logs and physiological streams, setup and session duration, protocol deviations, and reports of tolerability/adverse events. Acceptability was treated as a dimension of QoE, captured through SUS scores, custom usability/acceptability questions, and participants' willingness to recommend the simulator for training or assessment.

Secondary exploratory endpoints included simulator-derived performance metrics (e.g., task duration, collision count, joystick control variability), self-report QoE questionnaires, and exploratory correlations with WST and MoCA. These were interpreted as preliminary signals of alignment rather than validation.

The methodology was structured into three stages:

- Pre-assessment (demographics, cognitive and mobility screening),
- Simulator session (task execution with physiological monitoring), and
- Post-assessment (questionnaires and feedback).

All procedures followed the ethical standards approved by the Technological University of the Shannon (TUS) Research Ethics Committee. Participants were provided with written and verbal information prior to signing informed consent. The study adhered to the Declaration of Helsinki and

the STARD 2015 guidelines (J. F. Cohen et al., 2016) for pilot diagnostic accuracy studies. All data were anonymised and securely stored in compliance with applicable data protection regulations.

5.1.4 Study 3: E-WATS – Powered Wheelchair Simulator System

The virtual wheelchair simulator used in this study is an updated version of E-WATS system as per Figure 5.1. This system combines two primary components: the wheelchair simulator (Figure 5.1 A) and devices for capturing user response data (Figure 5.1 B), which is controlled using a wheelchair joystick adapted for simulator use on a computer. It collects implicit physiological data through tools such as the smartwatch from Empatica (EmbracePlus and E4), HD1080p Logi Camera integrated with the OpenFace library, Mindband and OpenVibe Interface. Additionally, it gathers explicit self-report data and feedback from semi-structured interviews conducted with users interacting with the wheelchair simulator. Together, these data sources provide complementary insights into user experience and performance at a feasibility level, combining implicit physiological and behavioural measures with explicit self-report and interview feedback.

The E-WATS was updated after lab-based findings. The simulator has evolved to better assess power mobility and cognitive performance. Adjustments to task complexity, sensor integration, and evaluation algorithms have improved its ability to generate meaningful, interpretable metrics aligned with clinical benchmarks.

This study focuses on assessing the feasibility of capturing and exploring relationships between simulator-derived metrics, such as task completion time and error rates, and reference standards including the WST and MoCA. A control group of non-wheelchair users was included to provide a baseline for evaluating simulator sensitivity at an exploratory level, helping to determine whether performance differences may reflect prior mobility experience or cognitive variation.

Finally, semi-structured interviews conducted after simulator use provide qualitative insights that inform further system refinement. This iterative process supports the development of a simulator that is adaptable and potentially transferable to clinical contexts, while recognising that further validation studies are required before routine clinical use.

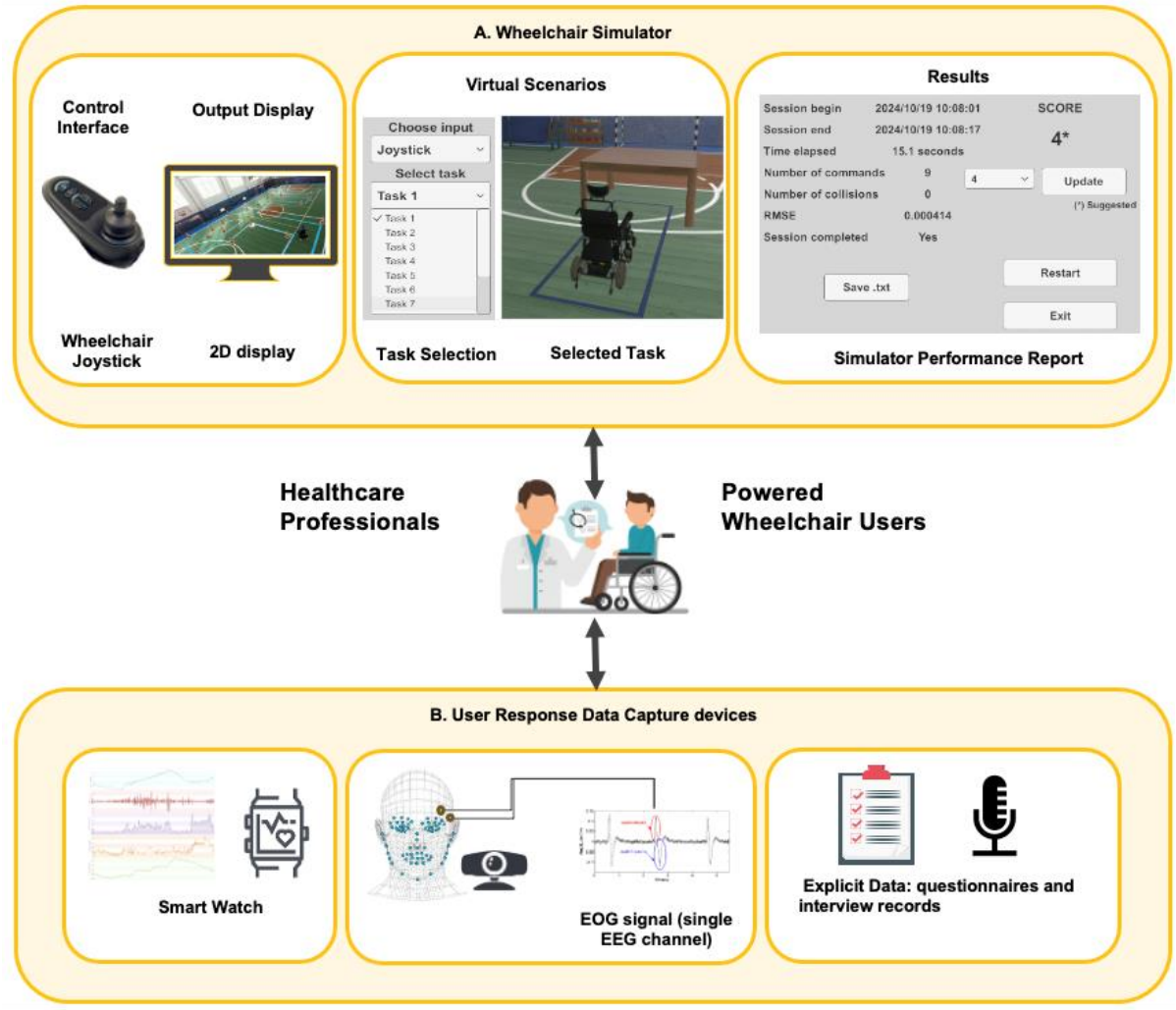


Figure 5.1: E-WATS: Wheelchair Simulator System. (A) Diagram of the simulator's core components, including the computer, virtual environment, and control interface (joystick controller). (B) Devices for capturing user responses and physiological data.

5.1.5 Participants

The recruitment and implementation took place at two community centres operated by TUS in Athlone and the IWA centres, located in Athlone and Cork, Ireland. These IWA's centres provide services and mobility training programmes to individuals with physical disabilities and were selected to ensure ecological validity by embedding the simulator in familiar and routine environments.

Simulator sessions coordinated with participants' weekly schedules and held in quiet, private rooms within the centres to provide a safe and controlled testing space. IWA staff facilitated participant pre-screening, scheduling, and consent, while all simulator setup, real-time monitoring, and data collection were managed by the primary investigator.

5.1.5.1 Participant Recruitment and Profile

The wheelchair user group consisted of individuals with diverse diagnoses, mobility needs, and levels of prior experience operating powered wheelchairs. Cognitive profiles also varied, as assessed through MoCA. The control group included individuals with no mobility impairments and no prior experience using powered mobility devices. This control group served as a baseline reference for exploratory sensitivity analyses, helping to assess whether simulator-derived metrics could distinguish between novice and experienced mobility groups at a feasibility level. The comparative design therefore provided preliminary insights into simulator performance, usability, and physiological responses across groups with distinct mobility backgrounds.

5.1.5.2 Inclusion and Exclusion Criteria

Participants were eligible for inclusion if they:

- Were 18 years of age or older.
- Had sufficient functional vision, hearing, and hand control to use the simulator.
- Were able to provide informed consent.
- For the wheelchair user group: had current or recent experience using a powered wheelchair.
- For the control group: had no self-reported physical or cognitive impairments.

Exclusion criteria included:

- Diagnosed neurological or musculoskeletal conditions affecting safe simulator use;
- Severe cognitive impairment (MoCA < 10).
- History of epilepsy, uncorrected visual impairment, or motion sickness that might be aggravated by simulator use.
- Prior participation in earlier simulator research conducted by the study team.

Participant screening was conducted in collaboration with IWA staff and included administration of the MoCA, WST questionnaire, and health review. All participants received a short joystick familiarization session prior to starting the simulator tasks.

5.1.6 Study 3 Procedure

The study followed a structured protocol (Figure 5.2) beginning with participant recruitment and informed consent. Participants were briefed on the study procedures, and demographic data, including age, gender, diagnosis, and prior exposure to simulators or virtual reality, were collected as outlined in the Appendix E.

Simulator sessions were integrated into the weekly routines at the IWA centres. Each participant completed a one-on-one session conducted in a private room, supervised by a trained researcher to ensure safety and task adherence. The simulation was displayed on a monitor and controlled using a powered wheelchair joystick mounted to a table or in the participant's wheelchair. Participants navigated through a series of tasks inspired by PMRT, including directional navigation, obstacle avoidance, and manoeuvring in confined spaces. All interactions were automatically logged for subsequent performance analysis.

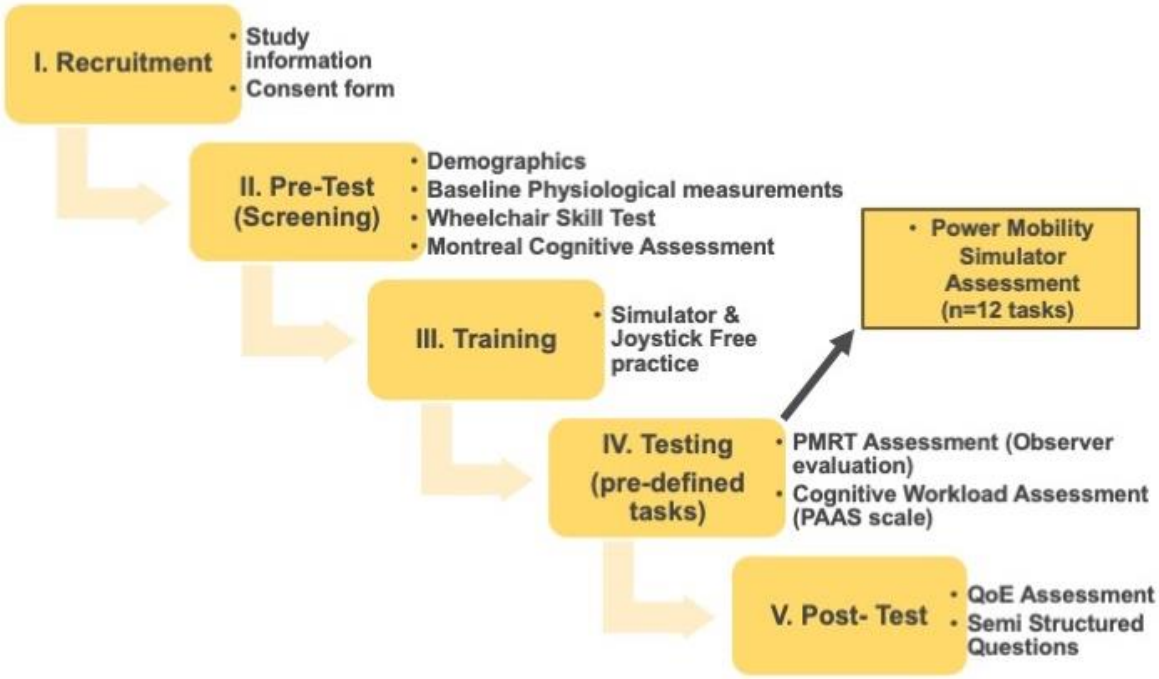


Figure 5.2: Field-based experiment protocol.

Prior to simulator use, participants completed a pre-assessment phase consisting of three key elements. MoCA was used to evaluate cognitive functioning, while WST-Q captured self-reported mobility skills in terms of performance and confidence. A five-minute resting baseline was also recorded using the Empatica E4 and EmbracePlus wristband, which monitored EDA, HR, and IBI to establish physiological and acceleration reference values.

During the simulation session, simulator-based performance metrics, including joystick input, total task time, task completion, number of collisions, and trajectory deviations, were continuously recorded. Simultaneously, physiological signals were monitored via the wearable devices. Immediately after completing the simulator tasks, participants rated their perceived cognitive workload using the Paas Cognitive Load Rating Scale.

In the post-assessment phase, participants completed a set of validated self-report questionnaires. SUS assessed interface ease of use; IPQ captured perceived immersion and spatial presence; SAM measured emotional state across valence, arousal, and dominance dimensions; and NASA-TLX provided a breakdown of perceived task load. Participants were also invited to share open-ended feedback on the simulator's usability, realism, and perceived challenges or improvements. All instruments were provided in either paper or digital format, based on participant preference.

5.1.7 Simulator-based Tasks

The virtual wheelchair simulator was configured to deliver a set of structured tasks designed to evaluate functional power mobility skills in alignment with clinical frameworks such as PMRT (Massengale et al., 2005) and WST-Q (Kirby et al., 2002). Each task simulated a real-world mobility challenge commonly encountered by powered wheelchair users and was selected based on its relevance to assessing directional control, spatial navigation, obstacle negotiation, and movement precision.

Tasks were presented sequentially within an immersive virtual environment and performed using a standard powered wheelchair joystick. The simulator recorded detailed performance metrics, including joystick activity, task duration, number of collisions, and deviations from optimal trajectories. To ensure alignment with clinical practice, the virtual tasks were adapted from standard PMRT and WST items, providing a controlled and repeatable simulation context. Task 9 involved a free-driving session with stop/start commands triggered via auditory cues. It was used before the main assessment to calibrate joystick parameters (e.g., acceleration and speed) based on individual user needs and to gather initial user feedback. The task was also strategically inserted mid-sequence to offer a short recovery period and reduce potential fatigue before more demanding tasks.

Task 9 was therefore excluded from the performance analysis. Table 5.1 summarises each simulator task and its mapping to the corresponding PMRT and WST-Q items. While all 26 items from the WST-Q were self-reported by participants, the virtual task designs were directly based on Items 8 to 13, which cover core powered mobility skills such as straight-line driving, turning, and manoeuvring. Mappings to additional items were inferred at an indirect level.

To avoid bias in quantitative comparisons, the PMRT mappings in this study were limited to the first 12 structured tasks defined by (Massengale et al., 2005). Following (Valentini et al., 2024), the final four PMRT tasks were excluded due to their unstructured nature and lack of standardisation, which may compromise measurement consistency. An overview of the virtual scenario and the full simulator

task sequence is illustrated in Figure 5.3. Optional auxiliary camera views (rear and top-down) were available to the user during any task to support spatial awareness and orientation as needed.

Table 5.1: Virtual Tasks with Corresponding PMRT and WST-Q Items

Task	Description	Corresponding PMRT item(s)**	Corresponding WST-Q Item (s)***
1	Approach and stop as close as possible to 1.5 × 1.5 m virtual table without collision.	**PMRT Item 1:** Approaching people/furniture without bumping into them	**WST Item 8**: Rolls forward. **WST Item 14**: Reaches objects (indirect) WST Item
2	Stop at three marked positions for 5 seconds each, with audio cues to proceed.	**PMRT Item 2:** Starting and stopping wheelchair at will	**WST Item 8**: Rolls forward. **WST Item 11**: Turns while moving forward
3	Cross a narrow (0.9 m) doorway and drive straight for 1 m.	**PMRT Item 3:** Passing through doorways without hitting walls	**WST Item 18**: Gets through hinged door (contextual).
4	Drive forward 2 m, turn 90° right, and continue 1 m.	**PMRT Item 4:** Turning around a 90° right-hand corner	**WST Item 11**: Turns while moving forward
5	Same as Task 4, but with a 90° left turn.	**PMRT Item 5:** Turning around a 90° left-hand corner	**WST Item 11**: Turns while moving forward
6	Drive straight forward for 4.5 m in an open area.	**PMRT Item 6:** Driving straight forward (15 ft) in an open area	**WST Item 8**: Rolls forward
7	Drive straight backward for 3 m with limited visual cues.	**PMRT Item 7:** Driving straight backward (10 ft) in an open area	**WST Item 9**: Rolls backward
8	Perform a series of turns including a 180° in-place U-turn.	PMRT Item 8: Turning 180°	**WST Item 10**: Turns in place. **WST Item 11**: Turns while moving forward. **WST Item 12**: Turns while moving backward
9*	Free driving session with stop/start commands via audio cues. *Used as joystick practice; not analysed.*	**PMRT Item 9:** Starting and stopping wheelchair upon request. Not applicable – practice only	Not applicable – practice only
10	Drive forward, turn 90° right, then 90° left, and continue forward.	**PMRT Item 10:** Turning right and left upon command	**WST Item 11**: Turns while moving forward
11	Drive 4.5 m in a narrow 1 m-wide corridor without hitting walls.	**PMRT Item 11:** Driving straight forward in a narrow corridor without hitting walls	**WST Item 8**: Rolls forward. **WST Item 18**: Gets through hinged door (contextual)
12	Zigzag through six staggered virtual chairs while avoiding collisions.	**PMRT Item 12:** Manoeuvre between objects	**WST Item 11**: Turns while moving forward. **WST Item 13**: Manoeuvres sideways

*Task 9 was a free driving session with audio stop/start cues, conducted before the assessment to calibrate joystick settings (e.g., acceleration, speed) and gather user feedback on control preferences.

** PMRT Framework: Only the first 12 structured PMRT.

*** WST Framework: Virtual tasks were directly based on Items 8–13, other mappings were indirect.

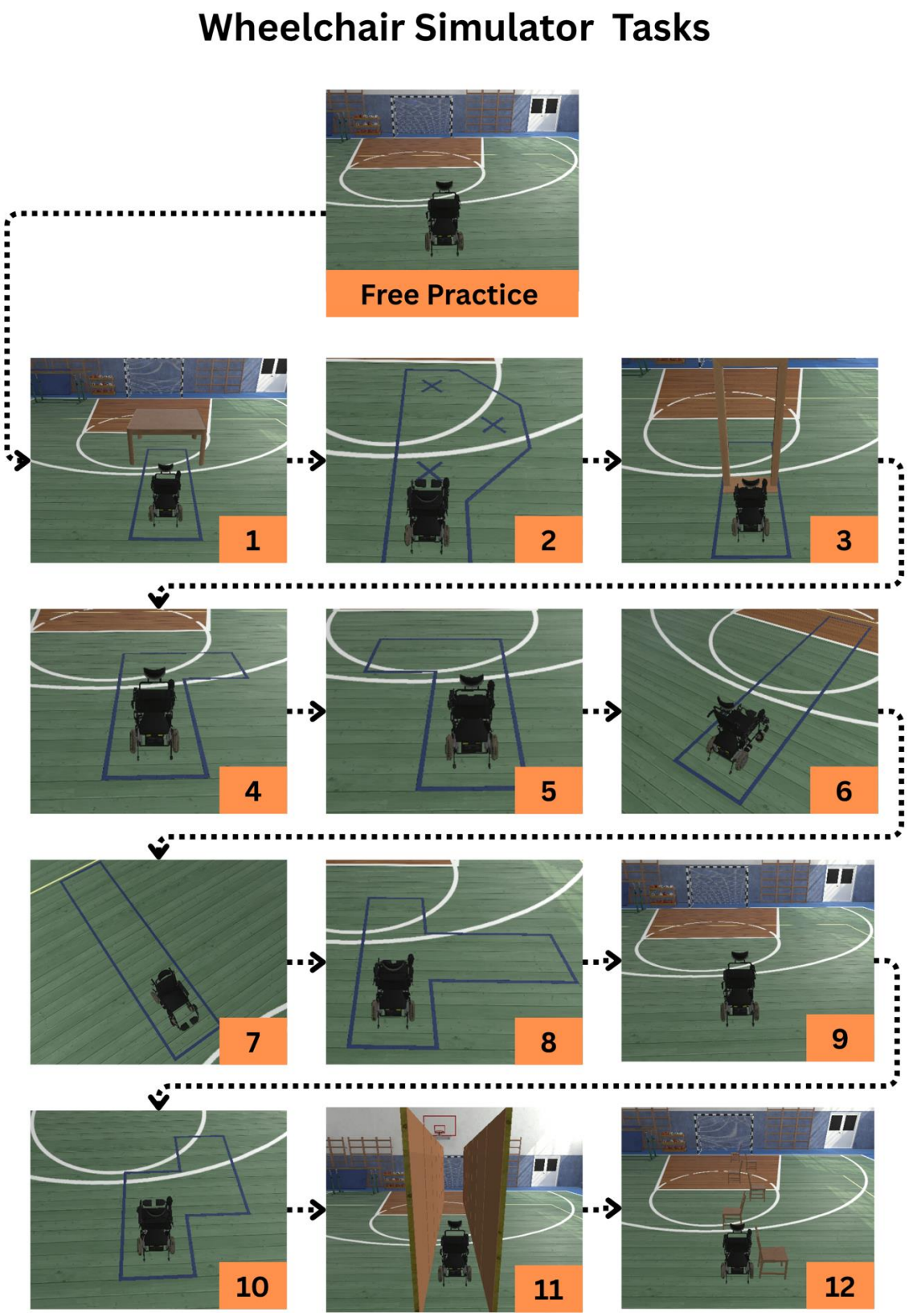


Figure 5.3: Sequence of. Tasks in E-WATS Simulator.

5.1.8 Assessments

Participants completed a structured set of assessments at three stages of the study, before, during, and after the simulator session, summarized in Table 5.2. These assessments evaluated power mobility performance, cognitive functioning, and QoE using clinical tools, simulator-based data, and subjective instruments.

The MoCA was administered at baseline to assess cognitive status. The WST-Q was completed once by wheelchair users and pre-/post- by controls, with only post-test scores used in analysis. Simulator tasks, modelled on PMRT and WST instruments, captured objective performance data, task time, collisions, joystick activity, and RMSE, to quantify driving accuracy. QoE was assessed using validated tools: SAM (Bradley & Lang, 1994), for emotional state, PAAS (F. G. W. C. Paas & Van Merriënboer, 1994a) and NASA-TLX (Hart, 2006) for cognitive workload, short version of SUS (Brooke, 1996), for usability, and short version of IPQ (Igroup Project Consortium, 2015) for sense of presence/immersion.

Heart rate was monitored continuously via a wearable wrist sensor, while jerk (a measure of motion smoothness) was computed from wrist accelerometer data to represent behavioural response. The simulator tasks served as the core of the mobility evaluation, capturing performance and physiological responses aligned with the WST and PMRT frameworks. These tasks provided a controlled yet ecologically valid foundation for linking observed behaviour with perceived usability, workload, and emotional engagement.

In addition to the standardized SUS and IPQ instruments, a set of custom usability questions was included to evaluate participants' perceptions of task realism, ease of use, skill development, and system acceptability. These items were developed based on prior literature and mapped to relevant constructs in usability and simulator-based training studies. Table 5.3 presents the full list of questions, indicating their associated instruments and the specific usability dimensions assessed.

Table 5.2: Outcomes Table

Outcome Category	Measure	Timing	Assessment Method
Primary Outcomes	Power Mobility Skills (performance)	During the test	Number of completed Virtual Tasks based on Power Mobility Road Test (PMRT) and Wheelchair Skill test (WST). Simulator-based metrics.
		Pre- test/ Post-test	Wheelchair Skill Test Questionnaire (WST-Q)
	Cognitive Functioning	Pre-test	Montreal Cognitive Assessment (MoCA)
Secondary Outcomes	Quality of Experience	Post-test	Emotional Responses Self-Assessment Manikin (SAM)
		Post-test	Session-level Cognitive Workload (NASA-TLX)
		During the test	Task-level Cognitive Workload (PAAS)
		Post-test	Usability – System Usability Scale (SUS)
		Post-test	Presence (IPQ)
		During the test	Heart Rate (Monitored using wearables devices)
		During the test	Jerk (wrist-based): Represent hand control during the task execution.

Table 5.3: Post-experience questionnaire items

	Question	Instrument / Construct	Dimension Assessed
1	How well do you believe your performance reflects your current power mobility skills (joystick control)?	QoE Interaction (custom)	Joystick accuracy / ecological validity
2	Did the tasks seem too easy or too difficult for your abilities?	QoE Interaction (custom)	Task difficulty
3	Do you feel your power mobility skills improved during the simulator session?	QoE Interaction (custom)	Skill performance improvement
4	After using the simulator, do you feel more confident in handling a power wheelchair?	QoE Interaction (custom)	Confidence improvement
5	How aware were you of the real-world surroundings while navigating in the virtual world?	IPQ	Involvement (INV)
6	I had a sense of acting in the virtual space, rather than operating from outside.	IPQ	Realism (REAL)
7	How much did your experience in the virtual environment seem consistent with your real-world experience?	IPQ	Acting in virtual environment (SP)
8	In the computer-generated world, I had a sense of "being there".	IPQ	General presence (PRES)
9	I felt present in the virtual space.	IPQ	Sense of being in VE (SP)
10	I would imagine people would learn to use the simulator quickly.	SUS	Learnability
11	I found the system very difficult to use.	SUS	Ease of use
12	I needed to learn a lot before I could get going with the system.	SUS	Ease of learning
13	Was the simulator easy to understand and navigate?	SUS	Navigation / intuitiveness
14	Did the interface feel user-friendly and intuitive?	SUS	Interface usability
15	Were the tasks and instructions clear and easy to follow?	QoE Service (custom)	Instruction clarity
16	How satisfied are you with your experience using the simulator?	QoE Service (custom)	Overall Satisfaction
17	Would you recommend the simulator as a training tool for new power wheelchair users?	QoE Service (custom)	Acceptability as training tool
18	Would you recommend the simulator as an assessment tool for new wheelchair users?	QoE Service (custom)	Acceptability as assessment tool

5.1.9 Outcomes

The outcomes of this field-based study are presented in alignment with the study's guiding objectives and are categorised into primary and secondary domains (see Table 5.2). These results contribute to evaluating the feasibility, usability, and potential clinical relevance of the simulator in real-world settings. It contributes directly to Sub-Research Question 2 and address Objectives 2.1, 2.2, and 2.3.

Primary outcomes addressed Objective 2.2, which focused on exploring whether simulator-derived metrics could complement standardised clinical assessments. Performance data were collected during the simulator tasks, including task completion time, collision count, joystick control variability, and RMSE. These were considered in relation to participants' scores on the WST and MoCA, offering a preliminary perspective on the extent to which simulator metrics might reflect clinically relevant aspects of mobility and cognitive function.

Secondary outcomes supported Objectives 2.1 and 2.3, focusing on evaluating feasibility, user acceptability, and the refinement of the QoE-based evaluation framework. These outcomes included self-reported assessments of usability (SUS), cognitive workload (NASA-TLX and PAAS), emotional response (SAM), and presence (IPQ). In addition, physiological (heart rate) and behavioural (wrist-based acceleration and jerk) data were recorded during simulator use to enrich the understanding of user experience and task engagement. Together, these measures provided insights into how users perceived the system and identified areas for protocol improvement.

Overall, the outcomes presented in this chapter reflect the multidimensional nature of the study's aims. The analysis highlights the potential of simulator-based tools to contribute to assessment in assistive contexts, while also identifying important considerations for further development, validation, and clinical integration.

5.1.10 Statistical Analysis

This pilot study followed STARD 2015 (J. F. Cohen et al., 2016) guidelines for feasibility and diagnostic evaluation report. Descriptive statistics were used to summarize participant demographics, dropout rates, and outcomes related to usability (SUS), cognitive workload (NASA-TLX, PAAS), immersion/presence (IPQ), emotional state (SAM), cognitive function (MoCA), and perceived mobility skills (WST-Q). Continuous variables were reported using mean, standard deviation, range, median, interquartile range (IQR); categorical variables were summarized using counts and percentages.

To examine the simulator's sensitivity, performance metrics (e.g., task duration, collision count) were compared between wheelchair users and controls. Exploratory correlations were conducted between simulator metrics and clinical scores (MoCA, WST-Q) to assess alignment and inform future validation.

5.1.11 Results

5.1.11.1 Participants Flow Overview

Seventeen powered wheelchair users were initially recruited for this pilot study. Of these, twelve met the eligibility criteria and commenced the study protocol. Ten participants successfully completed all virtual simulator tasks and post-assessments, constituting the final sample for questionnaire-based analysis. Among them, physiological data were retained from eight participants; two datasets were excluded due to signal loss or sensor malfunction.

In parallel, eleven non-disabled control participants were recruited. One individual was unable to complete the simulator protocol, resulting in a final control group of ten participants with complete data. Accordingly, twenty participants (10 wheelchair users, 10 controls) formed the final dataset for feasibility and comparative analysis. The detailed participant recruitment and retention process is illustrated in Figure 5.4.

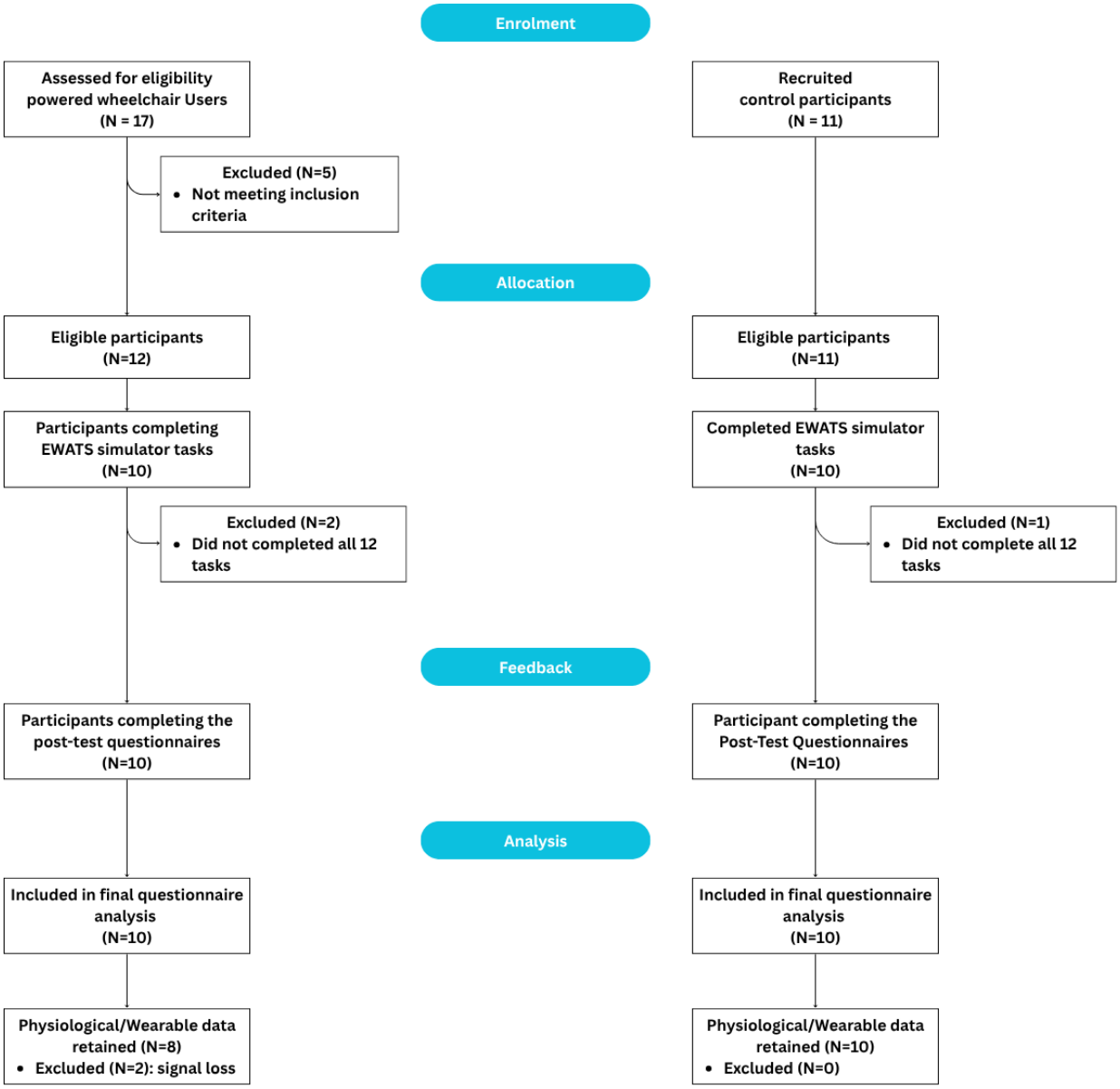


Figure 5.4: Flowchart of participant enrolment.

5.1.11.2 Demographics

The dataset analysed consisted of ten powered wheelchair users (6 females, 4 males; mean age = 54.7 ± 17.4 years) and ten non-disabled control participants (3 females, 7 males; mean age = 29 ± 7.6 years).

The wheelchair user group included individuals with a variety of primary diagnoses, such as stroke, multiple sclerosis, cerebral palsy, spina bifida, scoliosis, and musculoskeletal conditions. While this represents a spectrum of conditions, the small sample size limits the generalizability of the findings. The presence of a limited number of participants for each specific condition does not constitute a representative sample for a clinical population. Consequently, the findings from this study are interpreted as a pilot feasibility study, which provides a foundation for future, larger-scale research with more diverse and adequately sized cohorts.

Participants in the wheelchair user group have experience of using a powered wheelchair varied from 1 to 21 years (mean = 11.5 ± 6.7 years), reflecting both recent and long-term users. Only one participant in this group reported regular engagement with video games. In contrast, control participants had no prior experience using powered wheelchairs. However, the majority (8 out of 10) reported frequent video game use, indicating greater familiarity with virtual interfaces and game-based environments. Group-wise descriptive statistics, including means, standard deviations, and values ranges for each key variable, are summarised in Table 5.4. Detailed individual-level data, including demographic information, MoCA scores, WST-Q performance and confidence scores (real and simulated), and video game experience, are presented in Table 5.5.

Cognitive screening using MoCA revealed a broader range of scores in the wheelchair user group (range: 17–27), with three individuals scoring below the commonly used clinical threshold of 26. In the control group, MoCA scores were uniformly high (range: 27–30), with a median of 29.

Self-reported real-world WST-Q performance and confidence were generally lower among control participants, as expected due to their lack of wheelchair use experience, although some high-scoring individuals were observed. For simulator-based WST-Q tasks (questions 8–13), all participants completed the same virtual driving scenarios. Wheelchair users tended to show higher confidence and consistency in these tasks compared to controls.

Table 5.4: Study 3 demographics summary statistics

Group	F	M	Play video games (yes)	Play video game (no)	Mean Age (SD)	Mean PWC Use (SD)	MoCA (SD)	WST Perf. (SD)	WST Conf. (SD)	WST Q8-Q13 Perf. (SD)	WST Q8-13 Conf. (SD)
Users	6	4	1	9	54.7 (17.42)	11.50 (6.20)	23.7 (2.91)	79.67 (9.17)	77.52 (9.99)	94.44 (11.11)	93.89(6 .65)
Control	3	7	8	2	29 (7.56)	0	29 (0.82)	76.67 (18.66)	75.00 (22.70)	81.11 (19.46)	73.33 (24.54)

Table 5.5: Study 3 Full Participant Demographics and Scores

	Gender	Age	Diagnosis	PWC Use Year	Play video games	MoCA	WST Perf.	WST Conf.	WST Q8-13 Perf.	WST Q8-13 Conf.
U_1	M	35	Multiple Sclerosis (MS)	5	yes	25	85.90	80.00	100.0	88.89
U_2	F	67	Degenerative disease	8	no	22	69.23	73.33	66.67	94.44
U_3	F	69	Musculoskeletal	21	no	24	96.15	96.15	100.0	100.0
U_4	F	56	Scoliosis (Congenital)	1	no	17	74.67	74.36	83.33	94.44
U_5	F	29	Cerebral Palsy (CP)	19	no	26	79.71	70.83	100.0	94.44
U_6	F	30	Spine Bifida	16	no	26	73.33	74.36	100.0	100.0
U_7	M	60	Stroke (CVA)	14	no	25	74.36	66.67	100.0	83.33
U_8	M	64	Stroke (CVA)	12	no	22	80.56	80.56	100.0	100.0
U_9	F	79	Musculoskeletal	9	no	23	70.51	66.67	94.44	83.33
U_10	M	58	Stroke (CVA)	10	no	27	92.31	92.31	100	100
C_1	F	32	N/A	0	no	29	61.54	62.82	83.33	61.11
C_2	M	30	N/A	0	no	29	74.36	69.23	83.33	66.67
C_3	M	43	N/A	0	yes	29	94.87	100.0	100.0	100.0
C_4	M	18	N/A	0	yes	29	58.97	52.56	44.44	33.33
C_5	F	40	N/A	0	yes	30	58.97	52.56	66.67	61.11
C_6	M	21	N/A	0	yes	29	91.03	97.44	100	100
C_7	M	27	N/A	0	yes	27	100	100	100	100
C_8	M	25	N/A	0	yes	29	51.28	42.31	66.67	55.56
C_9	F	26	N/A	0	yes	29	100	100	100	100
C_10	M	27	N/A	0	yes	30	75.64	73.08	66.67	55.56

5.1.11.3 Simulator-based Performance Results

To explore group-level trends, simulator-derived metrics were averaged across participants within each group. This included total task completion time, number of collisions, number of commands, and mean RMSE values. Group-level descriptive statistics are summarised at Table 5.6. Individual participant data are presented in Table 5.7. Corresponding boxplot charts are presented in Figure 5.5.

While the control group appeared to complete tasks in less time and with fewer collisions and commands, as well as lower RMSE values, these differences should be interpreted with caution due to the small sample size and the high variability observed within the user group. This variation, particularly in time and RMSE, reflects the expected heterogeneity in motor and cognitive abilities among individuals with diverse functional profiles. To statistically examine group differences, Mann–Whitney U tests were performed for each metric (Table 5.6). Statistically significant differences were found in the number of collisions and mean RMSE, both of which were higher in the user group. These differences were also associated with large effect sizes. Although the difference in task completion time was not statistically significant, the observed effect size was moderate, indicating a potential trend warranting further investigation.

In summary, these preliminary results suggest that the simulator may be sensitive to differences in task performance between participants with and without prior wheelchair experience, warranting further investigation in larger and more diverse samples.

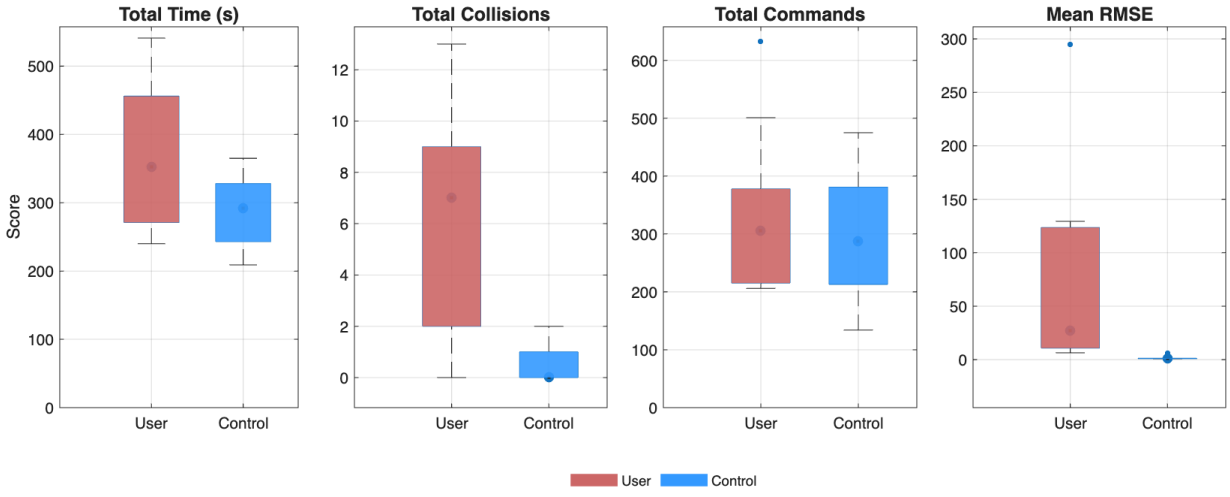


Figure 5.5: Study 3 group-wise comparisons for simulator-based task performance.

Table 5.6: Study 3 Group-wise comparisons of performance metrics

Metric	Group	Mean (SD)	Median [IQR]	U	Z	p-value	Effect size (r)
Elapse Time (s)	Users	371.7(109,05)	353 [271.00 – 456.00]	26.00	-1.81	0.075	0.40
	Control	288.7(50.32)	292.50 [243.00 – 328.00]				
Collisions	Users	6.3 (4.24)	7.00 [2.00 – 9.00]	7.50	-3.33	0.000*	0.74
	Control	0.4 (0.70)	0.00 [0.00 – 1.00]				
Commands	Users	338.9 (140.82)	305.50 [215.00 – 378.00]	44.50	-0.42	0.684	0.09
	Control	295 (102.15)	287.00 [213.00 – 381.00]				
RMSE	Users	67.95 (92.01)	27.28 [10.80 – 123.57]	0.00	-3.78	0.000*	0.85
	Control	1.77 (1.63)	1.10 [1.03 – 1.35]				

Table 5.7: Individual participant results for simulator-based task performance.

ID	Elapse Time (s)	Total Collisions	Total Commands	Mean RMSE
U_1	240.00	2.00	259.00	6.95
U_2	311.00	8.00	210.00	17.51
U_3	254.00	6.00	206.00	6.36
U_4	413.00	8.00	378.00	129.34
U_5	271.00	2.00	352.00	10.80
U_6	366.00	0.00	377.00	30.21
U_7	340.00	4.00	215.00	35.24
U_8	541.00	13.00	501.00	24.36
U_9	525.00	9.00	633.00	123.57
U_10	456.00	11.00	258.00	295.14
C_1	339.00	0.00	389.00	6.00
C_2	365.00	2.00	381.00	3.13
C_3	209.00	0.00	134.00	1.06
C_4	277.00	0.00	213.00	0.89
C_5	328.00	0.00	475.00	0.76
C_6	267.00	1.00	269.00	1.04
C_7	315.00	0.00	305.00	1.03
C_8	243.00	1.00	245.00	1.35
C_9	236.00	0.00	206.00	1.15
C_10	308.00	0.00	333.00	1.28

5.1.11.4 Usability and Presence

This section presents the results of the post-simulation questionnaire, which included 18 items assessing user experience across three primary domains: (i) QoE related to perceived interaction with the simulator (Q1–Q4), (ii) presence and immersion, measured via the IPQ (Q5–Q9), and (iii) system usability (SUS) and service-level Quality of Experience, encompassing usability metrics (Q10–Q14) and service perceptions such as satisfaction and acceptability (Q15–Q18). Descriptive statistics and between-group comparisons are summarized in Table 5.8.

5.1.11.5 Usability

Participants' responses across the usability-related questions indicated overall positive perceptions, with no statistically significant differences between wheelchair users and controls across most items. In the first set of questions (Q1–Q4), which targeted QoE feature related to perceived interaction, both groups reported comparable ratings concerning joystick accuracy, task difficulty, perceived skill improvement, and confidence gained. Although users tended to score slightly higher on joystick accuracy and confidence, and control groups on skill improvement, none of these differences reached statistical significance (see Table 5.8 and Figure 5.6).

The second set (Q10–Q14), derived from SUS, assessed aspects such as learnability, ease of use, and interface usability. Both groups provided moderate to high usability ratings. Controls gave higher ratings for learnability and ease of use, while users rated navigation and interface usability marginally better. Again, these differences were not statistically significant (see Table 5.8 and Figure 5.7).

Finally, the third set (Q15–Q18), addressing QoE feature related to service aspects, clarity, satisfaction, and system acceptability, showed slightly more divergent responses. Notably, satisfaction (Q16) was rated significantly higher by controls (M = 4.80) compared to users (M = 3.90, p = .0011, r = .61). While both groups reported high levels of acceptability for training and assessment use, controls showed a ceiling effect (median = 5.00), whereas users' ratings were more distributed. Overall, these results suggest that both groups found the system usable and acceptable, with minor differences in perceived satisfaction and ease of use (Figure 5.8).

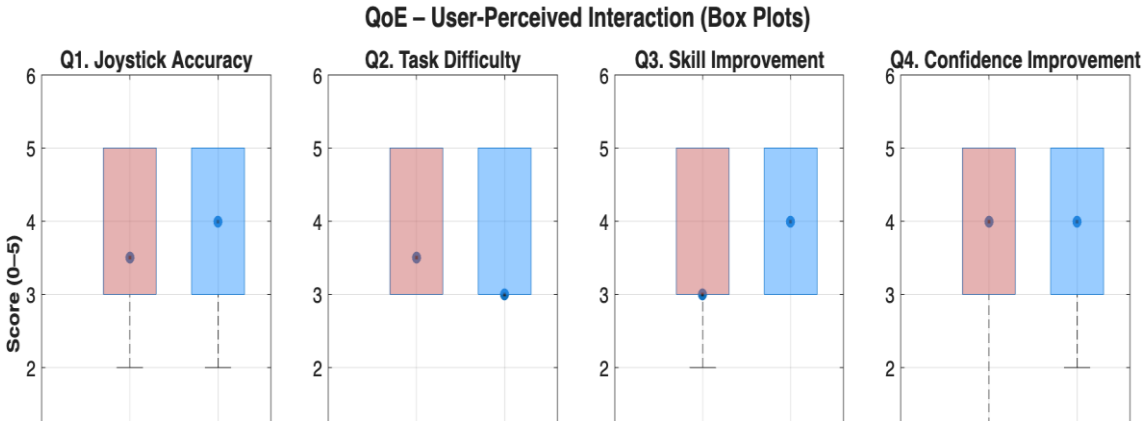


Figure 5.6: Study 3 QoE Interaction Items results (Q1-Q4).

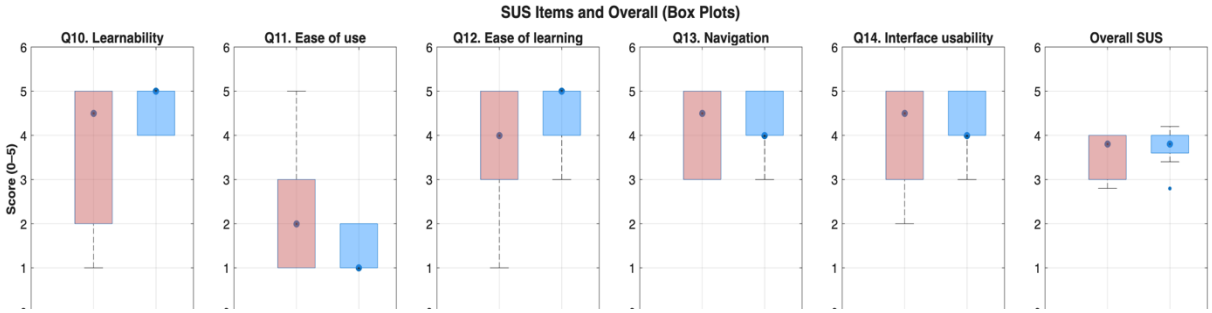


Figure 5.7: Study 3 SUS Items and Overall score (Q10-Q14).

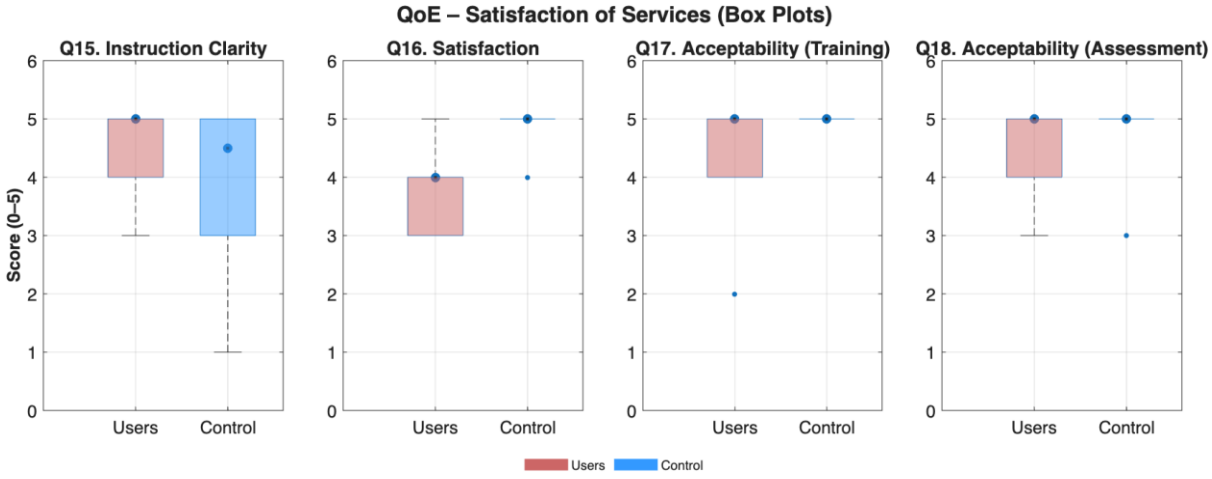


Figure 5.8: Study 3 QoE Satisfaction Items (Q15-Q18).

Table 5.8: Study 3 Group-wise comparisons of QoE, Usability and Presence (IPQ).

Metric	Group	Mean (SD)	Median [IQR]	U	Z	p-value	Effect size (r)
Q1. QoE interaction - Joystick Accuracy	Users	3.80 (1.14)	3.50 [3.00 - 5.00]	47.50	-0.20	0.853	0.04
	Control	3.70 (1.16)	4.00 [3.00 - 5.00]				
Q2. QoE Interaction - Task Difficulty	Users	3.90 (0.99)	3.50 [3.00 - 5.00]	44.50	-0.47	0.684	0.10
	Control	3.70 (0.95)	3.00 [3.00 - 5.00]				
Q3. QoE Interaction - Skill Performance Improvement	Users	3.70 (1.16)	3.00 [3.00 - 5.00]	39.50	-0.85	0.436	0.19
	Control	4.10 (0.88)	4.00 [3.00 - 5.00]				
Q4. QoE Interaction - Confidence Improvement	Users	3.80 (1.32)	4.00 [3.00 - 5.00]	48.50	-0.12	0.912	0.03
	Control	3.90 (1.20)	4.00 [3.00 - 5.00]				
Q5. IQP - Involvement (INV)	Users	2.00 (1.41)	1.00 [1.00 - 3.00]	26.00	-1.92	0.075	0.43
	Control	3.10 (1.10)	3.00 [3.00 - 4.00]				
Q6. IPQ - Realism (REAL)	Users	3.10 (1.45)	3.00 [2.00 - 4.00]	27.00	-1.80	0.089	0.40
	Control	4.20 (0.79)	4.00 [4.00 - 5.00]				
Q7. IPQ - Acting in VR (SP)	Users	4.00 (0.94)	4.00 [3.00 - 5.00]	37.00	-1.03	0.353	0.23
	Control	3.40 (1.26)	3.50 [3.00 - 4.00]				
Q8. IPQ - General Presence	Users	3.30 (1.42)	3.50 [3.00 - 4.00]	42.50	-0.59	0.579	0.13
	Control	3.10 (1.10)	3.00 [3.00 - 4.00]				
Q9. IPQ - Sense of being in VE (SP)	Users	3.30 (1.42)	3.50 [3.00 - 4.00]	44.00	-0.47	0.684	0.10
	Control	3.10 (1.20)	3.00 [2.00 - 4.00]				
Q10. SUS - Learnability	Users	3.70 (1.57)	4.50 [2.00 - 5.00]	37.00	-1.09	0.353	0.24
	Control	4.60 (0.52)	5.00 [4.00 - 5.00]				
Q11. SUS - Ease of use	Users	2.30 (1.42)	2.00 [1.00 - 3.00]	29.00	-1.75	0.123	0.39
	Control	1.30 (0.48)	1.00 [1.00 - 2.00]				
Q12. SUS - Ease of learning	Users	3.60 (1.35)	4.00 [3.00 - 5.00]	32.00	-1.44	0.19	0.32
	Control	4.40 (0.84)	5.00 [4.00 - 5.00]				
Q13. SUS - Navigation/ intuitiveness	Users	4.10 (0.99)	4.50 [3.00 - 5.00]	48.50	-0.12	0.912	0.03
	Control	4.20 (0.63)	4.00 [4.00 - 5.00]				
Q14. SUS - Interface usability	Users	4.00 (1.15)	4.50 [3.00 - 5.00]	47.00	-0.24	0.853	0.05
	Control	4.20 (0.79)	4.00 [4.00 - 5.00]				
Q15. QoE Service - Instruction clarity	Users	4.40 (0.84)	5.00 [4.00 - 5.00]	42.00	-0.67	0.579	0.15
	Control	3.90 (1.45)	4.50 [3.00 - 5.00]				
Q16. QoE Service - Satisfaction	Users	3.90 (0.74)	4.00 [3.00 - 4.00]	17.00	-2.74	0.011*	0.61
	Control	4.80 (0.42)	5.00 [5.00 - 5.00]				
Q17. QoE Service - Acceptability (training)	Users	4.50 (0.97)	5.00 [4.00 - 5.00]	35.00	-1.83	0.28	0.41
	Control	5.00 (0.00)	5.00 [5.00 - 5.00]				
Q18. QoE Service - Acceptability (assessment)	Users	4.50 (0.71)	5.00 [4.00 - 5.00]	36.50	-1.35	0.315	0.30
	Control	4.80 (0.63)	5.00 [5.00 - 5.00]				

*. The difference is significant at the 0.05 level.

5.1.11.6 Sense of Presence (IPQ)

Overall, responses varied between groups, with control participants generally reporting higher levels of realism and involvement, while wheelchair users scored higher on items related to the feeling of acting within and being present in the virtual environment. Figure 5.9 and Figure 5.10 provide visual summaries of IPQ item responses. The radar plot (Figure 5.9) displays the distribution of perceived presence features across groups, while the boxplot (Figure 5.10) highlights differences in mean scores between Controls and Users participants.

The involvement (Q5) and realism (Q6) dimensions showed the largest differences. Controls rated both aspects higher on average, although these differences did not reach statistical significance ($p = .075$ and $p = .089$, respectively). Interestingly, wheelchair users reported a stronger sense of acting within the virtual space (Q7), with a higher median score than controls, suggesting they may have experienced a stronger embodied interaction despite the lower realism. Scores for general presence (Q8) and sense of being present (Q9) were relatively similar across both groups, with no significant differences observed.

These findings indicate that while both groups experienced a moderate sense of presence, the control group may have perceived the environment as more realistic and immersive, whereas users may have felt more actively engaged in navigating the virtual space.

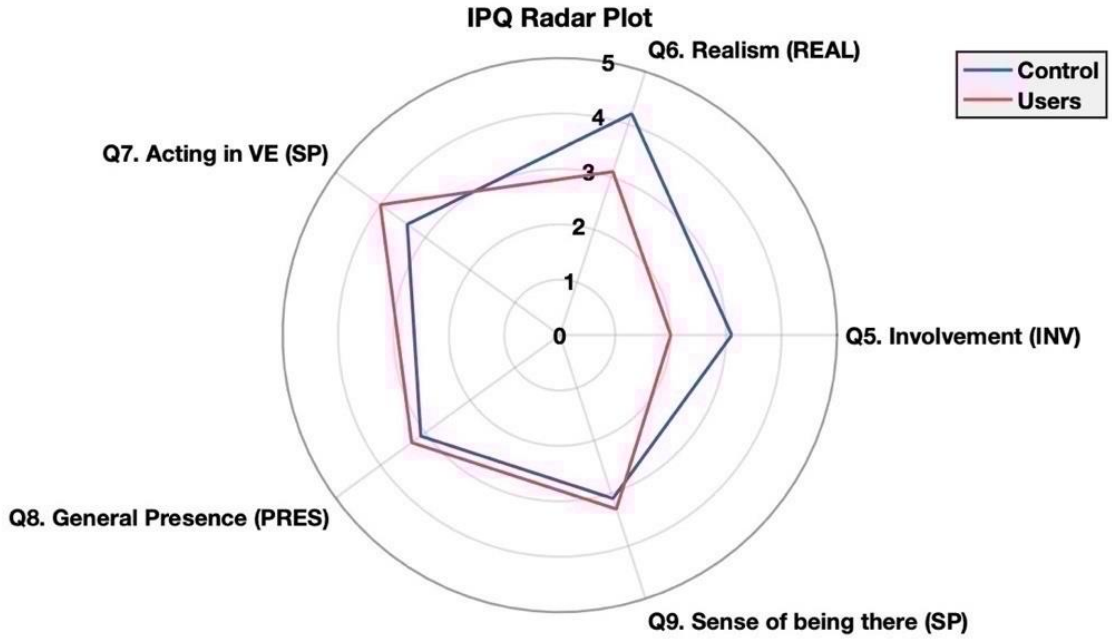


Figure 5.9: Study 3 Radar plot of IPQ scores by group.

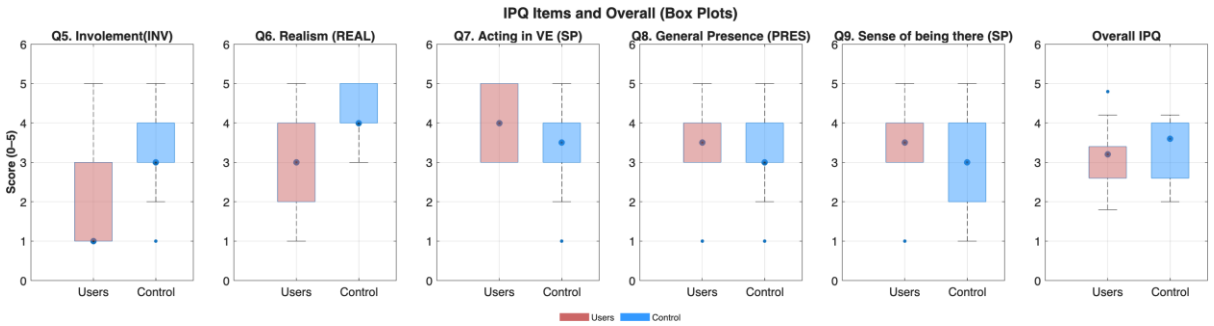


Figure 5.10: Study 3 Boxplot of IPQ items by group.

5.1.11.7 Cognitive Workload (NASA-TLX and Paas scales)

5.1.11.7.1 Cognitive workload at session level (NASA-TLX)

Only the raw NASA-TLX scores were analysed to compare overall cognitive workload across groups. Descriptive results (Table 5.9) show that both groups reported comparable median and mean Raw TLX values, with no statistically significant differences across the individual dimensions or overall workload (all $p > .05$). Figure 5.12 presents a radar plot of mean scores for each NASA-TLX subscale. Figure 5.11 provides a box plot of group means with standard deviations. Wheelchair users tended to report slightly higher scores in mental demand, effort , and performance demand, whereas control participants reported higher frustration and temporal demand. However, these tendencies were not supported by significant Mann–Whitney U test results, and all effect sizes were small ($r \leq 0.30$).

Overall, the findings suggest that perceived cognitive workload at session level was similar between groups, though variability was observed within the user group across several workload dimensions. This pattern is consistent with prior wheelchair simulator studies that reported moderate levels of cognitive workload in desktop-based simulator designs compared to immersive settings.

For example, (L. Devigne et al., 2017) reported a mean NASA-TLX score of M=27.2, SD=18.2 during joystick-based virtual tasks, indicating manageable workload levels. Vailland et al. (2020) (Vailland et al., 2020) found increased NASA-TLX performance sub-scores when vestibular feedback was added to VR simulations, no vestibular feedback (M=7.06, SD= 5.76) versus multisensory feedback (M=4.06, SD=4.78). Fraudet et al. (2024) (Fraudet et al., 2024) observed elevated workload across virtual circuits of increasing difficulty, circuit-1 (M=18.33, SD=21.03), circuit-2 (M=15.33, SD=17.81) and circuit-3 (M=19.17, SD=19.60). (Kamaraj, Dicianno, et al., 2016b) reported higher workload demand under immersive VR conditions (M=49.39, SD=25) compared to standard PC screen interfaces (M=48.49, SD=24) and real world (M=22.4, SD=17.3).

Collectively, these findings underscore how simulator features influence perceived workload, though scores generally remain within moderate ranges, consistent with the present results. To further explore task-level differences in cognitive mental workload, the following section presents Paas mental effort ratings collected after each simulator task.

Table 5.9: Study 3 Group-wise comparisons of NASA-TLX items

Metric	Group	Mean (SD)	Median [IQR]	U	Z	p-value	Effect size (r)
Nasa mental	Users	50.55 (30.94)	52.50 [27.50 – 83.00]	44.00	-0.45	0.684	0.10
	Control	45.00 (25.93)	40.00 [25.00 – 70.00]				
Nasa physical	Users	26.00 (28.56)	12.50 [10.00 – 40.00]	47.00	-0.23	0.853	0.05
	Control	25.50 (21.66)	20.00 [5.00 – 30.00]				
Nasa temporal	Users	9.00 (11.74)	5.00 [5.00 – 10.00]	35.00	-1.24	0.28	0.28
	Control	18.50 (20.01)	5.00 [5.00 – 25.00]				
Nasa performance	Users	28.50 (14.73)	22.50 [20.00 – 45.00]	41.50	-0.65	0.529	0.15
	Control	29.50 (25.54)	17.50 [15.00 – 35.00]				
Nasa effort	Users	65.25 (26.05)	65.00 [45.00 – 90.00]	32.50	-1.32	0.19	0.30
	Control	48.50 (29.63)	50.00 [25.00 – 70.00]				
Nasa frustration	Users	14.50 (14.03)	10.00 [5.00 – 15.00]	46.50	-0.28	0.796	0.06
	Control	20.00 (18.10)	10.00 [5.00 – 40.00]				
Nasa Raw TLX	Users	32.30 (14.35)	30.42 [25.83 – 37.50]	45.50	-0.34	0.739	0.08
	Control	31.17 (16.80)	28.33 [18.33 – 43.33]				

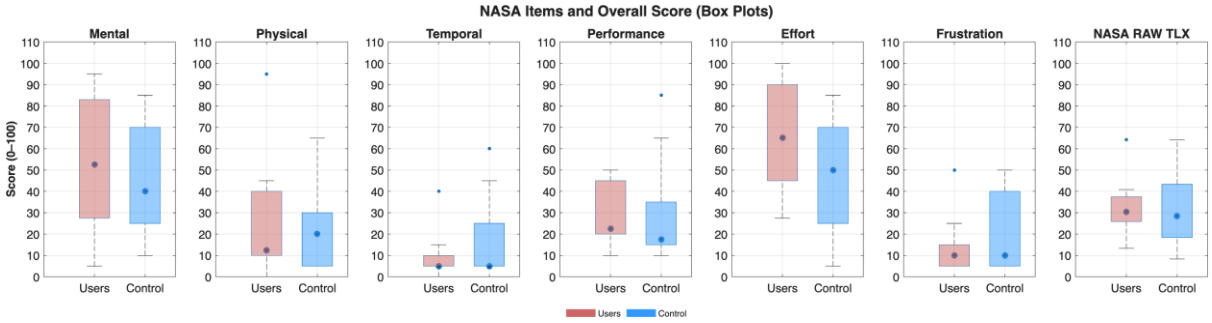


Figure 5.11: Study 3 Boxplot of NASA-TLX raw scores.

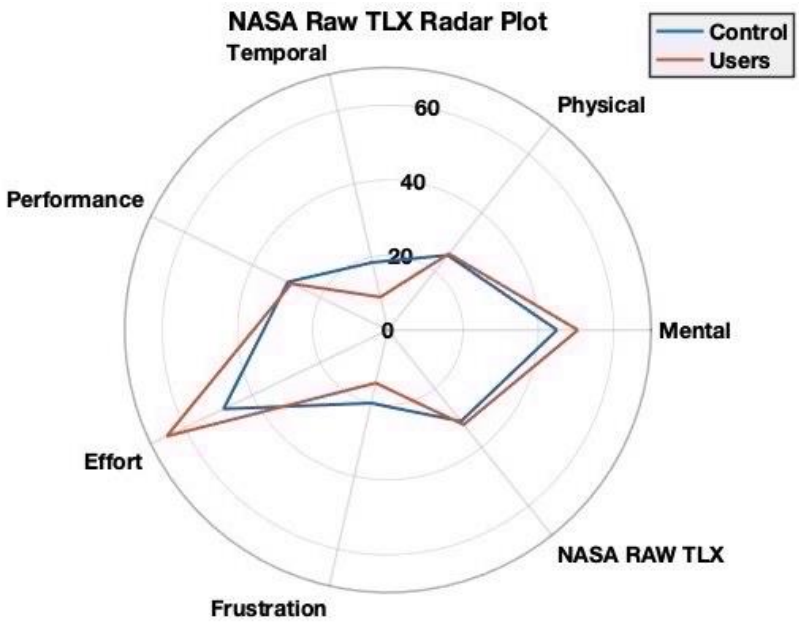


Figure 5.12: Study 3 Radar plot of NASA-TLX raw scores by group.

5.1.11.7.2 Cognitive mental workload at task level (Paas)

Paas mental effort scores per task were analysed to examine differences in perceived mental effort between groups across simulator tasks level. As shown in Table 5.10, group-wise comparisons revealed no statistically significant differences at the $p < .05$ level, though four tasks (Task 8, Task 10, Task 7 and Task 1) approached significance with medium effect sizes ($r = 0.44$, 0.38, 0.32, and 0.31 respectively). Wheelchair users reported higher mental effort than controls in most tasks, especially toward the end of the session (Tasks 7, 8, 10, and 12). These trends are visually depicted in Figure 5.13, which shows the trajectory of mean Paas scores across tasks for each group.

The Paas scale was selected for its simplicity and suitability for repeated assessments with minimal respondent burden, enabling mental effort to be captured directly after each task without interfering with the session flow. It is a widely used instrument in cognitive load research and has demonstrated sensitivity to both intrinsic and extraneous load factors in different contexts (Toy et al., 2020; Wiebe et al., 2010). This allowed a more granular evaluation of task-level workload, in contrast to the NASA-TLX, which provides an overall session-level measure but is more cognitively demanding to complete.

Overall, these results align with the NASA-TLX findings, suggesting comparable cognitive workload levels across groups, with slightly higher perceived mental effort reported by wheelchair users in specific task segments. While NASA-TLX captured an overall impression at the session level, the Paas scale provided more granular insights into task-level cognitive demands, revealing potential accumulative or task-specific workload differences that may not have been reflected in global scores.

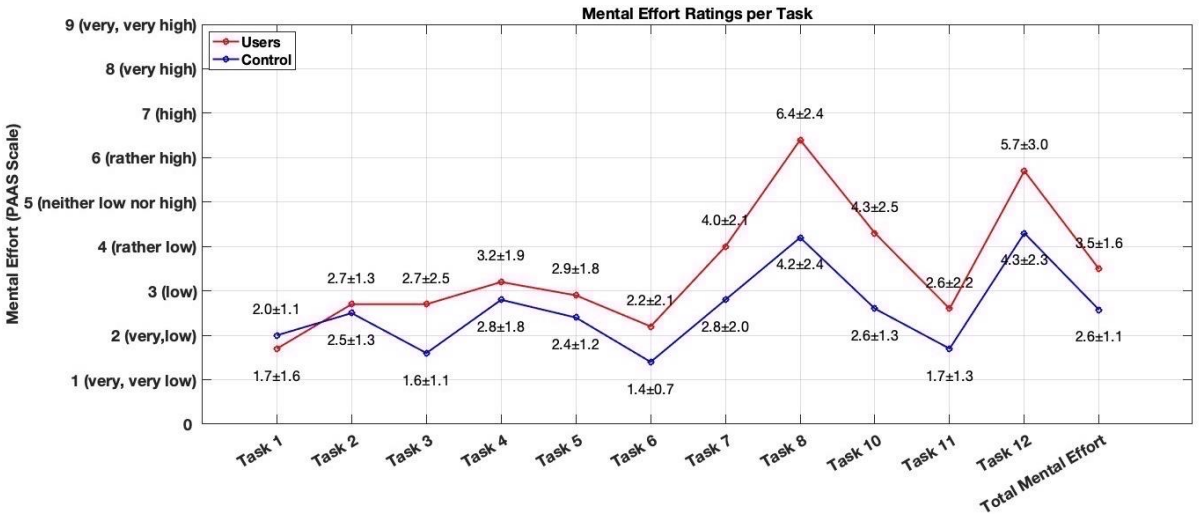


Figure 5.13: Study 3 Paas mental effort line plot per task.

Table 5.10: Study 3 Group-wise comparisons of Paas mental effort per task

Paas Mental Effort	Group	Mean (SD)	Median [IQR]	U	Z	p-value	Effect size (r)
Task 1	Users	1.70 (1.64)	1.00 [1.00 – 1.00]	34.00	-1.37	0.247	0.31
	Control	2.00 (1.05)	2.00 [1.00 – 3.00]				
Task 2	Users	2.70 (1.34)	2.50 [2.00 – 4.00]	45.50	-0.35	0.739	0.08
	Control	2.50 (1.27)	2.00 [2.00 – 3.00]				
Task 3	Users	2.70 (2.45)	2.00 [1.00 – 3.00]	34.50	-1.29	0.247	0.29
	Control	1.60 (1.07)	1.00 [1.00 – 2.00]				
Task 4	Users	3.20 (1.87)	2.50 [2.00 – 5.00]	43.50	-0.50	0.631	0.11
	Control	2.80 (1.75)	2.50 [1.00 – 4.00]				
Task 5	Users	2.90 (1.79)	2.50 [1.00 – 4.00]	43.00	-0.54	0.631	0.12
	Control	2.40 (1.17)	2.50 [1.00 – 3.00]				
Task 6	Users	2.20 (2.15)	1.50 [1.00 – 2.00]	39.00	-0.95	0.436	0.21
	Control	1.40 (0.70)	1.00 [1.00 – 2.00]				
Task 7	Users	4.00 (2.05)	4.50 [2.00 – 5.00]	31.50	-1.42	0.165	0.32
	Control	2.80 (2.04)	2.50 [1.00 – 4.00]				
Task 8	Users	6.40 (2.37)	7.00 [6.00 – 8.00]	24.00	-1.98	0.052	0.44
	Control	4.20 (2.44)	4.50 [2.00 – 6.00]				
Task 10	Users	4.30 (2.45)	3.50 [3.00 – 6.00]	28.00	-1.70	0.105	0.38
	Control	2.60 (1.26)	2.50 [2.00 – 3.00]				
Task 11	Users	2.60 (2.22)	1.50 [1.00 – 5.00]	38.50	-0.98	0.393	0.22
	Control	1.70 (1.34)	1.00 [1.00 – 2.00]				
Task 12	Users	5.70 (2.98)	7.00 [3.00 – 7.00]	34.00	-1.23	0.247	0.27
	Control	4.30 (2.31)	4.00 [3.00 – 6.00]				

5.1.11.8 Emotion (SAM)

Emotional responses were assessed using SAM across three dimensions: valence (pleasure), arousal, and dominance. As presented in Table 5.11, the median and mean scores indicate generally positive emotional experiences for both groups and none of domains reach statistical significant difference with a small-to-medium effect size (r<.34). Figure 5.14 illustrates group-wise mean emotional responses. The control group showed higher valence and similar levels of arousal and dominance when compared to wheelchair users, suggesting slightly more positive affect with comparable valence and perceived control (dominance).

Table 5.11: Study 3 Group-wise comparisons of SAM scales

Metric	Group	Mean (SD)	Median [IQR]	U	Z	p-value	Effect size (r)
SAM valence	Users	6.30 (1.64)	5.50 [5.00 – 7.00]	31.00	-1.51	0.165	0.34
	Control	7.40 (1.58)	7.00 [7.00 – 9.00]				
SAM Arousal	Users	6.10 (1.20)	6.00 [5.00 – 7.00]	45.00	-0.41	0.739	0.09
	Control	6.10 (2.23)	7.00 [5.00 – 7.00]				
SAM Dominance	Users	7.40 (1.84)	8.00 [5.00 – 9.00]	42.50	-0.59	0.579	0.13
	Control	7.00 (1.89)	7.50 [5.00 – 9.00]				

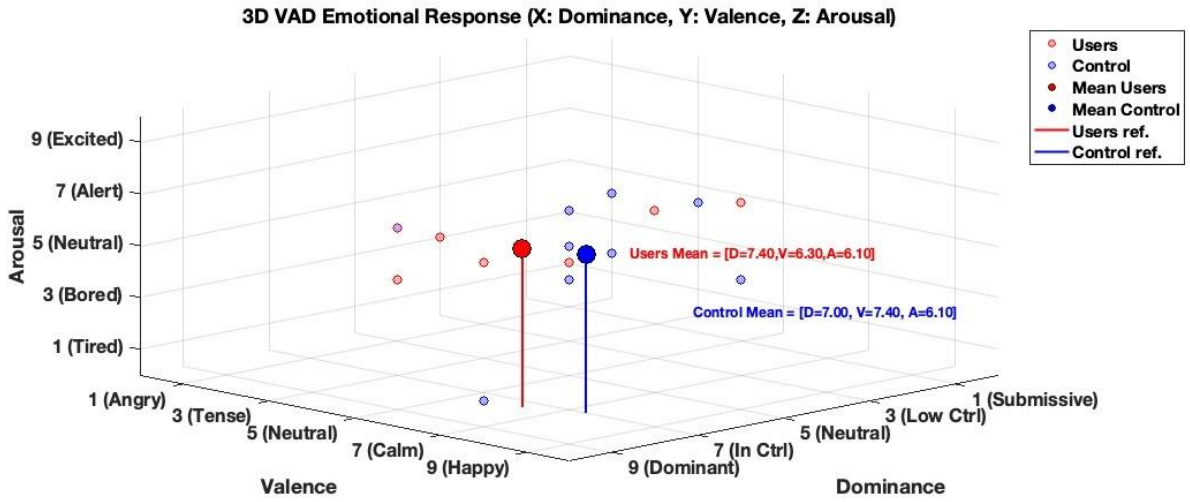


Figure 5.14: Study 3 VAD (Valence–Arousal–Dominance) items by group.

In summary, these results suggest that participants experienced moderate to high arousal and positive emotional states during simulator use, with no substantial differences in emotional valence, arousal or dominance between wheelchair users and controls. The observed variation in valence may reflect individual preferences or comfort levels with the virtual tasks and environment, rather than systematic differences in emotional strain or engagement. These emotional patterns are further

examined alongside physiological indicators in the following section, which analyses group-wise heart rate dynamics during simulator tasks.

5.1.11.9 Heart Rate

Heart rate (HR) metrics were examined across participant groups and task conditions using non-parametric comparisons. The descriptive statistics and Mann–Whitney U test results are summarized in Table 5.12 (session-level comparison), Table 4.13 and Table 7.13 (per-task comparison) in the Appendix.

Session-level analysis indicated no significant difference in baseline HR between users and control participants ($p = 0.696$, $r = 0.11$), suggesting comparable initial physiological states. During the simulation phase, mean HR values were higher among users than controls, with this difference approaching statistical significance ($p = 0.068$, $r = 0.44$), pointing to a moderate effect and potentially heightened arousal (stress or engagement) in the user group.

The most notable finding emerged in the HR change from baseline to task performance. Controls exhibited a significantly larger reduction in HR (M = –15.54 bpm, SD = 7.73) compared to users (M = –1.56 bpm, SD = 7.70), with this difference reaching statistical significance ($p = 0.003$) and associated with a large effect size ($r = 0.67$). Although based on a small sample, this trend suggests distinct patterns of autonomic regulation, whereby users maintained elevated HR during simulation, while controls showed greater decreases, possibly reflecting adaptation or lower task demand.

Table 5.12: Study 3 Group-wise comparisons of HR at session level

Metric (bpm)	Group	Mean(SD)	Median[IQR]	U	Z	p-value	Effect size (r)
HR baseline	User	91.99 (13.68)	87.59 [81.56 - 106.71]	35.00	-0.44	0.696	0.11
	Control	95.39 (8.06)	97.10 [95.00 - 98.65]				
HR session	User	90.43 (12.57)	87.68 [80.32 - 103.45]	19.00	-1.87	0.068	0.44
	Control	79.85 (5.54)	80.74 [74.74 - 83.13]				
HR change	User	-1.56 (7.70)	-2.52 [-4.95 - 1.20]	8.00	-2.84	0.003*	0.67
	Control	-15.54 (7.73)	-16.93 [-21.43 - -11.80]				
HR SD	User	1.65 (0.74)	1.77 [1.02 - 2.29]	34.00	-0.53	0.633	0.13
	Control	1.44 (0.63)	1.34 [0.87 - 1.99]				
*. The difference is significant at the 0.05 level.							

This divergence may reflect different mechanisms of autonomic control between groups. Sustained HR among wheelchair users could signal heightened sympathetic activation, possibly linked to

increased effort, attentional focus, or anticipatory stress when navigating tasks that mimic real-world mobility challenges. In contrast, the larger HR reduction observed in controls may indicate more efficient parasympathetic reactivation, suggesting that the simulator tasks imposed relatively lower physiological or cognitive demands on this group. However, due the small sample size, these trends should be interpreted as exploratory signals rather than definitive evidence of distinct autonomic profiles. Future studies with larger and more diverse cohorts will be required to establish whether HR change can serve as a reliable biomarker of engagement, task demand, or adaptive regulation in simulator-based mobility assessments.

No statistically significant group differences were observed in HR variability (HR SD), with both groups showing comparable levels of HR fluctuation. These group-level findings are visually illustrated in Figure 5.15, which presents boxplots of the HR metrics across users and controls.

Per-task level analysis (Table 7.13 at Appendix) revealed patterns consistent with the overall group comparisons. Across nearly all simulation tasks, users maintained a higher mean heart rate than controls, especially during tasks requiring directional navigation (e.g., turning or reactive manoeuvring). The control group exhibited more pronounced heart rate decreases during simulator use, reflected by consistently more negative HR change values. These exploratory patterns suggest that users may have experienced greater task-related engagement or anticipatory stress, while controls appeared to adapt more quickly to simulator tasks..

Table 7.13 shows that while some tasks revealed statistically significant between-group differences in HR change and HR variability (SD), these effects were not consistent across all tasks. In some instances, users displayed significantly higher HR SD, potentially reflecting greater intra-individual heart rate fluctuations in response to simulator demands. This may suggest increased physiological arousal or adaptive responses to task complexity, particularly in later tasks. However, this pattern was not uniform. For example, in Task 8, the control group exhibited a higher mean HR SD, although the difference was not statistically significant. Such variability implies that HR SD may be influenced more by individual and task-specific factors, such as novelty or cognitive demand, than by consistent group-level differences in autonomic regulation.

Overall, HR SD values varied across tasks without a consistent group pattern, limiting their interpretive value as standalone indicators of physiological arousal. These task-level trends are visualized in Figure 5.16, which depicts the evolution of HR values from baseline through each simulation task for both groups.

In summary, simulator exposure elicited divergent cardiovascular responses in this pilot sample. Users tended to sustain elevated HR while controls showed decreases, suggesting potential differences in task engagement, cognitive load, or familiarity with the virtual environment. These findings highlight HR change as a potentially sensitive marker of physiological adaptation during virtual mobility assessment, but confirmatory conclusions cannot be drawn at this stage. Larger-scale studies are required to determine its robustness and clinical applicability.

Figure 5.15: Study 3 Boxplot for HR related metrics.

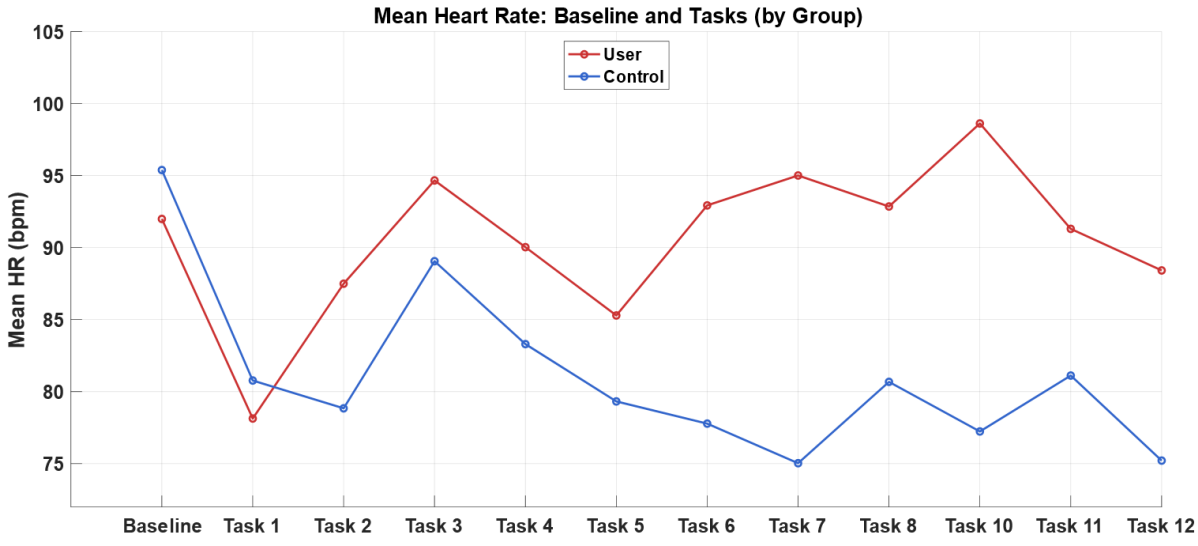


Figure 5.16: Study 3 HR line plot from baseline to tasks.

Table 5.13: Study 3 Summary of group-wise comparisons of HR at task level

Task	Metric	Group	Mean(SD)	Median [IQR]	U	Z	p-value	Effect size (r)
	HR Baseline	User	91.99 (13.68)	87.59 [81.56 - 106.71]	35.00	-0.44	0.696	0.10
		Control	95.39 (8.06)	97.1 [95.00 - 98.65]				
2	HR SD	User	3.38 (2.73)	2.89 [1.25 - 5.23]	15.00	-2.22	0.027*	0.52
		Control	2.24 (1.57)	1.92 [1.13 - 2.63]				
3	HR Task	User	94.67 (14.77)	94.83 [79.47 - 109.60]	14.00	-2.31	0.021*	0.54
		Control	89.06 (14.67)	91.37 [76.87 - 92.94]				
	HR SD	User	0.54 (0.5)	0.31 [0.18 - 0.94]	9.00	-2.75	0.004*	0.65
		Control	0.47 (0.47)	0.28 [0.09 - 0.71]				
4	HR Change	User	-1.96 (11.2)	-5.8 [-9.25 - 5.22]	14.00	-2.31	0.021*	0.54
		Control	-12.09 (12.6)	-13.63 [-21.05 - -5.89]				
5	HR Change	User	-6.7 (15.51)	-9.66 [-16.13 - 1.15]	17.00	-2.04	0.043*	0.48
		Control	-16.07 (12.27)	-18.59 [-27.88 - -6.83]				
6	HR Change	User	0.94 (11.37)	1.8 [-5.72 - 6.08]	10.00	-2.67	0.006	0.63
		Control	-17.61 (13.17)	-18.8 [-27.77 - -7.56]				
	HR SD	User	0.42 (0.32)	0.39 [0.12 - 0.66]	3.00	-3.29	0.000*	0.77
		Control	0.36 (0.36)	0.19 [0.08 - 0.62]				
7	HR Task	User	95.01 (17.74)	91.69 [81.88 - 106.67]	11.00	-2.58	0.009*	0.61
		Control	75.03 (9.07)	77.56 [64.50 - 81.16]				
	HR Change	User	3.02 (12.83)	1.67 [-5.48 - 14.72]	7.00	-2.93	0.002*	0.69
		Control	-20.36 (8.93)	-20.33 [-22.62 - -12.74]				
	HR SD	User	1.03 (1.59)	0.58 [0.20 - 0.80]	16.00	-2.13	0.034*	0.50
		Control	0.26 (0.23)	0.19 [0.08 - 0.39]				
8	HR Task	User	92.86 (18.65)	90.32 [78.84 - 111.00]	13.00	-2.40	0.016*	0.57
		Control	80.68 (11.73)	78.01 [73.98 - 84.11]				

*. The mean difference is significant at the 0.05 level.

5.1.11.10 Acceleration and Jerk (wrist control)

Empatica E4 wristband accelerometer data were used to explore movement dynamics during simulator tasks. Specifically, acceleration magnitude and jerk-based metrics, such as root mean square (RMS), mean and standard deviation, were calculated to describe the intensity, stability, and variability of upper limb motion during driving. Based on previous studies in the literature described at that uses accelerometers(e.g., smartphone-based) or amplitudes of joystick movements for estimating stability(smoothness of movement), these measures may provide indirect insights into participants' motor control and interaction style while using the joystick(P. S. Archambault et al., 2012; Gacem et al., 2020). The general hypothesis guiding this exploratory analysis was that experienced wheelchair users would show more stable and efficient motion patterns, while non-users might demonstrate more abrupt or inconsistent control due to lack of familiarity.

Although group-level differences at the session level were not statistically significant, task-specific analysis revealed noteworthy effects. In Task 1, control group exhibited significantly higher RMS jerk ($p = 0.034$, $r = 0.50$), suggesting more forceful or sudden movements early in the session. These differences were not consistent across all tasks, potentially reflecting either early-stage adaptation by the control group or a stable regulation pattern among experienced users.

The progression of these wrist movement metrics across tasks is presented in Figure 5.17 (boxplots) and Figure 5.18 (line plots), while detailed statistical results are available in Table 7.14 (group-level comparisons) and Table 7.15 (per-task comparisons) at Appendix.

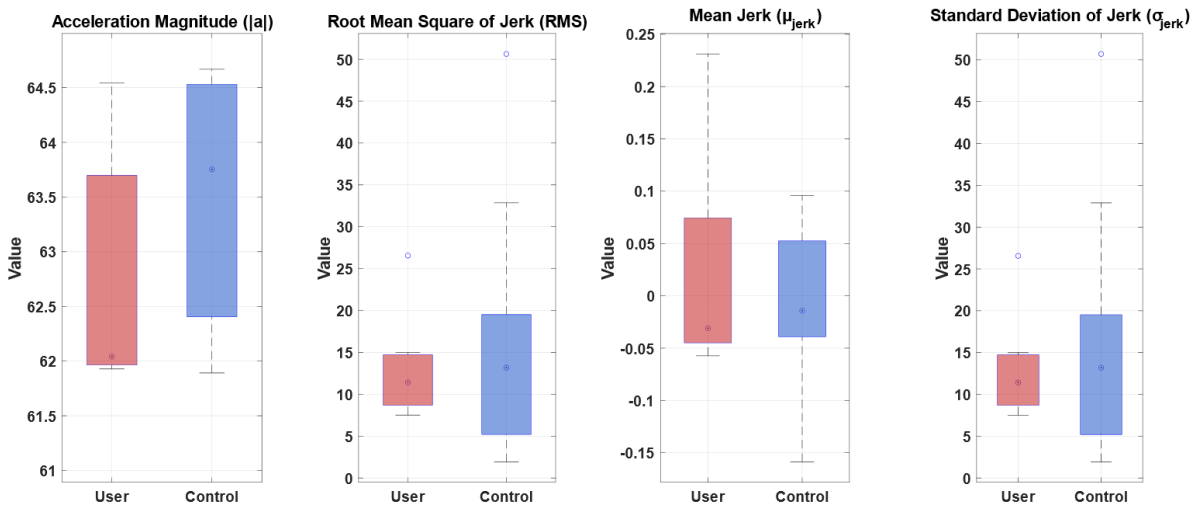


Figure 5.17: Study 3 Wrist Acceleration and Jerk related metrics at Session level.

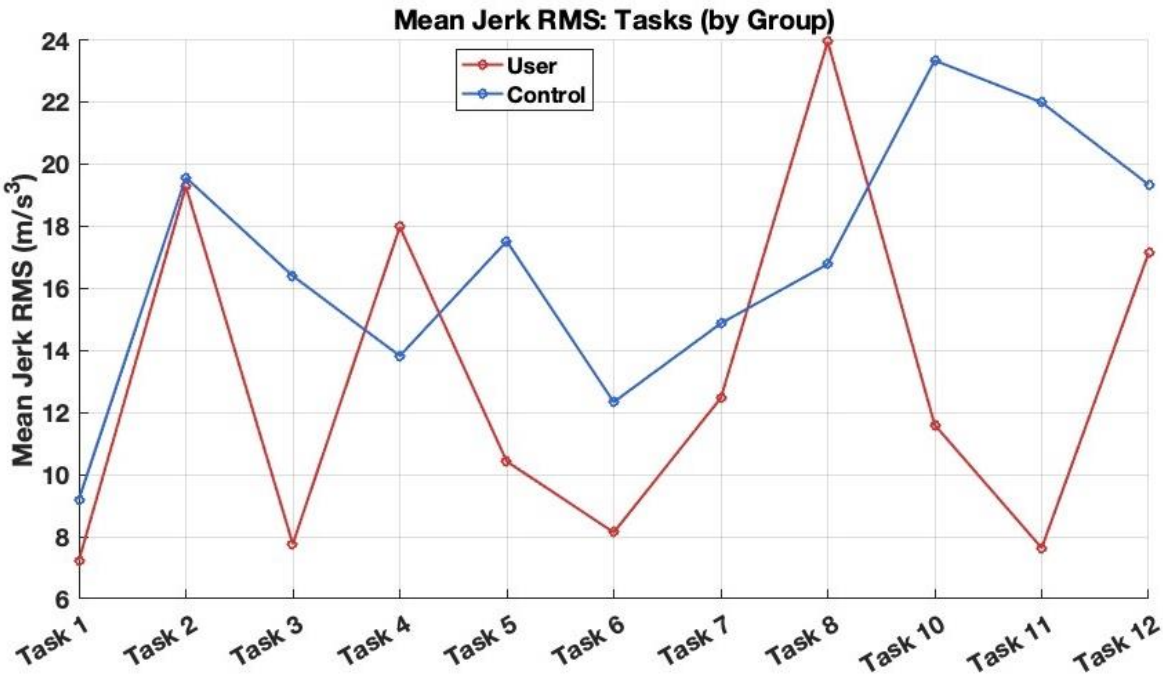


Figure 5.18: Study 3 Wrist Acceleration and Jerk related metrics line at tasks level.

5.1.11.11 Exploratory correlation analysis between Simulator-based performance metrics and clinical assessments (MoCA and WST-Q)

The Spearman correlation matrix (Figure 5.19) shows that among the simulator-derived performance metrics, MoCA scores were significantly negatively correlated with Total Collisions and Mean RMSE. These results suggest that participants with higher cognitive scores tended to perform more efficiently in the simulator, indicating the potential of simulator-based metrics to reflect cognitive functioning relevant to powered mobility.

However, the matrix also reveals no positive correlation, and in some cases, even weak negative correlations, between MoCA scores and WST-Q confidence ratings, a finding that diverges from earlier studies in the literature using traditional clinical assessments (Pellichero, Best, et al., 2021; Pellichero, Kenyon, et al., 2021), where higher cognitive scores (MoCA) were generally linked to greater self-reported mobility confidence (WST-Q), study with N = 30. In the current mixed sample, this pattern may reflect the presence of two distinct participant groups: experienced wheelchair users with some degree of cognitive impairment (lower MoCA but higher confidence) and able-bodied participants with higher cognitive scores but little or no wheelchair experience (higher MoCA but lower confidence). This group contrast likely suppresses or inverts the expected positive association between cognitive ability and perceived confidence, emphasizing the importance of sample stratification in future work. To address these effects and strengthen the interpretation of simulator-

based assessments, future studies should aim to increase sample size and consider analysing wheelchair users and able-bodied participants separately, particularly when interpreting subjective outcomes like confidence.

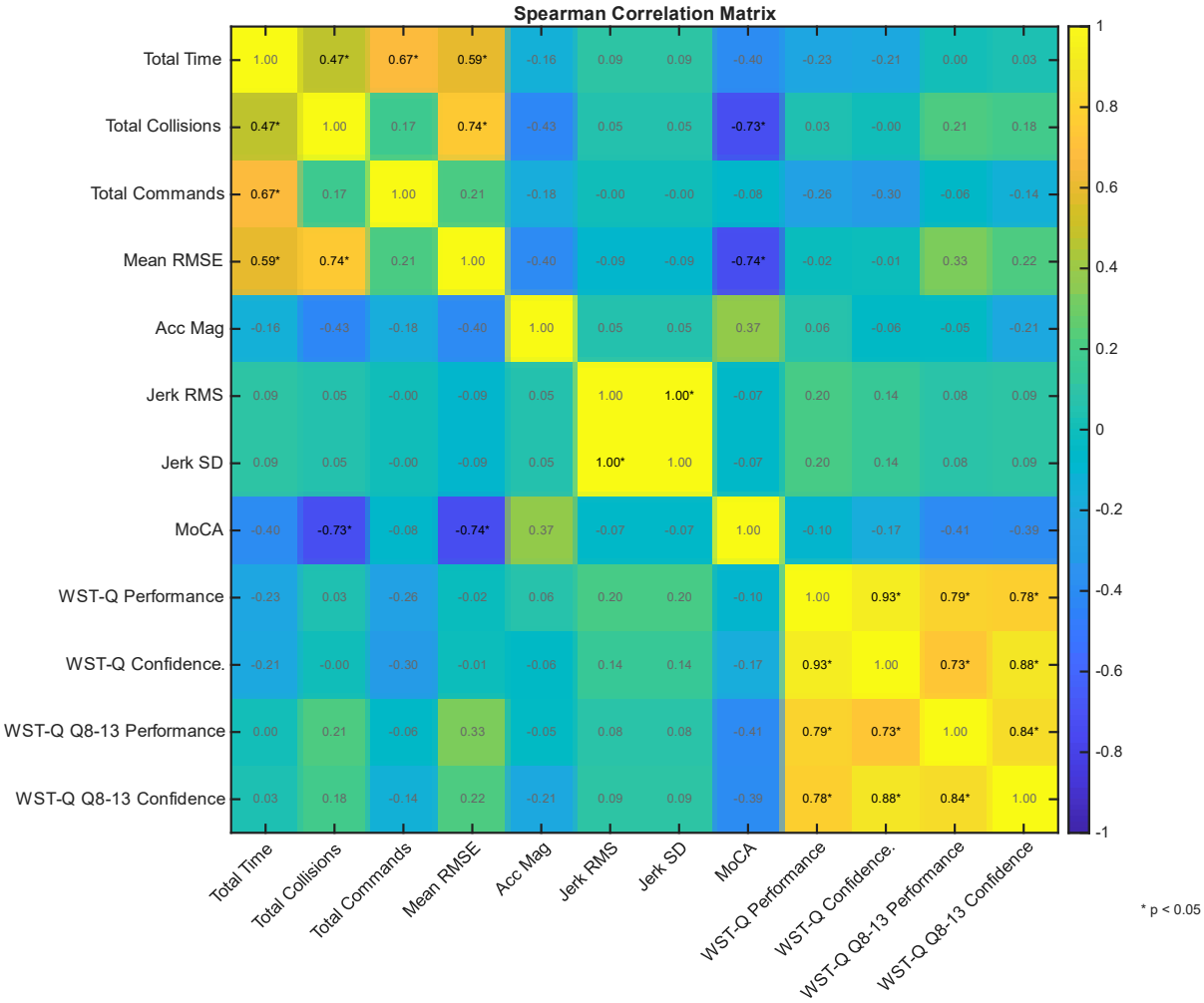


Figure 5.19: Spearman correlation matrix showing significant negative correlations between MoCA and both Collisions and RMSE.

Summary

In summary, these findings not only illustrate the challenges of interpreting simulator-based data in mixed populations but also highlight the broader need for structured, adaptable approaches that can accommodate user diversity and contextual variation. Recognising these challenges, the next section introduces a conceptual framework developed as a direct response to the insights gained through this study. This framework, EMPOWER-SIM, aims to provide a multidimensional foundation for designing, implementing, and evaluating simulator-based interventions in the context of powered mobility provision.

5.2 EMPOWER-SIM: Preliminary guidelines for simulator-based assessment

This section presents the EMPOWER-SIM (Empowering Power Wheelchair Users through Simulation) framework, , a preliminary framework shaped by the findings of the lab and pilot field studies. At this stage, EMPOWER-SIM should not be regarded as a validated clinical framework, but rather as a conceptual model and set of provisional guidelines. Its purpose is to capture the lessons learned from feasibility testing and translate them into practical directions for future research, protocol development, and validation work. Figure 5.20 outlines the four domains of EMPOWER-SIM, each reflecting themes that emerged during the studies. These domains should be interpreted as starting points rather than final recommendations.

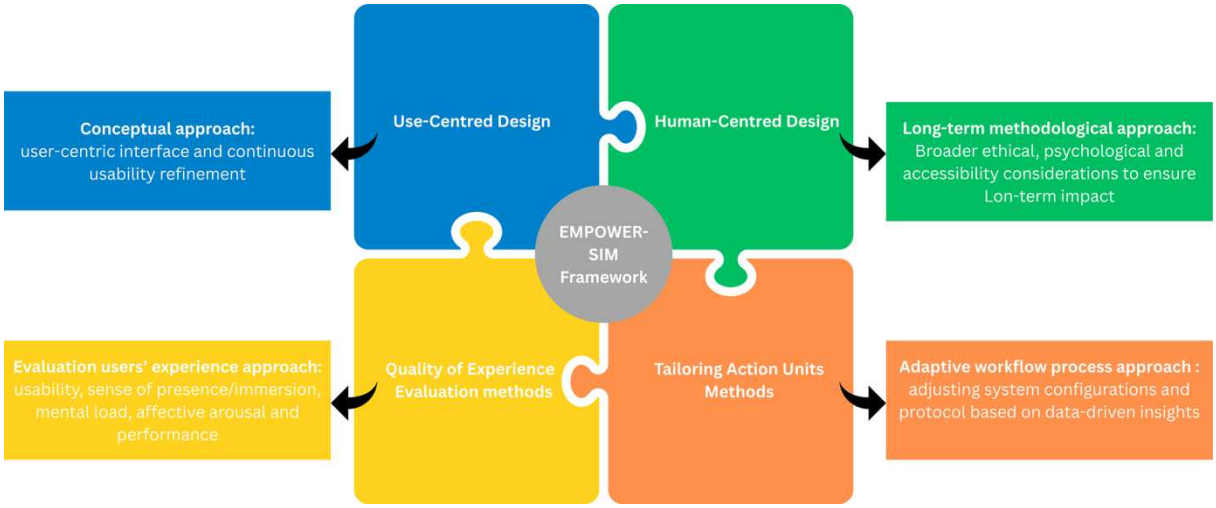


Figure 5.20: EMPOWER-SIM Framework core elements.

1.User-Centred Design (UCD domain)

The User-Centred Design (UCD) addresses the initial design and usability aspects of simulator development. It focuses on accessibility, task clarity, and interaction design. Lab findings highlighted the need for clear task instructions, adjustable immersion levels, and joystick settings that could be tailored to individual comfort. These insights provide preliminary design considerations but require further testing with larger and more diverse user groups.

2. Human-Centred Design (HCD domain)

In the Human-Centred Design (HCD) domain introduces a broader methodological perspective focused on long-term impact. It supports ongoing development through participatory engagement with diverse stakeholders (e.g., wheelchair users, healthcare professionals, and support staff) alongside

iterative prototyping and inclusive feedback mechanisms. These experiences suggest that meaningful integration depends on participatory design and contextual awareness, though the evidence base is still limited to small-scale pilot testing..

3. Quality of Experience (QoE domain)

QoE is central to understanding how users interact with the simulator. Across studies, self-report tools (SUS, IPQ, NASA-TLX, SAM) and physiological measures (EDA, HR) were collected. While these data provided useful insights into usability, workload, and emotional engagement, the small sample sizes and exploratory analyses mean that patterns should be viewed as indicative rather than conclusive.

4. Tailoring Actions domain

Tailoring Actions domain highlights the need for adjustable simulator settings (e.g., control sensitivity, task difficulty). Pilot findings showed variability in cognitive abilities, prior experience, and confidence among participants. These differences suggest that fixed configurations are unlikely to suit all users, reinforcing the importance of personalisation. At present, tailoring within EMPOWER-SIM is conceptual; robust adaptive mechanisms remain to be developed and tested.

5.2.1 EMPOWER-SIM: Proposed Clinical Workflow and Future Directions

Figure 5.21, presents a draft workflow illustrating how EMPOWER-SIM might be applied in practice. This process is structured into three interconnected domains:

1. Understanding the User

This initial phase involves identifying the user's position within the wheelchair provision continuum (e.g., new user, experienced user, or rehabilitation phase). Information is gathered on individual needs, preferences, and goals to guide intervention planning. This stage ensures a comprehensive understanding of the user, forming the basis for all subsequent tailoring steps.

2. Adapting Support

In this phase, the simulator is iteratively configured based on the user profile. Task scenarios, control modalities, and settings are adapted. QoE is assessed through a combination of self-reported and physiological feedback. These insights guide ongoing refinements to optimise usability, performance, and engagement.

3. Navigating Challenges

Based on the insights from performance data and user feedback, additional adjustments are made. These may include simulator configuration changes, personalised training plan updates, or referrals

to external supports such as occupational therapy or home modifications. The process is cyclical, allowing for continuous response to evolving needs.

While useful as a conceptual guide, this workflow is an early proposal that requires refinement and empirical validation before it can inform clinical protocols. The framework currently serves three exploratory functions:

- Assessment: simulator tasks estimate users' mobility and cognitive abilities aligned with stablished clinical assessment tools.
- Training: task complexity can be gradually increased based on performance reports.
- Follow-up: the same standardised tasks can be repeated to monitor progress or identify emerging support needs.

In summary, EMPOWER-SIM s presented here as an early, feasibility-informed framework. It captures preliminary lessons from lab and field studies and translates them into provisional guidelines. At this stage, EMPOWER-SIM should be regarded as a stepping stone toward more structured protocols and larger validation studies, not as a ready-to-use clinical tool.

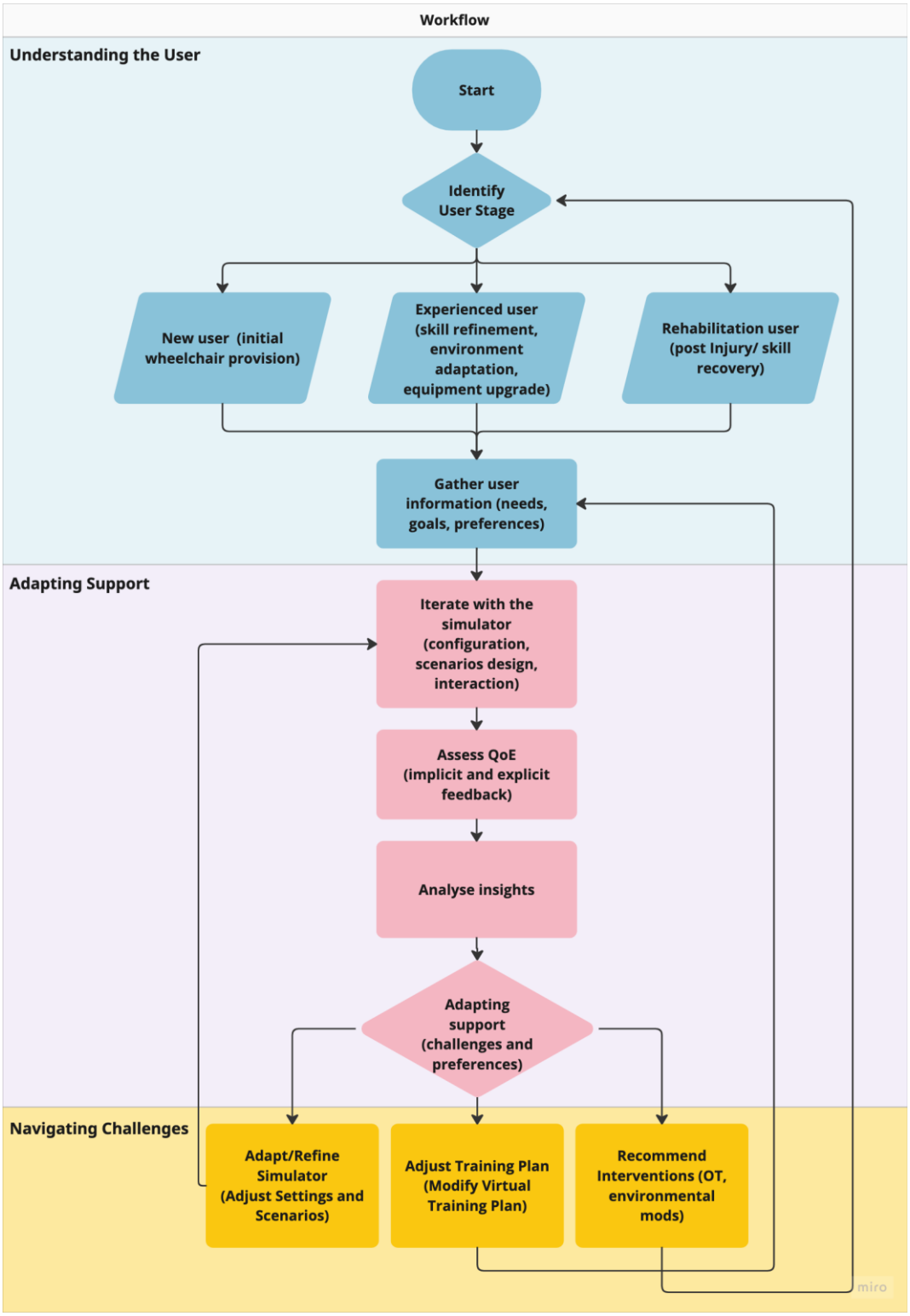


Figure 5.21: EMPOWER-SIM Framework workflow.

5.3 Discussion

This pilot study examined the feasibility of using a desktop-based wheelchair simulator for preliminary assessment and training across a small, mixed sample of powered wheelchair users and non-disabled controls. The study integrated subjective, behavioural, and physiological measures to explore whether simulator interactions could provide meaningful insights into users' cognitive, emotional, and functional responses.

The study demonstrated that it was possible to recruit participants, embed simulator sessions into community centres, and collect a multidimensional dataset. This confirms that the simulator can be deployed in real-world assistive contexts, although challenges such as heterogeneous user profiles, technical issues with physiological data capture, and modest ecological fidelity of the desktop platform were evident. The findings should therefore be interpreted as early indicators rather than confirmatory evidence.

5.3.1 Key Findings

Simulator-derived metrics, such as collisions and trajectory deviation (RMSE), appeared sensitive to differences between groups, with effect sizes suggesting meaningful contrasts. Although task completion time did not differ significantly, the trend toward longer durations among wheelchair users was consistent with their diverse clinical backgrounds and mobility needs. These results indicate that simulator metrics may capture clinically relevant aspects of mobility performance, though replication with larger samples is necessary.

Cognitive workload assessments showed no significant group-level differences on global measures (NASA-TLX), but task-specific ratings (Paas scale) suggested that some users experienced higher mental effort during later tasks. This pattern underscores the value of combining session-level and task-level measures to detect subtle differences in demand over time, though these observations remain tentative given the small sample.

Emotional responses measured via SAM were neutral to positive, with no significant group differences, suggesting that the simulator was acceptable and did not induce undue stress. This supports its feasibility for use in extended assessments or training, but further study is needed to understand how emotional engagement may change with more immersive setups.

Heart rate analysis revealed divergent patterns: users maintained elevated HR while controls showed decreases, with HR change differences reaching statistical significance. While this finding is consistent with literature linking HR responses to task engagement and challenge (Zorzi et al., 2023a), the small

sample and variability mean that these results should be treated as exploratory. They do, however, highlight the feasibility of capturing cardiovascular responses in this context.

Wrist-based jerk and acceleration metrics were also explored but did not yield consistent group differences. Variability across tasks suggests that such measures may be influenced by novelty, adaptation, or individual strategies. This reinforces the need for more systematic exploration of motor interaction metrics with larger datasets and refined instrumentation.

Importantly, exploratory correlations revealed that higher cognitive scores (MoCA) were associated with lower simulator collisions and RMSE, suggesting that simulator metrics may align with real-world cognitive functioning. However, the expected positive association between MoCA and WST-Q confidence (Pellichero, Best, et al., 2021; Pellichero, Kenyon, et al., 2021) was not observed, likely due to the contrasting profiles of experienced users with cognitive impairments and controls with no mobility experience but intact cognitive function. This highlights the interpretive complexity of mixed samples and suggests that future work should increase the sample size and stratify participants by mobility and cognitive status.

5.3.2 Limitations

Several limitations should be acknowledged when interpreting the findings of this pilot study. First, the sample size was small and not representative of the broader powered wheelchair user population. The heterogeneity of the user group, a range of primary conditions, cognitive abilities, and wheelchair experience, contributed to high intra-group variability, which may have reduced the statistical power to detect subtler effects. While this diversity reflects real-world clinical contexts, it complicates direct comparisons and limits the generalisability of findings.

Second, the mixed population, which included both wheelchair users and non-disabled controls, introduces interpretive complexity. Although useful for identifying broad group differences, the contrasting profiles between participants, particularly in terms of mobility experience and cognitive function may have influenced subjective responses such as perceived workload and confidence. This may explain the absence of an expected correlation between MoCA and WST-Q confidence, and underscores the need for stratified analyses in future studies.

Third, while physiological and wrist motion data added valuable depth to the assessment, however, signal quality limitations (e.g., sensor loss) and the exploratory nature of the analysis meant that these measures were not consistently robust across all participants or tasks. The wrist-worn accelerometers,

in particular, captured only one dimension of motor interaction and may not fully reflect joystick handling subtleties or compensatory strategies.

Finally, the desktop-based nature of the simulator limited immersion and sensory fidelity, which may have influenced user engagement and presence scores. Although this platform was selected for preventing cybersickness, its reduced realism compared to immersive VR systems may constrain its ability to fully reflect real-world challenges of powered mobility.

5.4 Summary

The pilot study demonstrated that a desktop-based wheelchair simulator is feasible for assessing cognitive, emotional, and functional responses. By combining clinical assessments, performance metrics, and physiological signals, the study provided initial evidence that the simulator can reflect individual differences. While the results should be interpreted cautiously due to the small sample size and variability, key findings suggest the simulator's potential as a complementary assessment tool.

The study's findings directly informed the development of initial concept of the EMPOWER-SIM framework, which is proposed as a structured, person-centred model for future research and development. This framework is not a validated framework but offers a direction for enhancing the clinical relevance, accessibility, and personalization of mobility simulation technologies.

Part V CONCLUSION

Chapter 6 Conclusion and Future Work

6.1 Conclusion

This PhD research investigated the integration of a virtual wheelchair simulator into clinical and real-world assessment and training. The work's core objective was to enhance the objectivity, personalization, and clinical relevance of simulator-based tools by proposing an evaluation framework and clinical pilot study protocol. By employing multimodal user evaluation, this research directly addressed a key gap in the field: the weak alignment of simulator metrics with established clinical tools. This was achieved by integrating simulator tasks that mirrored those in the Wheelchair Skills Test (WST) and Power Mobility Road Test (PMRT), and by examining the role of cognitive function using the MoCA.

The project followed a structured progression, beginning with the development and refinement of a proof-of-concept simulator system in controlled laboratory settings. This was followed by a field pilot feasibility study involving power wheelchair users and clinical partners in IWA centres. Across these phases, A multidimensional dataset was collected across these phases, combining subjective feedback, performance metrics, physiological responses, and wrist kinematics. This multimodal approach was a direct effort to address another critical gap about the underuse of wearable data as objective indicators of user response.

The findings from both phases support that simulator-based metrics, such as task performance, driving errors, and cognitive load, can meaningfully reflect user profiles and functional differences. While challenges related to sample heterogeneity, motion artifacts in physiological data, and limited statistical power were encountered, the overall results support the feasibility, acceptability, and potential clinical utility of simulator-based interventions. These outcomes informed the development of EMPOWER-SIM framework concept, proposed here as a preliminary, feasibility-informed set of guidelines to support tailored, QoE-informed simulator use in power mobility assessment and training.

6.2 Reflection of Research Questions

This research was guided by the following overarching research question:

> ***"How can a virtual wheelchair simulator be integrated into clinical settings—using Irish Wheelchair Association (IWA) centres as a use case—for power mobility training and assessment, by defining protocols and metrics that support structured, safe, and clinically applicable use across a diverse population of power wheelchair users?"***

To address this question, the research was structured around two sub-research questions (SRQs) and a set of specific objectives that were progressively explored across controlled laboratory and field-based studies.

> ***SRQ1: How can a virtual wheelchair simulator be designed and tested in a controlled environment to establish a clinically relevant proof of concept that supports multidimensional assessment, incorporating immersive technologies, physiological signals, subjective feedback, and Quality of Experience (QoE) evaluation?***

The first component of the research investigated how a virtual wheelchair simulator could be designed and tested in a controlled environment. Through lab-based studies, the project evaluated the effects of immersive display settings, motion dynamics, and multimodal feedback on user experience. Subjective (e.g., usability, emotional response, cognitive workload) and objective (e.g., electrodermal activity, heart rate) measures were collected to establish a multidimensional Quality of Experience (QoE) assessment approach. The findings indicated that immersive design features, particularly combining smooth motion profiles and headset-based displays, contributed to improved user comfort and engagement while helping to manage simulator-induced discomfort. These studies addressed SRQ1 and demonstrated the potential for capturing relevant experiential and physiological responses through virtual tasks.

> ***SRQ2: How can the proof-of-concept simulator be transferred into clinical settings, using Irish Wheelchair Association (IWA) centres as a use case, and how can protocols and evaluation methods be developed to test the feasibility its components, reflect the perspectives of wheelchair users, and support standardised implementation?***

The second component focused on how the proof-of-concept simulator could be transferred into clinical settings. A field pilot study conducted across two IWA centres examined the feasibility, acceptability, and sensitivity of the simulator-based protocol when applied with real-world users. This phase addressed SRQ2 and demonstrated that the system could be embedded into routine

rehabilitation activities without disruption. Comparative analyses between simulator-based performance metrics and conventional assessment tools (e.g., Wheelchair Skills Test and Montreal Cognitive Assessment) showed initial evidence of alignment, particularly in the relationship between cognitive scores and simulator driving performance. Moreover, the pilot study collected feedback from participants and capture practical considerations for simulator-based task standardisation and tailoring actions, which informed a preliminary evaluation framework model, EMPOWER-SIM, and clinical pilot study protocol.

Overall, the results from both study phases confirm that simulator-based assessment could potentially support structured and clinically applicable evaluation of power mobility skills when grounded in QoE principles and designed with real-world contexts in mind. The integration of multimodal metrics, subjective feedback, and task-based data within a tailored protocol aligns with current clinical practices while offering new opportunities for objective, repeatable, and user-centred assessment. While limitations remain, such as sample size and generalisability, the findings support the potential of the simulator as a complementary tool in power wheelchair provision and training services. The main research question was therefore addressed through an iterative and multi-perspective investigation, resulting in both technical contributions and a conceptual framework to guide future clinical interventions.

6.3 Limitations

While this PhD research offers meaningful contributions toward the integration of virtual wheelchair simulators into clinical contexts, several limitations must be acknowledged. First, the sample size across studies was modest, particularly in the clinical field study. Although mixed-methods designs and multimodal metrics strengthened the depth of analysis, small participant numbers limit statistical power and generalisability. Recruiting powered wheelchair users with diverse clinical conditions added heterogeneity, which, while reflective of real-world variability, may have also introduced confounding effects difficult to control with limited sample stratification.

Second, the early laboratory studies relied solely on able-bodied participants, whose familiarity with digital systems (e.g., video games) may not accurately reflect the target clinical population. This approach was necessary for initial feasibility testing and system calibration but may have limited the ecological validity of certain findings during those phases.

Third, physiological data collection, particularly electrodermal activity and heart rate, was affected by motion artefacts and sensor reliability, largely due to the characteristics of the wearable devices used. Although wearables were chosen for non-intrusiveness and field applicability, data quality varied

across sessions, leading to the exclusion of some samples and limiting fine-grained physiological analyses. Additionally, due to practical constraints, repeated-measures designs were not implemented, preventing within-subject tracking of learning curves and longitudinal adaptation.

Finally, while the simulator was aligned with validated clinical assessments (e.g., PMRT, WST and MoCA), direct clinical validation of diagnostic accuracy or predictive validity was beyond the scope of this exploratory work. The current framework supports feasibility and acceptability but further validation is needed to support clinical decision-making.

6.4 Future Work & Research Opportunities

Building upon the findings of this PhD, some future research directions are proposed to support the clinical translation and broader validation of simulator-based assessments in power mobility provision. These next steps are designed to move the research from a conceptual framework to a robust, clinically validated tool.

An immediate next step involves conducting a follow-up feasibility study. This will first require the recruitment of additional powered wheelchair users to create a more robust dataset. Following data collection, healthcare professionals at IWA centres will review the records to provide clinical insights and evaluate the simulator's metrics. Clinicians could review previously collected simulator-based records from powered wheelchair users and evaluate each participant’s driving performance using the Power Mobility Road Test (PMRT) assessment tool. This retrospective analysis will not only provide clinical insights into the interpretability of simulator metrics but also enable a preliminary investigation of inter-rater reliability. By assessing agreement between professional raters using the same simulator-derived data, this phase will test the robustness of the evaluation protocol and inform refinements in metric presentation and reporting.

This healthcare professional review phase complements the dataset already collected in Study 3 (Phase 1) and aims to complete a two-part feasibility assessment. Together, data from wheelchair users and clinicians will support the design of a full-scale clinical trial. This future trial is envisioned as a multicentre study implemented across Irish community and rehabilitation centres, applying a stratified sampling strategy to capture a diverse range of user needs, diagnoses, and cognitive-motor profiles. Such a trial would allow systematic validation of the simulator as a clinical assessment and training tool, including its usability, acceptability, and alignment with conventional clinical instruments.

In parallel, future research should incorporate a repeated-measures design to examine the effects of task repetition, learning curves, and simulator responsiveness to intervention or user progression. These designs would strengthen the evidence base for the simulator's use in both short-term assessment and long-term training pathways.

In addition, the dataset collected in Study 3 includes multimodal data that were not yet fully analysed, such as head movement patterns and additional physiological signals (e.g., electrodermal activity, inter-beat intervals). These data offer further opportunities to investigate behavioural and physiological indicators of user engagement, stress, or cognitive demand during simulator tasks. Future analyses may contribute to refining the evaluation framework by identifying new metrics or confirming the relevance of existing ones. However, these investigations will require careful preprocessing and validation, particularly given the presence of motion artefacts and individual variability. Addressing these technical challenges in future work may enhance the interpretability of implicit signals and support their integration into more comprehensive and clinically relevant user profiles.

Altogether, these future directions aim to build upon the technical, clinical, and methodological foundations established in this thesis, contributing to the gradual development of a more standardised and adaptable framework for simulator-based assessment and training within powered mobility technologies.

REFERENCES

A. R. de Sá, A., Morère, Y., & Naves, E. L. M. (2022). Skills assessment metrics of electric powered wheelchair driving in a virtual environment: a survey. *Medical & Biological Engineering & Computing Volume*, *60*, 323–335. https://doi.org/https://doi.org/10.1007/s11517-022-02500-8

Abbas, J. R., Chu, M. M. H., Jeyarajah, C., Isba, R., Payton, A., McGrath, B., Tolley, N., & Bruce, I. (2023). Virtual reality in simulation-based emergency skills training: A systematic review with a narrative synthesis. *Resuscitation Plus*, *16*, 100484. https://doi.org/10.1016/J.RESPLU.2023.100484

Abeele, V. Vanden, Nacke, L. E., Mekler, E. D., & Johnson, D. (2016). Design and preliminary validation of the player experience inventory. *CHI PLAY 2016 - Proceedings of the Annual Symposium on Computer-Human Interaction in Play Companion*, 335–341. https://doi.org/10.1145/2968120.2987744;CTYPE:STRING:BOOK

Abeele, V. Vanden, Spiel, K., Nacke, L., Johnson, D., & Gerling, K. (2020). Development and validation of the player experience inventory. *International Journal of Human Computer Studies*, *135*. https://doi.org/10.1016/J.IJHCS.2019.102370

Agency for Healthcare Research and Quality. (2024). *Fundamental Concepts for Understanding Cognitive Load*. https://www.ahrq.gov/diagnostic-safety/resources/issue-briefs/dxsafety-cognitive-load2.html

Ahmad, Z., & Khan, N. (2022). A Survey on Physiological Signal-Based Emotion Recognition. *Bioengineering*, *9*(11). https://doi.org/10.3390/bioengineering9110688

Alshaer, A., Hoermann, S., & Regenbrecht, H. (2013). Influence of peripheral and stereoscopic vision on driving performance in a power wheelchair simulator system. *2013 International Conference on Virtual Rehabilitation, ICVR 2013*, 152–164. https://doi.org/10.1109/ICVR.2013.6662066

Alshaer, A., O'Hare, D., Hoermann, S., & Regenbrecht, H. (2016). The impact of the visual representation of the input device on driving performance in a power wheelchair simulator. *In 11th International Conference on Disability, Virtual Reality & Associated Technologies (ICDVRAT)*.

Alshaer, A., Regenbrecht, H., & O'Hare, D. (2015). Investigating visual dominance with a virtual driving task. *2015 IEEE Virtual Reality Conference, VR 2015 - Proceedings*, 145–146. https://doi.org/10.1109/VR.2015.7223337

Amudha, A., & B. William Dharma, R. (2016). Efficient E-Learning by Dint of Cognitive Abilities. *I-Manager's Journal of Educational Technology*, *13*(2), 7. https://doi.org/10.26634/JET.13.2.8159

ANVISA. (2022). *RDC 657/2022*. https://www.gov.br/anvisa/pt-br/assuntos/produtosparasaude/temas-em-destaque/arquivos/2024/rdc-657-2022-en.pdf

Archambault, P., Gagnon, D., Routhier, F., & Miller, W. (2016). Effectiveness of power wheelchair simulator training, delivered at home, on wheelchair driving skills. *Annals of Physical and Rehabilitation Medicine*, *59*, e37–e38. https://doi.org/10.1016/J.REHAB.2016.07.087

Archambault, P. S., Blackburn, É., Reid, D., Routhier, F., & Miller, W. C. (2017). Development and user validation of driving tasks for a power wheelchair simulator. *Disability and Rehabilitation*, *39*(15), 1549–1556. https://doi.org/10.1080/09638288.2016.1226423;WEBSITE:WEBSITE:TFOPB;PAGEGROUP:STRING:PUBLICATION

Archambault, P. S., Chong, J. N. F., Sorrento, G., Routhier, F., & Boissy, P. (2011). Comparison of powered wheelchair driving performance in a real and in a simulated environment. *2011 International Conference on Virtual Rehabilitation, ICVR 2011*. https://doi.org/10.1109/ICVR.2011.5971807

Archambault, P. S., Routhier, F., Hamel, M., & Boissy, P. (2008). Analysis of movement to develop a virtual reality powered- Wheelchair simulator. *2008 Virtual Rehabilitation, IWVR*, 133–138. https://doi.org/10.1109/ICVR.2008.4625149

Archambault, P. S., Tao, G., Torkia, C., Boissy, P., Lemay, M., Reid, D., Routhier, F., Ryan, S. E., & Woodhouse, J. (2013). Development of a new virtual environment for a power wheelchair simulator: A user-centered approach. *2013 International Conference on Virtual Rehabilitation, ICVR 2013*, 216–217. https://doi.org/10.1109/ICVR.2013.6662129

Archambault, P. S., Tremblay, S., Cachecho, S., Routhier, F., & Boissy, P. (2012). Driving performance in a power wheelchair simulator. *Disability and Rehabilitation: Assistive Technology*, *7*(3), 226–233. https://doi.org/10.3109/17483107.2011.625072

Arlati, S., Colombo, V., Ferrigno, G., Sacchetti, R., & Sacco, M. (2020). Virtual reality-based wheelchair simulators: A scoping review. In *Assistive Technology* (Vol. 32, Issue 6, pp. 294–305). Bellwether Publishing, Ltd. https://doi.org/10.1080/10400435.2018.1553079

Ashley, D., Ashley, K., Alqasemi, R., & Dubey, R. (2017). Semi-autonomous mobility assistance for power wheelchair users navigating crowded environments. *IEEE International Conference on Rehabilitation Robotics*, 1025–1030. https://doi.org/10.1109/ICORR.2017.8009384

Athif, M., Rathnayake, B. L. K., Nagahapitiya, S. M. D. B. S., Samarasinghe, S. A. D. A. K., Samaratunga, P. S., Peiris, R. L., & De Silva, A. C. (2020). Using Biosignals for Objective Measurement of Presence in Virtual Reality Environments. *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS*, *2020-July*, 3035–3039. https://doi.org/10.1109/EMBC44109.2020.9176022

Augmented Reality and Virtual Reality in Medical Devices | FDA. (n.d.). Retrieved March 27, 2025, from https://www.fda.gov/medical-devices/digital-health-center-excellence/augmented-reality-and-virtual-reality-medical-devices#list

Azbel-Jackson, L., Butler, L. T., Ellis, J. A., & van Reekum, C. M. (2016). Stay calm! Regulating emotional responses by implementation intentions: Assessing the impact on physiological and subjective arousal. *Cognition and Emotion*, *30*(6), 1107–1121. https://doi.org/10.1080/02699931.2015.1049515;PAGE:STRING:ARTICLE/CHAPTER

Balk, S. A., Bertola, M. A., & Inman, V. W. (2013). Simulator Sickness Questionnaire: Twenty Years Later. *Proceedings of the 7th International Driving Symposium on Human Factors in Driver Assessment, Training, and Vehicle Design : Driving Assessment 2013*, 257–263. https://doi.org/10.17077/drivingassessment.1498

Baltrusaitis, T., Robinson, P., & Morency, L. P. (2016). OpenFace: An open source facial behavior analysis toolkit. *2016 IEEE Winter Conference on Applications of Computer Vision, WACV 2016*. https://doi.org/10.1109/WACV.2016.7477553

Baltrusaitis, T., Zadeh, A., Lim, Y. C., & Morency, L. P. (2018). OpenFace 2.0: Facial behavior analysis toolkit. *Proceedings - 13th IEEE International Conference on Automatic Face and Gesture Recognition, FG 2018*. https://doi.org/10.1109/FG.2018.00019

Baños, R. M., Botella, C., Alcañiz, M., Liaño, V., Guerrero, B., & Rey, B. (2004). Immersion and emotion: Their impact on the sense of presence. *Cyberpsychology and Behavior*, *7*(6), 734–741. https://doi.org/10.1089/CPB.2004.7.734;PAGEGROUP:STRING:PUBLICATION

Banu, A. R. S., & Nagaveni, V. (2023). Assessment of Sympathetic and Parasympathetic Activities of Nervous System from Heart Rate Variability Using Machine Learning Techniques. *SN Computer Science*, *4*(5), 1–10. https://doi.org/10.1007/S42979-023-02062-Y/METRICS

Bañuelos-Lozoya, E., González-Serna, G., González-Franco, N., Fragoso-Diaz, O., & Castro-Sánchez, N. (2021). A Systematic Review for Cognitive State-Based QoE/UX Evaluation. *Sensors 2021, Vol. 21, Page 3439*, *21*(10), 3439. https://doi.org/10.3390/S21103439

Benaim, C., Wauquiez, G., Pérennou, D., Piscicelli, C., Lucas-Pineau, B., Bonnin-Koang, H. Y., Vuadens, P., Binquet, C., Bourredjem, A., & Devilliers, H. (2022). Cognitive assessment scale for stroke patients (CASP): A multicentric validation study. *Annals of Physical and Rehabilitation Medicine*, *65*(3), 101594. https://doi.org/10.1016/J.REHAB.2021.101594

Benedek, M., & Kaernbach, C. (2010). A continuous measure of phasic electrodermal activity. *Journal of Neuroscience Methods*, *190*(1), 80–91. https://doi.org/10.1016/J.JNEUMETH.2010.04.028

Benford, F. (2017). Use of powered mobility for a young adult with profound and multiple learning disabilities: a practice analysis. *British Journal of Occupational Therapy*, *80*(8), 517–520. https://doi.org/10.1177/0308022617698169

Bigras, C., Kairy, D., & Archambault, P. S. (2019). Augmented feedback for powered wheelchair training in a virtual environment. *Journal of NeuroEngineering and Rehabilitation*, *16*(1), 1–12. https://doi.org/10.1186/S12984-019-0482-3/FIGURES/11

Bigras, C., Owonuwa, D. D., Miller, W. C., & Archambault, P. S. (2020). A scoping review of powered wheelchair driving tasks and performance-based outcomes. *Disability and Rehabilitation: Assistive Technology*, *15*(1), 76–91. https://doi.org/10.1080/17483107.2018.1527957

Blanco-Ríos, M. A., Candela-Leal, M. O., Orozco-Romo, C., Remis-Serna, P., Vélez-Saboyá, C. S., Lozoya-Santos, J. de J., Cebral-Loureda, M., & Ramírez-Moreno, M. A. (2024). Real-time EEG-based emotion recognition for neurohumanities: perspectives from principal component analysis and tree-based algorithms. *Frontiers in Human Neuroscience*, *18*, 1319574. https://doi.org/10.3389/FNHUM.2024.1319574/BIBTEX

Bowen, D. J., Kreuter, M., Spring, B., Cofta-Woerpel, L., Linnan, L., Weiner, D., Bakken, S., Kaplan, C. P., Squiers, L., Fabrizio, C., & Fernandez, M. (2009). How We Design Feasibility Studies. *American Journal of Preventive Medicine*, *36*(5), 452–457. https://doi.org/10.1016/j.amepre.2009.02.002

Bowman, D. A., Gabbard, J. L., & Hix, D. (2002). A Survey of Usability Evaluation in Virtual Environments: Classification and Comparison of Methods. *Presence: Teleoperators and Virtual Environments*, *11*(4), 404–424. https://doi.org/10.1162/105474602760204309

Bradley, M. M., & Lang, P. J. (1994). Measuring emotion: The self-assessment manikin and the semantic differential. *Journal of Behavior Therapy and Experimental Psychiatry*. https://doi.org/10.1016/0005-7916(94)90063-9

Brooke, J. (1996). SUS - A quick and dirty usability scale. In *Usability Evaluation in Industry* (p. 252). CRC Press.

Brown, P., Spronck, P., & Powell, W. (2022). The simulator sickness questionnaire, and the erroneous zero baseline assumption. *Frontiers in Virtual Reality*, *3*, 945800. https://doi.org/10.3389/FRVIR.2022.945800/BIBTEX

Brünken, R., Plass, J. L., & Leutner, D. (2003). Direct measurement of cognitive load in multimedia learning. *Educational Psychologist*, *38*(1), 53–61. https://doi.org/10.1207/S15326985EP3801_7

Butler, C., Okamoto, G. A., & McKay, T. M. (1984). Motorized wheelchair driving by disabled children. *Archives of Physical Medicine and Rehabilitation*, *65*(2), 95–97. https://pubmed.ncbi.nlm.nih.gov/6230066/

Callet, P. Le, Möller, S., Perkis, A., Brunnström, K., Beker, S., Moor, K. De, Dooms, A., Egger, S., Garcia, M.-N., Hoßfeld, T., Jumisko-Pyykkö, S., Keimel, C., Larabi, C., Lawlor, B., Callet, P. Le, Möller, S., Pereira, F., Pereira, M., Perkis, A., … Zgank, A. (2013). *Qualinet White Paper on Definitions of Quality of Experience*. https://doi.org/10.34894/VQ1DJA

Carlsson, A., & Lundälv, J. (2019). Acute injuries resulting from accidents involving powered mobility devices (PMDs)—Development and outcomes of PMD-related accidents in Sweden. *Traffic Injury Prevention*, *20*(5), 484–491. https://doi.org/10.1080/15389588.2019.1606910

Caruso, T. J., Rama, A., Uribe-Marquez, S., & Mitchell, J. D. (2025). Pro-Con Debate: Virtual Reality Compared to Augmented Reality for Medical Simulation. *Anesthesia and Analgesia*, *140*(6), 1264–1272. https://doi.org/10.1213/ANE.0000000000007057

CENSO. (2010). *Cartilha do Censo 2010: Pessoas com Deficiência.* http://www.ibge.gov.br/estadosat/temas.php?tema=censodemog2010_defic

Chakrabarti, S., Biswas, N., Karnani, K., Padul, V., Jones, L. D., Kesari, S., & Ashili, S. (2023). Binned Data Provide Better Imputation of Missing Time Series Data from Wearables. *Sensors*, *23*(3), 1454. https://doi.org/10.3390/S23031454/S1

Chang, Z., Lichtenstein, P., D'Onofrio, B. M., Sjölander, A., & Larsson, H. (2014). Serious Transport Accidents in Adults With Attention-Deficit/Hyperactivity Disorder and the Effect of Medication: A Population-Based Study. *JAMA Psychiatry*, *71*(3), 319–325. https://doi.org/10.1001/JAMAPSYCHIATRY.2013.4174

Chen, C. W., Chatzimisios, P., Dagiuklas, T., Atzori, L., Varela, M., Skorin-Kapov, L., De Moor, K., & Reichl, P. (2015). QoE-Defining a User-Centric Concept for Service Quality. In *Multimedia Quality of Experience (QoE)*. https://doi.org/10.1002/9781118736135.ch2

Chen, M., Xu, Y., Sirois, A., Li, Y., Dempski, R., Smith, G., Oda, Y., Telliel, Y., Lewis, E. S., & Wolkowicz, K. L. (2023). WheelUp! Developing an Interactive Electric-power Wheelchair Virtual Training Environment. *IEEE Conference on Computatonal Intelligence and Games, CIG*. https://doi.org/10.1109/COG57401.2023.10333203

Chen, S., & Epps, J. (2014). Using task-induced pupil diameter and blink rate to infer cognitive load. *Human-Computer Interaction*, *29*(4), 390–413. https://doi.org/10.1080/07370024.2014.892428

Cinaz, B., La Marca, R., Arnrich, B., & Tröster, G. (2010). Towards Continuous Monitoring of Mental Workload. *International Workshop on Ubiquitous Health and Wellness*, *September*.

Client Satisfaction Questionnaire (CSQ-8) Parent and Child | Measures Library. (n.d.). Retrieved May 24, 2025, from https://elcentro.sonhs.miami.edu/research/measures-library/csq-8/index.html

Cohen, J. (2013). Statistical Power Analysis for the Behavioral Sciences. *Statistical Power Analysis for the Behavioral Sciences*. https://doi.org/10.4324/9780203771587

Cooper, R. A., Cooper, R., Tolerico, M., Guo, S., Ding, D., & Pearlman, J. (2006). Advances in Electric-Powered Wheelchairs. *Topics in Spinal Cord Injury Rehabilitation*, *11*(4), 15–29. https://doi.org/10.1310/ACUK-KFYP-ABEQ-A30C

COPM | Canadian Occupational Performance Measure. (n.d.). Retrieved May 24, 2025, from https://www.thecopm.ca/

Crichlow, L. R., Fernie, G. R., Campos, J. L., & Grant, P. R. (2012). *A FULL MOTION MANUAL WHEELCHAIR SIMULATOR FOR REHABILITATION RESEARCH*.

CSO. (2022). *Press Statement Census 2022 Results Profile 4 - Disability, Health and Carers - CSO - Central Statistics Office*. https://www.cso.ie/en/csolatestnews/pressreleases/2023pressreleases/pressstatementcensus2022resultsprofile4-disabilityhealthandcarers/

Cullen, B., O'Neill, B., & Evans, J. J. (2008). Neuropsychological predictors of powered wheelchair use: a prospective follow-up study. *Clinical Rehabilitation*, *22*(9), 836–846. https://doi.org/10.1177/0269215508091873

Curran, V. R., Xu, X., Aydin, M. Y., & Meruvia-Pastor, O. (2023). Use of Extended Reality in Medical Education: An Integrative Review. *Medical Science Educator*, *33*(1), 275–286. https://doi.org/10.1007/S40670-022-01698-4/METRICS

da Silveira, A. C., Lima de Souza, M., Ghinea, G., & Saibel Santos, C. A. (2024). Physiological Data for User Experience and Quality of Experience: A Systematic Review (2018–2022). *International*

Journal of Human-Computer Interaction. https://doi.org/10.1080/10447318.2024.2311972/ASSET/E14D2B7B-F976-4667-A621-EEA61B910699/ASSETS/GRAPHIC/HIHC_A_2311972_F0004_B.JPG

Dalhousie University. (2023). Wheelchair Skills Program Manual Version 5.4.2. In *Halifax*. Halifax. https://wheelchairskillsprogram.ca/en/skills-manual-forms/

Darji, J., Biswas, N., Padul, V., Gill, J., Kesari, S., & Ashili, S. (2024). Efficient use of binned data for imputing univariate time series data. *Frontiers in Big Data*, *7*, 1422650. https://doi.org/10.3389/FDATA.2024.1422650/BIBTEX

Dawson, D., Chan, R., & Kaiserman, E. (1994). Development of the Power-Mobility Indoor Driving Assessment for Residents of Long-Term Care Facilities: A Preliminary Report. *Http://Dx.Doi.Org/10.1177/000841749406100507*, *61*(5), 269–276. https://doi.org/10.1177/000841749406100507

Day, T. W., & John, N. W. (2019). Training powered wheelchair manoeuvres in mixed reality. *2019 11th International Conference on Virtual Worlds and Games for Serious Applications, VS-Games 2019 - Proceedings*, 1DUUMY. https://doi.org/10.1109/VS-GAMES.2019.8864515

Deitz, J., Jaffe, K. M., Wolf, L. S., Massagli, T. L., & Anson, D. (1991). Pediatric Power Wheelchairs: Evaluation of Function in the Home and School Environments. *Assistive Technology*, *3*(1), 24–31. https://doi.org/10.1080/10400435.1991.10132177

Deligkaris, P., Panagopoulou, E., Montgomery, A. J., & Masoura, E. (2014). Job burnout and cognitive functioning: A systematic review. *Work and Stress*, *28*(2), 107–123. https://doi.org/10.1080/02678373.2014.909545

Dennison, M. S., Wisti, A. Z., & D'Zmura, M. (2016). Use of physiological signals to predict cybersickness. *Displays*, *44*, 42–52. https://doi.org/10.1016/J.DISPLA.2016.07.002

Diamond, A. (2013). Executive functions. *Annual Review of Psychology*, *64*(Volume 64, 2013), 135–168. https://doi.org/10.1146/ANNUREV-PSYCH-113011-143750/CITE/REFWORKS

Diamond, S. R., & Royce, J. R. (1980). Cognitive Abilities As Expressions Of Three "Ways Of Knowing." *Multivariate Behavioral Research*, *15*(1), 31–56. https://doi.org/10.1207/S15327906MBR1501_3

Dicianno, B. E., Joseph, J., Eckstein, S., Zigler, C. K., Quinby, E. J., Schmeler, M. R., Schein, R. M., Pearlman, J., & Cooper, R. A. (2019). The future of the provision process for mobility assistive

technology: a survey of providers. *Disability and Rehabilitation: Assistive Technology*, *14*(4), 338–345. https://doi.org/10.1080/17483107.2018.1448470

Dolan, M. J., & Henderson, G. I. (2017). Control devices for electrically powered wheelchairs: prevalence, defining characteristics and user perspectives. *Disability and Rehabilitation: Assistive Technology*, *12*(6), 618–624. https://doi.org/10.1080/17483107.2016.1201154

Domingues, I., Pinheiro, J., Silveira, J., Francisco, P., Jutai, J., & Correia Martins, A. (2019). Psychosocial Impact of Powered Wheelchair, Users' Satisfaction and Their Relation to Social Participation. *Technologies 2019, Vol. 7, Page 73*, *7*(4), 73. https://doi.org/10.3390/TECHNOLOGIES7040073

Duncan, C. C., Barry, R. J., Connolly, J. F., Fischer, C., Michie, P. T., Näätänen, R., Polich, J., Reinvang, I., & Van Petten, C. (2009). Event-related potentials in clinical research: Guidelines for eliciting, recording, and quantifying mismatch negativity, P300, and N400. In *Clinical Neurophysiology* (Vol. 120, Issue 11, pp. 1883–1908). https://doi.org/10.1016/j.clinph.2009.07.045

Eager, D., Pendrill, A. M., & Reistad, N. (2016). Beyond velocity and acceleration: Jerk, snap and higher derivatives. *European Journal of Physics*. https://doi.org/10.1088/0143-0807/37/6/065008

Edwards, K., & McCluskey, A. (2010). A survey of adult power wheelchair and scooter users. *Disability and Rehabilitation: Assistive Technology*, *5*(6), 411–419. https://doi.org/10.3109/17483101003793412

Empatica. (2025a). *Decoding wearable sensor signals - what to expect from your E4 Data | Blog | Empatica*. https://www.empatica.com/blog/decoding-wearable-sensor-signals-what-to-expect-from-your-e4-data.html

Empatica. (2025b). *Empatica | Medical devices, AI and algorithms for remote patient monitoring*. https://www.empatica.com/

Eoghan Hynes, A., Murray, N., Flynn, R., & Lee, B. (2023). *A Quality of Experience Evaluation of Text and 3D Instruction Formats in Augmented Reality Applications*.

Erdfelder, E., FAul, F., Buchner, A., & Lang, A. G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, *41*(4), 1149–1160. https://doi.org/10.3758/BRM.41.4.1149

EU. (2017). *Regulation - 2017/745 - EN - Medical Device Regulation - EUR-Lex*. https://eur-lex.europa.eu/eli/reg/2017/745/oj/eng

Fager, S. K. (2018). Alternative Access for Adults Who Rely on Augmentative and Alternative Communication. *Perspectives of the ASHA Special Interest Groups*, *3*(12), 6–12. https://doi.org/10.1044/PERSP3.SIG12.6

Faure, C., Routhier, F., Lettre, J., Choukou, M. A., & Archambault, P. S. (2023). Effectiveness of the miWe Simulator Training on Powered Wheelchair-driving Skills: A Randomized Controlled Trial. *Archives of Physical Medicine and Rehabilitation*, *104*(9), 1371–1377. https://doi.org/10.1016/j.apmr.2023.04.022

Fisher, G. G., Chaffee, D. S., Tetrick, L. E., Davalos, D. B., & Potter, G. G. (2017). Cognitive functioning, aging, and work: A review and recommendations for research and practice. *Journal of Occupational Health Psychology*, *22*(3), 314–336. https://doi.org/10.1037/OCP0000086

Fishleigh, L., Taylor, R., Hale, G., & Bowers, D. S. (2024). Factors that affect powered wheelchair use for an adult population: a systematic review. *Disability and Rehabilitation: Assistive Technology*. https://doi.org/10.1080/17483107.2024.2304122

Fraudet, B., Leblong, E., Piette, P., Nicolas, B., Gouranton, V., Babel, M., Devigne, L., Pasteau, F., & Gallien, P. (2024). Evaluation of power wheelchair driving performance in simulator compared to driving in real-life situations: the SIMADAPT (simulator ADAPT) project—a pilot study. *Journal of NeuroEngineering and Rehabilitation*, *21*(1). https://doi.org/10.1186/s12984-024-01354-5

Friedman, N. P., & Miyake, A. (2017). Unity and diversity of executive functions: Individual differences as a window on cognitive structure. *Cortex*, *86*, 186–204. https://doi.org/10.1016/J.CORTEX.2016.04.023

Furumasu, J., Guerette, P., & Tefft, D. (1996). The development of a powered wheelchair mobility program for young children. *Technology and Disability*, *5*(1), 41–48. https://doi.org/10.3233/TAD-1996-5106;SUBPAGE:STRING:ABSTRACT;JOURNAL:JOURNAL:TADA;WEBSITE:WEBSITE:SAGE;REQUESTEDJOURNAL:JOURNAL:TADA;WGROUP:STRING:PUBLICATION

Furumasu, J., Guerette, P., & Tefft, D. (2004). Relevance of the Pediatric Powered Wheelchair Screening Test for children with cerebral palsy. *Developmental Medicine & Child Neurology*, *46*(7), 468–474. https://doi.org/10.1111/J.1469-8749.2004.TB00507.X

Gacem, A., Monacelli, E., Wang, T., Rabreau, O., & Al-ani, T. (2020). Assessment of wheelchair skills based on analysis of driving style. *Cognition, Technology and Work*, *22*(1), 193–207. https://doi.org/10.1007/S10111-019-00563-6/METRICS

Gambiraza, M., Kesedzic, I., Sarlija, M., Popovic, S., & Cosic, K. (2021). Classification of Cognitive Load based on Oculometric Features. *2021 44th International Convention on Information, Communication and Electronic Technology, MIPRO 2021 - Proceedings*, 377–382. https://doi.org/10.23919/MIPRO52101.2021.9597067

Garner, T. D., & Ricard, M. D. (2022). Effects of Trunk Functional Capacity on the Control of Angular Momentum During Manual Wheelchair Braking. *The Open Sports Sciences Journal*, *15*(1). https://doi.org/10.2174/1875399X-V15-E2208150

Geers, A. M., Prinsen, E. C., van der Pijl, D. J., Bergsma, A., Rietman, J. S., & Koopman, B. F. J. M. (2023). Head support in wheelchairs (scoping review): state-of-the-art and beyond. *Disability and Rehabilitation: Assistive Technology*, *18*(5), 564–587. https://doi.org/10.1080/17483107.2021.1892840

Gefen, N., Rigbi, A., Archambault, P. S., & Weiss, P. L. (2019). Comparing children's driving abilities in physical and virtual environments. *Disability and Rehabilitation: Assistive Technology*, *16*(6), 653–660. https://doi.org/10.1080/17483107.2019.1693644;WGROUP:STRING:PUBLICATION

Gefen, N., Rigbi, A., & Weiss, P. L. (2022). Reliability and validity of pediatric powered mobility outcome measures. *Disability and Rehabilitation: Assistive Technology*, *17*(8), 882–887. https://doi.org/10.1080/17483107.2020.1819449;JOURNAL:JOURNAL:IIDT20

Giannakakis, G., Grigoriadis, D., Giannakaki, K., Simantiraki, O., Roniotis, A., & Tsiknakis, M. (2022). Review on Psychological Stress Detection Using Biosignals. *IEEE Transactions on Affective Computing*, *13*(1), 440–460. https://doi.org/10.1109/TAFFC.2019.2927337

Goncalves, F., Trenoras, L., Monacelli, E., & Schmid, A. (2014). Motion adaptation on a wheelchair driving simulator. *2014 2nd Workshop on Virtual and Augmented Assistive Technology, VAAT 2014; Co-Located with the 2014 Virtual Reality Conference - Proceedings*, 17–22. https://doi.org/10.1109/VAAT.2014.6799463

Gowran, R. J., Bray, N., Goldberg, M., Rushton, P., Saab, M. B. A., Constantine, D., Ghosh, R., & Pearlman, J. (2021). Understanding the Global Challenges to Accessing Appropriate Wheelchairs: Position Paper. *International Journal of Environmental Research and Public Health 2021, Vol. 18, Page 3338*, *18*(7), 3338. https://doi.org/10.3390/IJERPH18073338

Grant, P. R., & Haycock, B. (2008). Effect of jerk and acceleration on the perception of motion strength. *Journal of Aircraft*. https://doi.org/10.2514/1.33757

Gullett, N., Zajkowska, Z., Walsh, A., Harper, R., & Mondelli, V. (2023). Heart rate variability (HRV) as a way to understand associations between the autonomic nervous system (ANS) and affective states: A critical review of the literature. *International Journal of Psychophysiology*, *192*, 35–42. https://doi.org/10.1016/J.IJPSYCHO.2023.08.001

Guy-Evans, O., & McLeod, S. (2025). *Autonomic Nervous System (ANS): What It Is and How It Works*. https://www.simplypsychology.org/autonomic-nervous-system.html

H. Montenegro-Couto, E., A. Hernandez-Ossa, K., L. C. Bissoli, A., Sime, M., & F. Bastos-Filho, T. (2018). *TOWARDS AN ASSISTIVE INTERFACE TO COMMAND ROBOTIC WHEELCHAIRS AND INTERACT WITH ENVIRONMENT THROUGH EYE GAZE*. https://doi.org/10.29327/COBECSEB.78867

Hafid, N., & Inoue, T. (2006). Assessment of driving skills using virtual reality: Comparative survey on experts and unskilled users of electric wheelchairs. *Technology and Disability*, *18*(4), 217–226. https://doi.org/10.3233/TAD-2006-18409

Hafid Niniss, & A Nadif. (2000). The 3rd international conference on disability, virtual reality and associated technologies. *Proceedings of the 3rd International Conference on Disabilities, Virtual Reality and Associated Technologies*, 9–14. https://www.researchgate.net/publication/228720597_Simulation_of_the_behaviour_of_a_powered_wheelchair_using_virtual_reality

Hart, S. G. (2006). NASA-task load index (NASA-TLX); 20 years later. *Proceedings of the Human Factors and Ergonomics Society*.

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. *Advances in Psychology*, *52*(C), 139–183. https://doi.org/10.1016/S0166-4115(08)62386-9

Hasdai, A., Jessel, A. S., & Weiss, P. L. (1998). Use of a Computer Simulator for Training Children With Disabilities in the Operation of a Powered Wheelchair. *The American Journal of Occupational Therapy*, *52*(3), 215–220. https://doi.org/10.5014/AJOT.52.3.215

Headleand, C. J., Day, T., Pop, S. R., Ritsos, P. D., & John, N. W. (2015). Challenges and technologies for low cost wheelchair simulation. *Eurographics Workshop on Visual Computing for Biology and Medicine, VCBM 2015*, 207–208. https://doi.org/10.2312/vcbm.20151225

Headleand, C. J., Day, T., Pop, S. R., Ritsos, P. D., & John, N. W. (2016). A Cost-Effective Virtual Environment for Simulating and Training Powered Wheelchairs Manoeuvres. *Studies in Health Technology and Informatics*, *220*, 134–141. https://doi.org/10.3233/978-1-61499-625-5-134

Hebbar, P. A., Bhattacharya, K., Prabhakar, G., Pashilkar, A. A., & Biswas, P. (2021). Correlation Between Physiological and Performance-Based Metrics to Estimate Pilots' Cognitive Workload. *Frontiers in Psychology*, *12*, 555446. https://doi.org/10.3389/FPSYG.2021.555446/BIBTEX

Hernandez-Ossa, K. A., Longo, B., Montenegro-Couto, E., Alejandra Romero-Laiseca, M., Frizera-Neto, A., & Bastos-Filho, T. (2017). Development and pilot test of a virtual reality system for electric powered wheelchair simulation. *2017 IEEE International Conference on Systems, Man, and Cybernetics, SMC 2017*, *2017-January*, 2355–2360. https://doi.org/10.1109/SMC.2017.8122974

Hernandez-Ossa, K. A., Montenegro-Couto, E. H., Longo, B., Bissoli, A., Sime, M. M., Lessa, H. M., Enriquez, I. R., Frizera-Neto, A., & Bastos-Filho, T. (2020). Simulation System of Electric-Powered Wheelchairs for Training Purposes. *Sensors 2020, Vol. 20, Page 3565*, *20*(12), 3565. https://doi.org/10.3390/S20123565

Hertzog, M. A. (2008). Considerations in determining sample size for pilot studies. *Research in Nursing & Health*, *31*(2), 180–191. https://doi.org/10.1002/NUR.20247

Hoenig, H., Pieper, C., Zolkewitz, M., Schenkman, M., & Branch, L. G. (2002). Wheelchair users are not necessarily wheelchair bound. *Journal of the American Geriatrics Society*, *50*(4), 645–654. https://doi.org/10.1046/j.1532-5415.2002.50158.x

Hoffmann, A., & Menozzi, M. (1999). Applying the Ishihara test to a PC-based screening system. *Displays, 20*(1), 39–47. https://doi.org/10.1016/S0141-9382(98)00053-5

Holger Hoffmann. (2015). *Violin Plot - File Exchange - MATLAB Central*. Https://Uk.Mathworks.Com/Matlabcentral/Fileexchange/45134-Violin-Plot. https://uk.mathworks.com/matlabcentral/fileexchange/45134-violin-plot

Horvers, A., Tombeng, N., Bosse, T., Lazonder, A. W., & Molenaar, I. (2021). Detecting Emotions through Electrodermal Activity in Learning Contexts: A Systematic Review. *Sensors 2021, Vol. 21, Page 7869*, *21*(23), 7869. https://doi.org/10.3390/S21237869

Hughes, A. M., Hancock, G. M., Marlow, S. L., Stowers, K., & Salas, E. (2019). Cardiac Measures of Cognitive Workload: A Meta-Analysis. *Https://Doi.Org/10.1177/0018720819830553*, *61*(3), 393–414. https://doi.org/10.1177/0018720819830553

Hynes, E., Flynn, R., Lee, B., & Murray, N. (2023). A QoE evaluation of augmented reality for the informational phase of procedure assistance. *Quality and User Experience*, *8*(1). https://doi.org/10.1007/s41233-023-00054-7

Igroup Project Consortium. (2015). *Igroup Presence Questionnaire*. Igroup.Org.

IJsselsteijn, W. A., Kort, Y. A. W. de, & Poels, K. (2013). *The Game Experience Questionnaire*. Technische Universiteit Eindhoven. https://research.tue.nl/en/publications/the-game-experience-questionnaire

Inman, D. P., Loge, K., Cram, A., & Peterson, M. (2011). Learning to Drive a Wheelchair in Virtual Reality. *Journal of Special Education Technology*, *26*(3), 21–34. https://doi.org/10.1177/016264341102600303;SUBPAGE:STRING:ABSTRACT;REQUESTEDJOURNAL:JOURNAL:JSTA;WEBSITE:WEBSITE:SAGE;WGROUP:STRING:PUBLICATION

Iskander, J., Hossny, M., & Nahavandi, S. (2018). A Review on Ocular Biomechanic Models for Assessing Visual Fatigue in Virtual Reality. *IEEE Access*, *6*, 19345–19361. https://doi.org/10.1109/ACCESS.2018.2815663

Islam, R., Lee, Y., Jaloli, M., Muhammad, I., Zhu, D., Rad, P., Huang, Y., & Quarles, J. (2020). Automatic Detection and Prediction of Cybersickness Severity using Deep Neural Networks from user's Physiological Signals. *Proceedings - 2020 IEEE International Symposium on Mixed and Augmented Reality, ISMAR 2020*, 400–411. https://doi.org/10.1109/ISMAR50242.2020.00066

ISO. (2007). *ISO 8589:2007 - Sensory analysis -- General guidance for the design of test rooms*. https://www.iso.org/standard/36385.html

ITU. (2017). *P.10 : New definitions for inclusion in Recommendation ITU-T P.10/G.100*. https://www.itu.int/rec/T-REC-P.10-202405-I!Amd2/en

ITU-T. (2025). *SG12 - Performance, quality of service and quality of experience*. https://www.itu.int/en/ITU-T/studygroups/2025-2028/12/Pages/default.aspx

ITU-T P. 809: Standardization Activities Targeting Gaming Quality of Experience – ACM SIGMM Records. (2018). https://records.sigmm.org/2021/03/24/itu-t-standardization-activities-targeting-gaming-quality-of-experience/

ITU-T P.913. (2018). *ITU-P.913 : Methods for the subjective assessment of video quality, audio quality and audiovisual quality of Internet video and distribution quality television in any environment*. https://www.itu.int/rec/T-REC-P.913/en

Jenkins, G. R., Vogtle, L. K., & Yuen, H. K. (2015). Factors Associated With the Use of Standardized Power Mobility Skills Assessments Among Assistive Technology Practitioners. *Assistive Technology : The Official Journal of RESNA*, *27*(4), 219–225. https://doi.org/10.1080/10400435.2015.1030515

Johanson, G. A., & Brooks, G. P. (2009). Initial Scale Development: Sample Size for Pilot Studies. *Http://Dx.Doi.Org/10.1177/0013164409355692*, *70*(3), 394–400. https://doi.org/10.1177/0013164409355692

John, N. W., Pop, S. R., Day, T. W., Ritsos, P. D., & Headleand, C. J. (2018). The Implementation and Validation of a Virtual Environment for Training Powered Wheelchair Manoeuvres. *IEEE Transactions on Visualization and Computer Graphics*, *24*(5), 1867–1878. https://doi.org/10.1109/TVCG.2017.2700273

Jones, M. A., McEwen, I. R., & Hansen, L. (2003). Use of Power Mobility for a Young Child With Spinal Muscular Atrophy. *Physical Therapy*, *83*(3), 253–262. https://doi.org/10.1093/PTJ/83.3.253

Kamaraj, D. C. (2020). *Quantifying Electric Powered Wheelchair Driving Ability* [University of Pittsburgh]. https://doi.org/http://d-scholarship.pitt.edu/id/eprint/39223

Kamaraj, D. C., Dicianno, B. E., Mahajan, H. P., Buhari, A. M., & Cooper, R. A. (2016a). Interrater Reliability of the Power Mobility Road Test in the Virtual Reality-Based Simulator-2. *Archives of Physical Medicine and Rehabilitation*, *97*(7), 1078–1084. https://doi.org/10.1016/j.apmr.2016.02.005

Kamaraj, D. C., Dicianno, B. E., Mahajan, H. P., Buhari, A. M., & Cooper, R. A. (2016b). Stability and Workload of the Virtual Reality-Based Simulator-2. *Archives of Physical Medicine and Rehabilitation*, *97*(7), 1085–1092. https://doi.org/10.1016/j.apmr.2016.01.032

Kamaraj, D. C., Mahajan, H. P., Dicianno, B. E., Terhorst, L., & Cooper, R. (2016). Discriminative Ability of The Quantitative Electric Powered Wheelchair Driving Metrics In VRSIM-2. *Archives of Physical Medicine and Rehabilitation*, *97*(10), e47. https://doi.org/10.1016/j.apmr.2016.08.139

Kaufeld, M., Bourdeinik, J., Prinz, L. M., Mundt, M., & Hecht, H. (2022). Emotions are associated with the genesis of visually induced motion sickness in virtual reality. *Experimental Brain Research*, *240*(10), 2757–2771. https://doi.org/10.1007/S00221-022-06454-Z/METRICS

Kennedy, R. S., Lane, N. E., Berbaum, K. S., & Lilienthal, M. G. (1993). Simulator Sickness Questionnaire: An Enhanced Method for Quantifying Simulator Sickness. *The International Journal of Aviation Psychology*, *3*(3), 203–220. https://doi.org/10.1207/s15327108ijap0303_3

Kenyon, L. K., Farris, J., Brockway, K., Hannum, N., & Proctor, K. (2015). Promoting Self-exploration and Function Through an Individualized Power Mobility Training Program. *Pediatric Physical Therapy*, *27*(2), 200–206. https://doi.org/10.1097/PEP.0000000000000129

Kenyon, L. K., Farris, J. P., Cain, B., King, E., & VandenBerg, A. (2018). Development and content validation of the power mobility training tool. *Disability and Rehabilitation: Assistive Technology*, *13*(1), 10–24. https://doi.org/10.1080/17483107.2016.1278468

Kenyon, L. K., Farris, J. P., Veety, L., Kleikamp, B., Harrington, K., Jenkinson, J., Montgomery, A., Otieno, S., Russell, I. M., & Zondervan, D. K. (2024). The IndieTrainer system: a small-scale trial exploring a new approach to support powered mobility skill acquisition in children. *Disability and Rehabilitation: Assistive Technology*. https://doi.org/10.1080/17483107.2024.2325563;PAGE:STRING:ARTICLE/CHAPTER

Kenyon, L. K., Jones, M., Breaux, B., Tsotsoros, J., Gardner, T., & Livingstone, R. (2020). American and Canadian therapists' perspectives of age and cognitive skills for paediatric power mobility: a qualitative study. *Disability and Rehabilitation. Assistive Technology*, *15*(6), 692–700. https://doi.org/10.1080/17483107.2019.1606858

Kenyon, L. K., Jones, M., Livingstone, R., Breaux, B., Tsotsoros, J., & Williams, K. M. (2018). Power mobility for children: a survey study of American and Canadian therapists' perspectives and practices. *Developmental Medicine and Child Neurology*, *60*(10), 1018–1025. https://doi.org/10.1111/dmcn.13960

Kirby, R. L., Miller, W. C., Routhier, F., Demers, L., Mihailidis, A., Polgar, J. M., Rushton, P. W., Titus, L., Smith, C., McAllister, M., Theriault, C., Thompson, K., & Sawatzky, B. (2015). Effectiveness of a Wheelchair Skills Training Program for Powered Wheelchair Users: A Randomized Controlled Trial. *Archives of Physical Medicine and Rehabilitation*, *96*(11), 2017-2026.e3. https://doi.org/10.1016/j.apmr.2015.07.009

Kirby, R. L., Smith, C., Parker, K., Han, L., Theriault, C. J., & Doucette, S. P. (2020). Practices and views of occupational therapists in Nova Scotia regarding wheelchair-skills training for clients and their caregivers: an online survey. *Disability and Rehabilitation: Assistive Technology*, *15*(7), 773–780. https://doi.org/10.1080/17483107.2020.1749890

Kirby, R. L., Swuste, J., Dupuis, D. J., MacLeod, D. A., & Monroe, R. (2002). The Wheelchair Skills Test: A pilot study of a new outcome measure. *Archives of Physical Medicine and Rehabilitation*, *83*(1), 10–18. https://doi.org/10.1053/APMR.2002.26823

Kirby, R. Lee. (2017). *Wheelchair skills assessment and training*. CRC Press, Taylor & Francis.

Kirk-Wade, E., Wong, H., & Stiebahl, S. (2024). *UK disability statistics: Prevalence and life experiences*. https://commonslibrary.parliament.uk/research-briefings/cbp-9602/

Kleiman, E. M., Bentley, K. H., Maimone, J. S., Lee, H. I. S., Kilbury, E. N., Fortgang, R. G., Zuromski, K. L., Huffman, J. C., & Nock, M. K. (2021). Can passive measurement of physiological distress help better predict suicidal thinking? *Translational Psychiatry*, *11*(1), 1–6. https://doi.org/10.1038/S41398-021-01730-Y;SUBJMETA=2811,477,631,648,706;KWRD=HUMAN+BEHAVIOUR,SCIENTIFIC+COMMUNITY

Kojić T., Voigt-Antons J.-N., & Möller, S. (2021). *Proposal for new Work Item P.IntVR: Subjective Test Method for Interactive Virtual Reality Applications.* Contribution Presented to the ITU-T SG12 as COM 12 – C 548 – E, Geneva, Switzerland. https://www.itu.int/md/T17-SG12-C-0548/en

Kothe, C., Yahya Shirazi, S., Stenner, T., Medine, D., Boulay, C., Grivich, M. I., Artoni, F., Mullen, T., Delorme, A., Makeig, S., & Author, C. (2025). The lab streaming layer for synchronized multimodal recording. *Imaging Neuroscience*, *3*, 2025. https://doi.org/10.1162/IMAG.A.136

Kourtesis, P., Collina, S., Doumas, L. A. A., & MacPherson, S. E. (2019). Validation of the Virtual Reality Neuroscience Questionnaire: Maximum Duration of Immersive Virtual Reality Sessions Without the Presence of Pertinent Adverse Symptomatology. *Frontiers in Human Neuroscience*, *13*, 489597. https://doi.org/10.3389/FNHUM.2019.00417/BIBTEX

Krieglstein, F., Beege, M., Rey, G. D., Ginns, P., Krell, M., & Schneider, S. (2022). A Systematic Meta-analysis of the Reliability and Validity of Subjective Cognitive Load Questionnaires in Experimental Multimedia Learning Research. *Educational Psychology Review*, *34*(4), 2485–2541. https://doi.org/10.1007/s10648-022-09683-4

Kyriaki, K., Koukopoulos, D., & Fidas, C. A. (2024). A Comprehensive Survey of EEG Preprocessing Methods for Cognitive Load Assessment. *IEEE Access*, *12*, 23466–23489. https://doi.org/10.1109/ACCESS.2024.3360328

L. Devigne, M. Babel, F. Nouviale, V. K. Narayanan, F. Pasteau, & P. Gallien. (2017). Design of an immersive simulator for assisted power wheelchair driving. *2017 International Conference on Rehabilitation Robotics (ICORR)*. https://doi.org/10.1109/ICORR.2017.8009379

Lam, J. F., Gosselin, L., & Rushton, P. W. (2018). Use of Virtual Technology as an Intervention for Wheelchair Skills Training: A Systematic Review. In *Archives of Physical Medicine and Rehabilitation* (Vol. 99, Issue 11, pp. 2313–2341). W.B. Saunders. https://doi.org/10.1016/j.apmr.2018.02.007

Leblong, E., Fraudet, B., Devigne, L., Babel, M., Pasteau, F., Nicolas, B., & Gallien, P. (2021). SWADAPT1: assessment of an electric wheelchair-driving robotic module in standardized circuits: a prospective, controlled repeated measure design pilot study. *Journal of NeuroEngineering and Rehabilitation*, *18*(1), 1–12. https://doi.org/10.1186/S12984-021-00923-2/TABLES/3

Lee, K., Sim, Y. S., Cho, H.-S., Eo, M., Yoon, S., Yoon, S., & Lim, W. (2024). *Binning as a Pretext Task: Improving Self-Supervised Learning in Tabular Domains*. https://arxiv.org/pdf/2405.07414

Lenhard, W., & Lenhard, A. (2022). *Computation of different effect sizes*. https://doi.org/10.13140/RG.2.2.17823.92329

Letts, L., Dawson, D., Bretholz, I., Kaiserman-Goldenstein, E., Gleason, J., McLellan, E., Norton, L., & Roth, C. (2007). Reliability and Validity of the Power-Mobility Community Driving Assessment. *Assistive Technology*, *19*(3), 154–163. https://doi.org/10.1080/10400435.2007.10131872

Li, C. X., Yang, Y., Chen, X. Y., Xu, J. D., Song, S., Fan, D. C., & Chen, F. (2014). Mental workload of young drivers during curve negotiation. *2014 International Conference on Connected Vehicles and Expo, ICCVE 2014 - Proceedings*, 291–292. https://doi.org/10.1109/ICCVE.2014.7297558

Li, W., Spachos, P., Chignell, M., Leon-Garcia, A., Zucherman, L., & Jiang, J. (2016). Understanding the relationships between performance metrics and QoE for Over-The-Top video. *2016 IEEE International Conference on Communications, ICC 2016*. https://doi.org/10.1109/ICC.2016.7511100

Liao, D., Shu, L., Liang, G., Li, Y., Zhang, Y., Zhang, W., & Xu, X. (2020). Design and Evaluation of Affective Virtual Reality System Based on Multimodal Physiological Signals and Self-Assessment Manikin. *IEEE Journal of Electromagnetics, RF and Microwaves in Medicine and Biology*, *4*(3), 216–224. https://doi.org/10.1109/JERM.2019.2948767

Lima, R., Osório, D., & Gamboa, H. (2020). Heart Rate Variability and Electrodermal Activity Biosignal Processing: Predicting the Autonomous Nervous System Response in Mental Stress. *Communications in Computer and Information Science*, *1211 CCIS*, 328–351. https://doi.org/10.1007/978-3-030-46970-2_16

Liu, L., Wang, J., & Chen, W. (2014). A Virtual Simulation and Driver Evaluation Platform for Smart Wheelchairs. *Communications in Computer and Information Science*, *461*, 307–318. https://doi.org/10.1007/978-3-662-45283-7_32

Livingstone, R. (2010). A critical review of powered mobility assessment and training for children. *Disability and Rehabilitation: Assistive Technology*, *5*(6), 392–400. https://doi.org/10.3109/17483107.2010.496097

Magalhães, M., Coelho, A., Melo, M., & Bessa, M. (2024). Measuring users' emotional responses in multisensory virtual reality: a systematic literature review. *Multimedia Tools and Applications*, *83*(14), 43377–43417. https://doi.org/10.1007/S11042-023-16918-1/

Mahajan, H. P. (2012). *DEVELOPMENT AND VALIDATION OF SIMULATORS FOR POWER WHEELCHAIR DRIVING EVALUATIONS* [University of Pittsburgh]. http://d-scholarship.pitt.edu/12080/1/Harshal_Mahajan_Dissertation.pdf

Mahajan, H. P., Dicianno, B. E., Cooper, R. A., & Ding, D. (2013). Assessment of wheelchair driving performance in a virtual reality-based simulator. *Journal of Spinal Cord Medicine*, *36*(4), 322–332. https://doi.org/10.1179/2045772313Y.0000000130

Mancinetti, M., Guttormsen, S., & Berendonk, C. (2019). Cognitive load in internal medicine: What every clinical teacher should know about cognitive load theory. In *European Journal of Internal Medicine* (Vol. 60, pp. 4–8). Elsevier B.V. https://doi.org/10.1016/j.ejim.2018.08.013

Mannion, R., & Exworthy, M. (2017). (Re) Making the Procrustean Bed? Standardization and Customization as Competing Logics in Healthcare. *International Journal of Health Policy and Management*, *6*(6), 301–304. https://doi.org/10.15171/IJHPM.2017.35

Mao, R. Q., Lan, L., Kay, J., Lohre, R., Ayeni, O. R., Goel, D. P., & SA, D. de. (2021). Immersive Virtual Reality for Surgical Training: A Systematic Review. *Journal of Surgical Research*, *268*, 40–58. https://doi.org/10.1016/J.JSS.2021.06.045

Martins, F. (2022). *Application of neural network to assess wheelchair driving abilities by power mobility road test* [Universidade Federal de Uberlândia]. https://doi.org/10.14393/ufu.te.2022.679

Martins, F. R. (2017). *Simulador para treinamento de cadeirantes em ambiente virtual acionado por comandos musculares e/ou visuais* [Universidade Federal de Uberlândia]. https://doi.org/10.14393/UFU.DI.2017.504

Martins, F. R., Naves, E. L. M., Morère, Y., & de Sá, A. A. R. (2022). Preliminary assessment of a multimodal electric-powered wheelchair simulator for training of activities of daily living. *Journal on Multimodal User Interfaces*, *16*(2), 193–205. https://doi.org/10.1007/S12193-021-00385-9/METRICS

Martins, R., & Carvalho, J. M. (2015). Eye blinking as an indicator of fatigue and mental load—a systematic review. *Occupational Safety and Hygiene III - Selected Extended and Revised Contributions from the International Symposium on Safety and Hygiene*. https://doi.org/10.1201/b18042-48

Massengale, S., Folden, D., McConnell, P., Stratton, L., Whitehead, V., & Whitehead, V. (2005). Effect of visual perception, visual function, cognition, and personality on power wheelchair use in adults. *Assistive Technology*, *17*(2), 108–121. https://doi.org/10.1080/10400435.2005.10132101

Mathis, K., & Joan Gowran, R. (2021). *A cross-sectional survey investigating wheelchair skills training in Ireland*. https://doi.org/10.1080/17483107.2021.2001058

Mattson, M. P. (2014). Superior pattern processing is the essence of the evolved human brain. *Frontiers in Neuroscience*, *8*(8 AUG), 98857. https://doi.org/10.3389/FNINS.2014.00265/XML/NLM

Maywald, A., & Stanley, M. (2015). Prescribing mobility scooters in Australia: Occupational therapists' narratives. *Australian Occupational Therapy Journal*, *62*(2), 86–92. https://doi.org/10.1111/1440-1630.12131

Mikolajewska, E. (2013). Wheelchair-related falls-a review. In *International Journal on Disability and Human Development* (Vol. 12, Issue 3, pp. 259–266). https://doi.org/10.1515/ijdhd-2012-0115

Möller, S., & Raake, A. (2013). Quality of Experience - Advanced Concepts, Applications and Methods. In *T-Labs Series in Telecommunication Services*.

Möller, S., Wältermann, M., & Garcia, M. N. (2014). Features of Quality of Experience. *T-Labs Series in Telecommunication Services*, 73–84. https://doi.org/10.1007/978-3-319-02681-7_5

Moreno, R., & Park, B. (2010). Cognitive load theory: Historical development and relation to other theories. In *Cognitive Load Theory* (pp. 7–28). Cambridge University Press. https://doi.org/10.1017/CBO9780511844744.003

Morère, Y., Bourhis, G., Cosnuau, K., Guilmois, G., Rumilly, E., & Blangy, E. (2018). ViEW: A wheelchair simulator for driving analysis. *Assistive Technology*, *32*(3), 125–135. https://doi.org/10.1080/10400435.2018.1503204

Morère, Y., Hadj Abdelkader, M. A., Meliani, S. M., & Bourhis, G. (2011). Powered wheelchair driving analysis on a simulator. *Assistive Technology Research Series*, *29*, 679–685. https://doi.org/10.3233/978-1-60750-814-4-679

Mortenson, W. Ben, Clarke, L. H., & Best, K. (2013). Prescribers' Experiences With Powered Mobility Prescription Among Older Adults. *The American Journal of Occupational Therapy*, *67*(1), 100–107. https://doi.org/10.5014/AJOT.2013.006122

Mortenson, W. Ben, Hurd Clarke, L., Goldsmith, C. H., Jang, S., & Kirby, R. L. (2018). Measurement properties of the Wheelchair Skills Test for scooters among experienced users. *Disability and Rehabilitation: Assistive Technology*, *13*(1), 60–65. https://doi.org/10.1080/17483107.2017.1280546

Mountain, A. D., Kirby, R. L., Eskes, G. A., Smith, C., Duncan, H., MacLeod, D. A., & Thompson, K. (2010). Ability of People With Stroke to Learn Powered Wheelchair Skills: A Pilot Study. *Archives of Physical Medicine and Rehabilitation*, *91*(4), 596–601. https://doi.org/10.1016/j.apmr.2009.12.011

Moya, V., Slawiñski, E., Mut, V., & Couto, E. H. M. (2017). Workload detection based on EEG device for teleoperation of a mobile robot. *17th Work. Inf. Process. Control. RPIC 2017*. https://doi.org/10.23919/RPIC.2017.8211636

Murphy, D., & Higgins, C. (2019). *Secondary Inputs for Measuring User Engagement in Immersive VR Education Environments*. https://arxiv.org/pdf/1910.01586

N. Hollender, C. Hofmann, M. Deneke, B. S. (2010). Integrating cognitive load theory and concepts of human-computer interaction. *Comput. Human Behav*, *26*(6), 1278–1288. https://doi.org/10.1016/j.chb.2010.05.031

Naismith, L. M., Cheung, J. J. H., Ringsted, C., & Cavalcanti, R. B. (2015). Limitations of subjective cognitive load measures in simulation-based procedural training. *Medical Education*, *49*(8), 805–814. https://doi.org/10.1111/medu.12732

Nasreddine, Z. S., Phillips, N. A., Bédirian, V., Charbonneau, S., Whitehead, V., Collin, I., Cummings, J. L., & Chertkow, H. (2005). The Montreal Cognitive Assessment, MoCA: A Brief Screening Tool For Mild Cognitive Impairment. *Journal of the American Geriatrics Society*, *53*(4), 695–699. https://doi.org/10.1111/J.1532-5415.2005.53221.X

Navarro, J., Hernout, E., Osiurak, F., & Reynaud, E. (2020). On the nature of eye-hand coordination in natural steering behavior. *PLoS ONE*, *15*(11 November). https://doi.org/10.1371/journal.pone.0242818

NDA. (2006). *Estimated Number of Users of Changing Places Toilets - Centre for Excellence in Universal Design*. https://universaldesign.ie/built-environment/universal-design-guidelines-for-changing-

places-toilets/universal-design-guidelines-for-changing-places-toilets-browsable/introduction-and-context/e

NHS. (2021). *Reference wheelchair standard and transport design - GOV.UK*. https://www.gov.uk/government/publications/reference-wheelchair-standard-and-transport-design

Nichols, S., & Patel, H. (2002). Health and safety implications of virtual reality: a review of empirical evidence. *Applied Ergonomics*, *33*(3), 251–271. https://doi.org/10.1016/S0003-6870(02)00020-0

Nilsson, L., & Durkin, J. (2014). Assessment of learning powered mobility use—Applying grounded theory to occupational performance. *Journal of Rehabilitation Research and Development*, *51*(6), 963–974. https://doi.org/10.1682/JRRD.2013.11.0237,

Nilsson, L., & Kenyon, L. (2022). Assessment and Intervention for Tool-Use in Learning Powered Mobility Intervention: A Focus on Tyro Learners. *Disabilities 2022, Vol. 2, Pages 304-316*, *2*(2), 304–316. https://doi.org/10.3390/DISABILITIES2020022

Nunnerley, J., Gupta, S., Snell, D., & King, M. (2017). Training wheelchair navigation in immersive virtual environments for patients with spinal cord injury–end-user input to design an effective system. *Disability and Rehabilitation: Assistive Technology*, *12*(4), 417–423. https://doi.org/10.1080/17483107.2016.1176259

Ouwehand, K., Kroef, A. van der, Wong, J., & Paas, F. (2021). Measuring Cognitive Load: Are There More Valid Alternatives to Likert Rating Scales? *Frontiers in Education*, *6*, 702616. https://doi.org/10.3389/FEDUC.2021.702616/BIBTEX

Paas, F., Camp, G., & Rikers, R. (2001). Instructional compensation for age-related cognitive declines: Effects of goal specificity in maze learning. *Journal of Educational Psychology*, *93*(1), 181–186. https://doi.org/10.1037/0022-0663.93.1.181

Paas, F. G. W. C., & Van Merriënboer, J. J. G. (1994a). Instructional control of cognitive load in the training of complex cognitive tasks. *Educational Psychology Review*, *6*(4), 351–371. https://doi.org/10.1007/BF02213420/METRICS

Paas, F. G. W. C., & Van Merriënboer, J. J. G. (1994b). Variability of Worked Examples and Transfer of Geometrical Problem-Solving Skills: A Cognitive-Load Approach. *Journal of Educational Psychology*, *86*(1), 122–133. https://doi.org/10.1037/0022-0663.86.1.122

Paas, F., Tuovinen, J. E., Tabbers, H., & Van Gerven, P. W. M. (2003). Cognitive load measurement as a means to advance cognitive load theory. *Educational Psychologist*, *38*(1), 63–71. https://doi.org/10.1207/S15326985EP3801_8

Palmisano, S., Stephenson, L., Davies, R. G., Kim, J., & Allison, R. S. (2024). Testing the 'differences in virtual and physical head pose' and 'subjective vertical conflict' accounts of cybersickness. *Virtual Reality*, *28*(1), 1–28. https://doi.org/10.1007/S10055-023-00909-6/METRICS

Panadero, C. F., De La Cruz, V., Carlos, B., Kloos, D., & Morán Núñez, D. (2014a). PhyMEL-WS: Physically Experiencing the Virtual World. Insights into Mixed Reality and Flow State on Board a Wheelchair Simulator. *JUCS - Journal of Universal Computer Science 20(12): 1629-1648*, *20*(12), 1629–1648. https://doi.org/10.3217/JUCS-020-12-1629

Panadero, C. F., De La Cruz, V., Carlos, B., Kloos, D., & Morán Núñez, D. (2014b). *PhyMEL-WS: Physically Experiencing the Virtual World. Insights into Mixed Reality and Flow State on Board a Wheelchair Simulator*.

Patel, K., Milosevic, M., Nakazawa, K., Popovic, M. R., & Masani, K. (2017). Wheelchair neuroprosthesis for improving dynamic trunk stability. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, *25*(12), 2472–2479. https://doi.org/10.1109/TNSRE.2017.2727072

Pellichero, A., Best, K., Leblond, J., Coignard, P., Sorita, É., & Routhier, F. (2021). Relationships between cognitive functioning and power wheelchair performance, confidence and life-space mobility among experienced power wheelchair users: An exploratory study. *Journal of Rehabilitation Medicine*, *53*(9), jrm00226–jrm00226. https://doi.org/10.2340/16501977-2869

Pellichero, A., Kenyon, L. K., Best, K. L., Lamontagne, M. E., Lavoie, M. D., Sorita, É., & Routhier, F. (2021). Relationships between cognitive functioning and powered mobility device use: A scoping review. In *International Journal of Environmental Research and Public Health* (Vol. 18, Issue 23). MDPI. https://doi.org/10.3390/ijerph182312467

Perkis, A., Timmerer, C., Baraković, S., Husić, J. B., Bech, S., Bosse, S., Botev, J., Brunnström, K., Cruz, L., De Moor, K., Saibanti, A. de P., Durnez, W., Egger-Lampl, S., Engelke, U., Falk, T. H., Gutiérrez, J., Hameed, A., Hines, A., Kojic, T., … Zadtootaghaj, S. (2020). *QUALINET White Paper on Definitions of Immersive Media Experience (IMEx)*. https://arxiv.org/pdf/2007.07032

Phadke, V., Harteveld, C., Jona, K., & Moghaddam, M. (2024). Exploring the Distinctive and Synergistic Affordances of AR and VR for Complex Psychomotor Training. *Proceedings - 2024 IEEE*

International Symposium on Mixed and Augmented Reality Adjunct, ISMAR-Adjunct 2024, 537–540. https://doi.org/10.1109/ISMAR-ADJUNCT64951.2024.00155

Pithon, T., Weiss, T., Richir, S., & Klinger, E. (2009). Wheelchair simulators: A review. *Technology and Disability*, *21*(1–2), 1–10. https://doi.org/10.3233/TAD-2009-0268

Poels, K., Kort, Y. A. W. de, & IJsselsteijn, W. A. (2007). *D3.3 : Game Experience Questionnaire: development of a self-report measure to assess the psychological impact of digital games*. Technische Universiteit Eindhoven. https://research.tue.nl/en/publications/d33-game-experience-questionnaire-development-of-a-self-report-me

Quiñones-Uriostegui, I., Alessi-Montero, A., Bueyes-Roiz, V., Nuñez-Carrera, L., Moreno-Hernández, A., Quinzaños-Fresnedo, J., & Rodríguez-Reyes, G. (2023). Wheelchair users' satisfaction after provision using the WHO 8-step guidelines: A pilot study. *The Journal of Spinal Cord Medicine*, *47*(5), 640. https://doi.org/10.1080/10790268.2023.2171627

Randolph, C., Tierney, M. C., Mohr, E., & Chase, T. N. (1998). The Repeatable Battery for the Assessment of Neuropsychological Status (RBANS): Preliminary Clinical Validity. *Journal of Clinical and Experimental Neuropsychology*, *20*(3), 310–319. https://doi.org/10.1076/JCEN.20.3.310.823

Reedy, G. B. (2015). Using Cognitive Load Theory to Inform Simulation Design and Practice. *Clinical Simulation in Nursing*, *11*(8), 355–360. https://doi.org/10.1016/j.ecns.2015.05.004

Regal, G., Schatz, R., Schrammel, J., & Suette, S. (2018). VRate: A Unity3D Asset for integrating Subjective Assessment Questionnaires in Virtual Environments. *2018 10th International Conference on Quality of Multimedia Experience, QoMEX 2018*. https://doi.org/10.1109/QOMEX.2018.8463296

Regulation - 2017/746 - EN - Medical Device Regulation - EUR-Lex. (n.d.). Retrieved March 27, 2025, from https://eur-lex.europa.eu/eli/reg/2017/746/oj

Reiter, U., Brunnström, K., De Moor, K., Larabi, M.-C., Pereira, M., Pinheiro, A., You, J., & Zgank, A. (2014). Factors Influencing Quality of Experience. In *Quality of experience : advanced concepts, applications and methods* (pp. 55–72). https://doi.org/10.1007/978-3-319-02681-7_4

Renard, Y., Lotte, F., Gibert, G., Congedo, M., Maby, E., Delannoy, V., Bertrand, O., & Lécuyer, A. (2010). OpenViBE: An Open-Source Software Platform to Design, Test, and Use Brain–Computer Interfaces in Real and Virtual Environments. *Presence: Teleoperators and Virtual Environments*, *19*(1), 35–53. https://doi.org/10.1162/pres.19.1.35

Roberts, J., Young, H., Andrew, K., McAlpine, A., & Hogg, J. (2012). The needs of carers who push wheelchairs. *Journal of Integrated Care*, *20*(1), 23–34. https://doi.org/10.1108/14769011211202265/FULL/XML

Rodrigues, T. B., Cathain, C. O., Connor, N. E. O., & Murray, N. (2022). A QoE Evaluation of Haptic and Augmented Reality Gait Applications via Time and Frequency-Domain Electrodermal Activity (EDA) Analysis. *Proceedings - 2022 IEEE International Symposium on Mixed and Augmented Reality Adjunct, ISMAR-Adjunct 2022*, 297–302. https://doi.org/10.1109/ISMAR-ADJUNCT57072.2022.00067

Rodriguez, N. (2018). Identifying Accessibility Conditions for Children with Multiple Disabilities: A Virtual Reality Wheelchair Simulator. *Adjunct Proceedings - 2018 IEEE International Symposium on Mixed and Augmented Reality, ISMAR-Adjunct 2018*, 370–372. https://doi.org/10.1109/ISMAR-ADJUNCT.2018.00107

Ronca, V., Martinez-Levy, A. C., Vozzi, A., Giorgi, A., Aricò, P., Capotorto, R., Borghini, G., Babiloni, F., & Di Flumeri, G. (2023). Wearable Technologies for Electrodermal and Cardiac Activity Measurements: A Comparison between Fitbit Sense, Empatica E4 and Shimmer GSR3+. *Sensors 2023, Vol. 23, Page 5847*, *23*(13), 5847. https://doi.org/10.3390/S23135847

Rushton, P. W., Kirby, R. L., Routhier, F., & Smith, C. (2016). Measurement properties of the Wheelchair Skills Test – Questionnaire for powered wheelchair users*. *Disability and Rehabilitation: Assistive Technology*, *11*(5), 400–406. https://doi.org/10.3109/17483107.2014.984778

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

Schober, P., & Schwarte, L. A. (2018). Correlation coefficients: Appropriate use and interpretation. *Anesthesia and Analgesia*, *126*(5), 1763–1768. https://doi.org/10.1213/ANE.0000000000002864

Sepich, N. C., Jasper, A., Fieffer, S., Gilbert, S. B., Dorneich, M. C., & Kelly, J. W. (2022). The impact of task workload on cybersickness. *Frontiers in Virtual Reality*, *3*, 943409. https://doi.org/10.3389/FRVIR.2022.943409/BIBTEX

Sheridan, T. B. (1996). Further Musings on the Psychophysics of Presence. *Presence: Teleoperators and Virtual Environments*, *5*(2), 241–246. https://doi.org/10.1162/PRES.1996.5.2.241

Shirley Ryan Abilitylab. (n.d.). *Rehabilitation Measures*. Retrieved April 14, 2025, from https://www.sralab.org/rehabilitation-measures

Sinsky, C. A., Bavafa, H., Roberts, R. G., & Beasley, J. W. (2021). Standardization vs Customization: Finding the Right Balance. *The Annals of Family Medicine*, *19*(2), 171–177. https://doi.org/10.1370/AFM.2654

Smith, E. M., Mortenson, W. Ben, Mihailidis, A., & Miller, W. C. (2022). Understanding the task demands for powered wheelchair driving: a think-aloud task analysis. *Disability and Rehabilitation: Assistive Technology*, *17*(6), 695–702. https://doi.org/10.1080/17483107.2020.1810335

Software as a Medical Device (SaMD) | FDA. (n.d.). Retrieved March 27, 2025, from https://www.fda.gov/medical-devices/digital-health-center-excellence/software-medical-device-samd

Solhjoo, S., Haigney, M. C., McBee, E., van Merrienboer, J. J. G., Schuwirth, L., Artino, A. R., Battista, A., Ratcliffe, T. A., Lee, H. D., & Durning, S. J. (2019). Heart Rate and Heart Rate Variability Correlate with Clinical Reasoning Performance and Self-Reported Measures of Cognitive Load. *Scientific Reports*, *9*(1), 1–9. https://doi.org/10.1038/S41598-019-50280-3;SUBJMETA

Somarathna, R., Elvitigala, D. S., Yan, Y., Quigley, A. J., & Mohammadi, G. (2023). Exploring User Engagement in Immersive Virtual Reality Games through Multimodal Body Movements. *Proceedings of the ACM Symposium on Virtual Reality Software and Technology, VRST*. https://doi.org/10.1145/3611659.3615687;PAGE:STRING:ARTICLE/CHAPTER

Sonar, A. V., Burdick, K. D., Begin, R. R., Resch, E. M., Thompson, E. M., Thacher, E., Searleman, J., Fulk, G., & Carroll, J. J. (2005). Development of a Virtual Reality-based power wheel chair simulator. *IEEE International Conference on Mechatronics and Automation, ICMA 2005*, 222–229. https://doi.org/10.1109/ICMA.2005.1626551

Spaeth, D. M., Mahajan, H., Karmarkar, A., Collins, D., Cooper, R. A., & Boninger, M. L. (2008). Development of a Wheelchair Virtual Driving Environment: Trials With Subjects With Traumatic Brain Injury. *Archives of Physical Medicine and Rehabilitation*, *89*(5), 996–1003. https://doi.org/10.1016/j.apmr.2007.11.030

Sprigle, S., Maurer, C., & Holowka, M. (2007). Development of Valid and Reliable Measures of Postural Stability. *The Journal of Spinal Cord Medicine*, *30*(1), 40–49. https://doi.org/10.1080/10790268.2007.11753913

Sprigle, S., Wootten, M., Sawacha, Z., & Theilman, G. (2003). Relationships Among Cushion Type, Backrest Height, Seated Posture, And Reach Of Wheelchair Users With Spinal Cord Injury. *The Journal of Spinal Cord Medicine*, *26*(3), 236–243. https://doi.org/10.1080/10790268.2003.11753690

Streijl, R. C., Winkler, S., & Hands, D. S. (2016). Mean opinion score (MOS) revisited: methods and applications, limitations and alternatives. *Multimedia Systems*. https://doi.org/10.1007/s00530-014-0446-1

Sue, S. (2007). Test distance vision using a Snellen chart. *Community Eye Heal*, *20*(63), 52.

Sung, J. H., Trace, Y., Peterson, E. W., Sosnoff, J. J., & Rice, L. A. (2019). Falls among full-time wheelchair users with spinal cord injury and multiple sclerosis: a comparison of characteristics of fallers and circumstances of falls. *Disability and Rehabilitation*, *41*(4), 389–395. https://doi.org/10.1080/09638288.2017.1393111

Sweller, J. (1994). Cognitive load theory, learning difficulty, and instructional design. *Learning and Instruction*, *4*(4), 295–312. https://doi.org/10.1016/0959-4752(94)90003-5

Sweller, J. (2011). Cognitive Load Theory. In *Psychology of Learning and Motivation - Advances in Research and Theory* (Vol. 55). Elsevier Inc. https://doi.org/10.1016/B978-0-12-387691-1.00002-8

Sweller, J. (2018). Measuring cognitive load. *Perspectives on Medical Education*, *7*(1). https://doi.org/10.1007/S40037-017-0395-4

Sydor, M., Krauss, A., & Krauss, H. (2017). Risk analysis for operating active wheelchairs in non-urban settings. *Annals of Agricultural and Environmental Medicine*, *24*(3), 532–536. https://doi.org/10.5604/12321966.1235166

Tao, G., & Archambault, P. S. (2015). Using a 3D hand motion controller in a virtual power wheelchair simulator for navigation-reaching. *International Conference on Virtual Rehabilitation, ICVR*, 137–138. https://doi.org/10.1109/ICVR.2015.7358615

Tao, G., & Archambault, P. S. (2016). Powered wheelchair simulator development: Implementing combined navigation-reaching tasks with a 3D hand motion controller. *Journal of NeuroEngineering and Rehabilitation*, *13*(1), 1–13. https://doi.org/10.1186/S12984-016-0112-2/TABLES/5

Tcha-Tokey, K., Christmann, O., Loup-Escande, E., & Richir, S. (2016). Proposition and Validation of a Questionnaire to Measure the User Experience in Immersive Virtual Environments. *International Journal of Virtual Reality*, *16*(1), 33–48. https://doi.org/10.20870/IJVR.2016.16.1.2880

Tefft, D., Guerette, P., & Furumasu, J. (1999). Cognitive predictors of young children's readiness for powered mobility. *Developmental Medicine & Child Neurology*, *41*(10), 665–670. https://doi.org/10.1111/J.1469-8749.1999.TB00520.X

Tian, F., Hua, M., Zhang, W., Li, Y., & Yang, X. (2021). Emotional arousal in 2D versus 3D virtual reality environments. *PLOS ONE*, *16*(9), e0256211. https://doi.org/10.1371/JOURNAL.PONE.0256211

Timmerer, C., Ebrahimi, T., & Pereira, F. (2015). Toward a new assessment of quality. *Computer*, *48*(3), 108–110. https://doi.org/10.1109/MC.2015.89

Tombaugh, T. N. (2004). Trail Making Test A and B: Normative data stratified by age and education. *Archives of Clinical Neuropsychology*, *19*(2), 203–214. https://doi.org/10.1016/S0887-6177(03)00039-8

Torkia, C., Reid, D., Korner-Bitensky, N., Kairy, D., Rushton, P. W., Demers, L., & Archambault, P. S. (2015). Power wheelchair driving challenges in the community: a users' perspective. *Disability and Rehabilitation: Assistive Technology*, *10*(3), 211–215. https://doi.org/10.3109/17483107.2014.898159

Torkia, C., Ryan, S. E., Reid, D., Boissy, P., Lemay, M., Routhier, F., Contardo, R., Woodhouse, J., & Archambault, P. S. (2019). Virtual community centre for power wheelchair training: Experience of children and clinicians. *Disability and Rehabilitation: Assistive Technology*, *14*(1), 46–55. https://doi.org/10.1080/17483107.2017.1392622;PAGE:STRING:ARTICLE/CHAPTER

Townsend, K., & Unsworth, C. A. (2019). The inter-rater reliability of the Powered Mobility Device Assessment Training Tool. *Australian Occupational Therapy Journal*, *66*(3), 393–400. https://doi.org/10.1111/1440-1630.12566

Toy, S., Miller, C. R., Guris, R. J. D., Duarte, S. S., Koessel, S., & Schiavi, A. (2020). Evaluation of 3 cognitive load measures during repeated simulation exercises for novice anesthesiology residents. *Simulation in Healthcare*, *15*(6), 388–396. https://doi.org/10.1097/SIH.0000000000000458

Tu, C. J., Liu, L., Wang, W., Du, H. P., Wang, Y. M., Xu, Y. B., & Li, P. (2017). Effectiveness and safety of wheelchair skills training program in improving the wheelchair skills capacity: a systematic review. *Clinical Rehabilitation*, *31*(12), 1573–1582. https://doi.org/10.1177/0269215517712043

Ute Jekosch. (2005). Voice and Speech Quality Perception. *Voice and Speech Quality Perception*. https://doi.org/10.1007/3-540-28860-0

Vailland, G., Devigne, L., Pasteau, F., Nouviale, F., Fraudet, B., Leblong, E., Babel, M., & Gouranton, V. (2021). VR based power wheelchair simulator: Usability evaluation through a clinically validated task with regular users. *Proceedings - 2021 IEEE Conference on Virtual Reality and 3D User Interfaces, VR 2021*, 420–427. https://doi.org/10.1109/VR50410.2021.00065

Vailland, G., Gaffary, Y., Devigne, L., Gouranton, V., Arnaldi, B., & Babel, M. (2020). Vestibular Feedback on a Virtual Reality Wheelchair Driving Simulator: A Pilot Study. *Proceedings of the 2020 ACM/IEEE International Conference on Human-Robot Interaction*, 171–179. https://doi.org/10.1145/3319502.3374825

Vailland, G., Grzeskowiak, F., Devigne, L., Gaffary, Y., Fraudet, B., Leblong, E., Nouviale, F., Pasteau, F., Breton, R. Le, Guegan, S., Gouranton, V., Arnaldi, B., & Babel, M. (2019). User-centered design of a multisensory power wheelchair simulator: Towards training and rehabilitation applications. *IEEE International Conference on Rehabilitation Robotics*, *2019-June*, 77–82. https://doi.org/10.1109/ICORR.2019.8779496

Valentini, C. A. M., Martins, F. R., de Sá, A. A. R., & Naves, E. L. M. (2024). Training protocol for driving power wheelchairs using virtual environment: preliminary results from a pilot study. *Health and Technology*, *14*(2), 339–350. https://doi.org/10.1007/s12553-023-00811-w

Van Voorhees, E. E., Dennis, P. A., Watkins, L. L., Patel, T. A., Calhoun, P. S., Dennis, M. F., & Beckham, J. C. (2022). Ambulatory Heart Rate Variability Monitoring: Comparisons Between the Empatica E4 Wristband and Holter ECG. *Psychosomatic Medicine*, *84*(2), 210. https://doi.org/10.1097/PSY.0000000000001010

Vincenzi, D., Mouloua, M., Hancock, P. A., Pharmer, J. A., & Ferraro, J. C. (2023). Human factors in simulation and training: Theory and methods. In *Human Factors in Simulation and Training: Theory and Methods*. CRC Press. https://doi.org/10.1201/9781003401360

Vlahovic, S., Suznjevic, M., & Skorin-Kapov, L. (2022). A survey of challenges and methods for Quality of Experience assessment of interactive VR applications. *Journal on Multimodal User Interfaces 2022 16:*, *16*(3), 257–291. https://doi.org/10.1007/S12193-022-00388-0

WHO. (2008). *Guidelines on the provision of in less resourced settings*.

WHO. (2011). *World Report on Disability*. https://www.who.int/teams/noncommunicable-diseases/sensory-functions-disability-and-rehabilitation/world-report-on-disability

WHO. (2022). Global report on assistive technology. *Global Report on Assistive TechnologySummary*, 1–11. https://www.who.int/publications/i/item/9789240049451

WHO. (2023). *Wheelchair provision guidelines*. https://www.who.int/publications/i/item/9789240074521

Widyanti, A., Hanna, Muslim, K., & Sutalaksana, I. Z. (2017). The sensitivity of Galvanic Skin Response for assessing mental workload in Indonesia. *Work*, *56*(1), 111–117. https://doi.org/10.3233/WOR-162479

Wiebe, E. N., Roberts, E., & Behrend, T. S. (2010). An examination of two mental workload measurement approaches to understanding multimedia learning. *Computers in Human Behavior*, *26*(3), 474–481. https://doi.org/10.1016/J.CHB.2009.12.006

Wilhelm, F. H., Pfaltz, M. C., & Grossman, P. (2006). *Distinguishing emotional from physical activation in ambulatory psychophysiological monitoring.* https://www.researchgate.net/publication/6966605

Witmer, B. G., & Singer, M. J. (1998). Measuring Presence in Virtual Environments: A Presence Questionnaire. *Presence: Teleoperators and Virtual Environments*, *7*(3), 225–240. https://doi.org/10.1162/105474698565686

World Vision. (2017). *The role of community development organizations in providing holistic wheelchair services*. https://www.google.com/url?sa=t&source=web&rct=j&opi=89978449&url=https://www.wvi.org/sites/default/files/The%25208%2520Steps%2520%2B%2520FINAL%2520Compressed%2520File.pdf&ved=2ahUKEwi46KT_h8uMAxXyVEEAHUk4PCsQFnoECCEQAQ&usg=AOvVaw1-6J-URNTTD7tjUdAyYmTM

Wu, H., Jiao, Y., Jiang, C., Wang, T., & Yu, J. (2025). Fatigue state evaluation of urban railway transit drivers using psychological, biological, and physical response signals. *IEEE Access*. https://doi.org/10.1109/ACCESS.2025.3533487

Xiang, H., Chany, A. M., & Smith, G. A. (2006). Wheelchair related injuries treated in US emergency departments. *Injury Prevention*, *12*(1), 8. https://doi.org/10.1136/IP.2005.010033

Xiong, X., & Dubin, J. A. (2010). A binning method for analyzing mixed longitudinal data measured at distinct time points. *Statistics in Medicine*, *29*(18), 1919–1931. https://doi.org/10.1002/sim.3953

Young, T. P., Guptill, M., Thomas, T., & Mellick, L. (2018). Effective Educational Videos in Emergency Medicine. *AEM Education and Training*, *2*, S17–S24. https://doi.org/10.1002/aet2.10210

Younis, H., Ramzan, F., Khan, J., & Ghani Khan, M. U. (2019). Wheelchair training virtual environment for people with physical and cognitive disabilities. *15th International Conference on Emerging Technologies, ICET 2019*. https://doi.org/10.1109/ICET48972.2019.8994550

Zatla, H., Hadj-Abdelkader, A., Morere, Y., & Bourhis, G. (2015). OPCM model application on a 3D simulator for powered wheelchair. *International Conference on Virtual Rehabilitation, ICVR*, 131–132. https://doi.org/10.1109/ICVR.2015.7358611

Zatla, H., Morère, Y., Hadj-Abdelkader, A., Bourhis, G., Demet, K., Guilmois, G., Bigaut, N., & Cosnuau, K. (2018). Preview Distance Index for the Analysis of Powered Wheelchair Driving. *IRBM*, *39*(3), 194–205. https://doi.org/10.1016/J.IRBM.2018.03.001

Zihisire Muke, P., Telec, Z., & Trawiński, B. (2022). Cognitive Load Measurement Using Arithmetic and Graphical Tasks and Galvanic Skin Response. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *13501 LNAI*, 836–850. https://doi.org/10.1007/978-3-031-16014-1_66

Zorzi, C., Covaci, A., & Marcelli, G. (2024). Using Virtual Reality to Complement Traditional Wheelchair Skills Training Methods: A Literature Review. *Applied Sciences*, *15*(1), 187. https://doi.org/10.3390/app15010187

Zorzi, C., Tabbaa, L., Covaci, A., Sirlantzis, K., & Marcelli, G. (2023a). A Standardized and Cost-Effective VR Approach for Powered Wheelchair Training. *IEEE Access*, *11*, 66921–66933. https://doi.org/10.1109/ACCESS.2023.3288424

Zorzi, C., Tabbaa, L., Covaci, A., Sirlantzis, K., & Marcelli, G. (2023b). Train vs. Play: Evaluating the Effects of Gamified and Non-Gamified Wheelchair Skills Training Using Virtual Reality. *Bioengineering 2023, Vol. 10, Page 1269*, *10*(11), 1269. https://doi.org/10.3390/BIOENGINEERING10111269

Chapter 7 Appendices

APPENDIX A. CHAPTER 4 LITERATURE REVIEW TABLE

Table 7.1 - List of QoE Assessment Methods

Assessment Type	Tool / Metric	Target QoE Dimension(s)	Notes
Subjective	System Usability Scale (SUS)	Usability, satisfaction	Standardized, robust, easy to deploy
	iGroup Presence Questionnaire (IPQ)	Immersion, spatial presence	Measures realism and user involvement
	Self-Assessment Manikin (SAM)	Emotional valence, arousal, dominance	Pictorial, low cognitive demand
	Simulator Sickness Questionnaire (SSQ)	Cybersickness, oculomotor discomfort	Safety-oriented; VR-specific
	Absolute Category Rating (ACR) + MOS	Overall comfort, control, visual quality	Simple scalar feedback; MOS used to summarize across users
	Game Experience Questionnaire (GEQ)	Immersion, flow, emotional experience	Gaming and interactive training contexts
	Player Experience Inventory (PXI)	Competence, affect, sensory and challenge	Covers deeper engagement and flow
	Virtual Reality Neuroscience Questionnaire	Comfort, realism, symptoms	Combines UX and simulator side effects
	Tcha-Tokey's VR UX Questionnaire	Flow, presence, skill, engagement	Broad-spectrum for VR UX and QoE
Implicit (Physiological)	Electrodermal Activity (EDA)	Arousal, attentional/mental effort	Real-time, wearable, continuous
	Heart Rate (HR), Heart Rate Variability (HRV)	Stress, engagement, emotional load	Sensitive to mental state variation
Implicit (Behavioural)	Task performance (errors, time, control)	Usability, load, frustration	Proxy for experience quality and control demand
	Eye-tracking (gaze, blink, saccades)	Attention, fatigue, task engagement	Suitable for adaptive systems
	Voice analysis (tone, pitch, hesitation)	Frustration, stress, emotional fluctuation	Passive, emerging modality
	Facial expression recognition	Engagement, frustration, boredom	Often integrated into modern HMDs and webcams
	Postural/motion behaviours (e.g., sway)	Discomfort, immersion, balance	Non-intrusive physical engagement indicator

APPENDIX B. LIST OF FEATURES FROM WHEELSIMPHYSIO-2023 DATASET

Table 7.2 - List of features from WheelSimPhysio-2023 dataset.

Data type	Classification	Metrics	Description
Demographic Information	Participant General Information	Participant	A unique identifier for each participant in the study
		Experiment	The specific session ID the participant took part in.
		metrics_type	denotes the type of metrics being collected (baseline, test or difference (test-baseline))
		gender	The gender of the participant
		age	the age of the participants (years)
		dominant_hand	Indicates whether the participant is right-handed, left-handed, or ambidextrous.
		play_games	whether the participant play video games
		drive	Whether the participant has car drive license
		vr_experience	The participant's prior experience using virtual reality technology
Physiological Features (Objective Implicit Metrics)	Inter-beat Interval (IBI)	meanIBI	The average time interval between successive heartbeats, measured in milliseconds (ms).
		sdnnIBI	The standard deviation of normal-to-normal (NN) intervals, reflecting overall heart rate variability.
		rmssdIBI	The root mean square of successive differences in NN intervals, indicating short-term variability.
		nn50IBI	The number of pairs of successive NN intervals differing by more than 50 ms, reflecting variability.
		pnn50IBI	The proportion of NN50 divided by the total number of NN intervals, representing heart rate variability.
	Hear Rate (HR)	meanHR	The average heart rate, measured in beats per minute (bpm).
		maxHR	The maximum heart rate recorded during the experiment.
		minHR	The minimum heart rate recorded during the experiment.
		RangeHR	The difference between the maximum and minimum heart rate.
		sdHR	The standard deviation of the heart rate over the experiment period.
	Electrodermal Activity (EDA)	meanSCRAmplitude	The average amplitude of skin conductance responses (SCR), reflecting sympathetic nervous system activity.

	also known as Galvanic Skin Response (GSR)	scrCount	The number of skin conductance responses recorded during the experiment.
		meanSCL	The average skin conductance level (SCL), representing overall arousal levels.
		meanSCRRiseTime	The average time it takes for SCR to reach its peak after stimulus onset.
		meanSCRRecoveryTime	The average time it takes for SCR to return to baseline after reaching its peak.
		F0SC	Frequency-domain features representing different frequency bands in the SCR signal. The spectral power in bandwidths 0.05 to 0.1
		F1SC	The spectral power in bandwidths 0.1 to 0.2
		F2SC	The spectral power in bandwidths 0.2 to 0.3
		F3SC	The spectral power in bandwidths 0.3 to 0.4
		meanFirstDerivative	The mean rate of change of the SCR signal over time, indicating the velocity of conductance changes.
		meanSecondDerivative	The mean acceleration of the SCR signal, showing the rapidity of conductance changes.
Wheelchair Simulator (System)	Performance	numCollisions	The number of collisions the participant made while navigating the wheelchair simulator.
		numCommandChanges	The number of times the participant altered control inputs during the simulation.
		totalTime	The total time taken by the participant to complete the simulation task.
Post-Experiment Feedback	Questionnaire (Subjective Data)	Valence (SAM))	Emotional valence, indicating how pleasant or unpleasant the participant felt during the experiment.
		Arousal (SAM)	The level of excitement or alertness the participant experienced during the experiment.
		Dominance (SAM)	The degree of control the participant felt over the environment and tasks during the experiment.
		Immersion (IPQ)	The level of immersion the participant experienced while using the wheelchair simulator and VR.
		Usability (SUS)	A measure of how usable or intuitive the participant found the system.
		Cognitive Load (NASA – TLX)	A measure of perceived workload across several dimensions using the NASA Task Load Index (TLX).
		Cybersickness (SSQ)	The Simulator Sickness Questionnaire (SSQ) score, assessing symptoms of discomfort from VR.

APPENDIX C. EDA FILTER PSEUDOCODE

Algorithm 1 Butterworth Low-Pass Filtering of EDA Signal

Input:
  eda = Raw EDA signal
  $fs = 4$ Hz : Sampling frequency
  $fc = 1$ Hz : Cutoff frequency
  $n = 5$: Initial filter order
  $type$ = 'low' : Filter type (low-pass)
Output: eda_out : Filtered EDA signal or empty if filtering fails

Step 1: Initialize output signal $eda\_out \leftarrow eda$
Step 2: Compute normalized cutoff frequency: $wn \leftarrow fc \ / \ (fs \ / \ 2)$
Step 3: Set initial filter order: $current\_order \leftarrow n$
Step 4: Compute Butterworth filter: $[B, A] \leftarrow \text{butter}(current\_order, wn, type)$
while all elements in $B < \epsilon$ (e.g., 1×10^{-12}) **and** $current\_order > 0$ **do**
  Step 5: Decrease filter order: $current\_order \leftarrow current\_order - 1$
  Step 6: Recompute filter: $[B, A] \leftarrow \text{butter}(current\_order, wn, type)$
end while
if length of $eda\_out > 3 \times n$ **then**
  Step 7: Apply zero-phase filtering: $eda\_out \leftarrow \text{filtfilt}(B, A, eda\_out)$
else
  Step 8: Set output to empty: $eda\_out \leftarrow \emptyset$
  Step 9: Display warning: "Signal too short for stable filter application"
end if
Return eda_out

Procedure for Butterworth low-pass filtering of EDA signals. To reduce high-frequency noise while preserving phasic activity in EDA signals, a zero-phase digital low-pass Butterworth filter was applied. This filtering approach ensures the preservation of signal components relevant to sympathetic arousal. The filter was implemented with the following parameters:

- Filter type: Low-pass
- Filter order: 5 (adaptively reduced if instability detected)
- Cutoff frequency: 1 Hz
- Sampling rate: 4 Hz

The cutoff frequency was normalized by dividing by the Nyquist frequency (fs/2), as required for digital filter design. The resulting normalized cutoff (wn) was used to compute the Butterworth filter coefficients (numerator B and denominator A). To ensure numerical stability, especially for short or

low-variance signals, an adaptive routine was implemented: if all elements of the coefficient vector B were below a small threshold (1e−12), the filter order was reduced iteratively until a stable design was obtained.

Once stable coefficients were determined, zero-phase filtering was applied using a forward and reverse filtering operation (filtfilt). This approach eliminates phase distortion, which is critical for preserving the timing of physiological responses in relation to experimental events. A minimum signal length criterion (>3× filter order) was enforced to ensure sufficient data for effective filtering. If this condition was not met, the filtering step was skipped, and a warning was issued.

This filter design was implemented to preserve the integrity of the EDA signal's phasic components, particularly those associated with stimulus-related sympathetic arousal, while minimizing contamination from high-frequency artefacts.

APPENDIX D. STUDY 1 AND STUDY 2 INFORMATION

a. INFORMATION SHEET

1. Project Title: A Quality of Experience Evaluation of Wheelchair Simulator in Virtual Reality Environments

2. Introduction

In this experiment, we aim to evaluate user quality of experience when using a traditional desktop setup and with a virtual reality head mounted display. Participants will be randomly divided into two groups (desktop and Headset group), a script will be used in the MATLAB software to randomly select which group the participants will be divided. Using the conventional screen allows users to view information and interact with virtual objects whilst being aware of their surroundings that is not relate with the simulator. Virtual reality head mounted displays project virtual objects to the wearers field of view, however they are not aware of their surroundings. We aim to capture data using both devices in order to determine if a user's quality of experience is enriched using an conventional desktop setup (common monitor) or virtual reality HMDs device.

I am inviting you to take part in a research experiment to be carried out in the Software Research Institute in Athlone Institute of Technology. The aim of this document is to explain why the research is being carried out and what it will involve.

If you are not clear on any points, please do not hesitate ask questions. Thank you for reading this.

3. What is the purpose of the project?

Virtual Reality (VR) technology has improved significantly in recent years to the point where recreated virtual scenes now provide the user with a sense of realism. With the improvement of graphics cards, increased computational power, 360-degree 3D HD cameras and many more technological advances, these are rapidly changing the future potential of VR. In this experiment, we aim to evaluate if a higher quality of experience is experienced using traditional desktop compared to VR Head mounted display device . The aim of this experiment is to evaluate quality of experience within PC monitor screen and HDM while using an wheelchair training simulator.

4. Do I have to take part?

It is entirely up to you to decide whether you wish to take part in this experiment. Refusal to take part is entirely at your discretion. If you decide to take part, you can keep this information sheet and will be required to sign a consent form.

5. What does the experiment involve?

This experiment should last a minimum of 25 mins and no more than 30 min. Participants will be seated in a laboratory in the AIT Engineering Building. The lab will consist of desktop computer, or VR headset, the non-invasive objective metric sensor (E4 and Mindwave)which the E4 will be attached to the participant's wrist and Mindwave will be put on the head. The test will involve exposing participants to a virtual scene using an conventional monitor or virtual reality head mounted display. The virtual scene consists of circuit with obstacles where the user must drive a wheelchair without commit collisions. The participant will be required to use USB joystick or keyboard to interact with the virtual environment while wearing non-intrusive equipment . The participant will be asked to fill out a questionnaire at the end of the experiment to give their thought on the quality of experience.

6. What do I have to do?

On the day of the test, participants will undergo a visual screening to ensure they are eligible for the test. The visual screening process involves testing the participant's visual clarity using a Snellen chart, and testing a participant's colour perception using the Ishihara test.

If you are pregnant or suspect that you may be pregnant, please let the administrator of the test know.

If you did not sleep at least 6 hours on the previous night, please let the administrator of test know.

If you consumed alcohol in the last 24 hours, please let the administrator of test know.

7. What are the possible disadvantages and risks of taking part?

Some users may feel some nausea when using a head mounted display (VR), but this soon goes away after removing the headset. Should a participant at any point feel a high level of nausea it is important to communicate this to the PI.

8. Will my taking part in this project be kept confidential?

Any information collected during this test will be strictly confidential. All data will be stored in a secure manor, and it will not be possible to recognise you from this experiment.

9. What will happen to the results of the research project?

The results of this experiment will be used to produce a paper for publication as part of my research.

10. Thanks!

Just like to say, thank you very much for your time and help with this experiment.

b. CONSENT FORM

Project Title: A Quality of Experience Evaluation of a Wheelchair Training Simulator in Virtual Reality Environments.
Name of Researcher: Débora Pereira Salgado
Please Tick the Box

1. **I am satisfied that I understand the information provided and have had enough time to consider the information.** ☐
2. I do not suffer from photosensitive epilepsy or any other form of epilepsy. ☐
3. I'm not pregnant and/or I am not experiencing any symptoms of pregnancy. ☐
4. I have not consumed alcohol beverages for the last 24 hours. ☐
5. I slept at least 6 hours on last 24 hours. ☐
6. I understand that my participation is voluntary and that I am free to withdraw at any time, without giving any reason, without my legal rights being affected. ☐
7. I understand that any data collected in the course of this study will be used for research purpose only and in the strictest confidence. Any information related to me will be discarded at the completion of this research. ☐
8. I agree to take part in the above study. ☐
9. I confirm that I have read the information sheet dated ___/___/___ for the above study and have had the opportunity to ask questions. ☐
10. Gender: _____
11. Age: _____

_____	_____	_____
Name of Participant	Date	Signature
_____	_____	_____
Name of Person taking consent (if different from researcher)	Date	Signature
_____	_____	_____
Researcher	Date	Signature

c. QUESTIONNAIRES

Questionnaire – Part 1

You will see some statements about experiences. Please indicate, whether or not each statement applies to your experience. There are no right or wrong answers, only your opinions counts. And please remember: Answer all these questions only referring to this one experience. Read through them to make sure you understand the statement. If you have any questions, please ask your administrator.

1. I found the system unnecessarily complex.

|_____|_____|_____|_____|_____|

Strongly Disagree Neutral Agree Strongly
Disagree Agree

2. I thought the system was easy to use.

|_____|_____|_____|_____|_____|

Strongly Disagree Neutral Agree Strongly
Disagree Agree

3. I would imagine that people would learn to use the wheelchair training simulator very quickly.

|_____|_____|_____|_____|_____|

Strongly Disagree Neutral Agree Strongly
Disagree Agree

4. I found the system very difficult to use.

|_____|_____|_____|_____|_____|

Strongly Disagree Neutral Agree Strongly
Disagree Agree

5. I needed to learn a lot of things before I could get going with this system.

|_____|_____|_____|_____|_____|

Strongly Disagree Neutral Agree Strongly
Disagree Agree

6. In the computer generated world I had a sense of "being there".

|_____|_____|_____|_____|_____|

Not at all Moderate Very much

7. I had a sense of acting in the virtual space, rather than operating something from outside.

|_____|_____|_____|_____|_____|

Strongly Disagree Neutral Agree Strongly
Disagree Agree

8. I felt present in the virtual space.

|_____|_____|_____|_____|_____|

Strongly Disagree Neutral Agree Strongly
Disagree Agree

9. How aware were you of the real world surrounding while navigating in the virtual world? (i.e. sounds, room temperature, other people, etc.)?

|_____|_____|_____|_____|_____|

Extreme Moderate Not aware
aware at all

10. How much did your experience in the virtual environment seem consistent with your real world experience?

|_____|_____|_____|_____|_____|

Extreme Moderate Not
Consistent consistent
at all

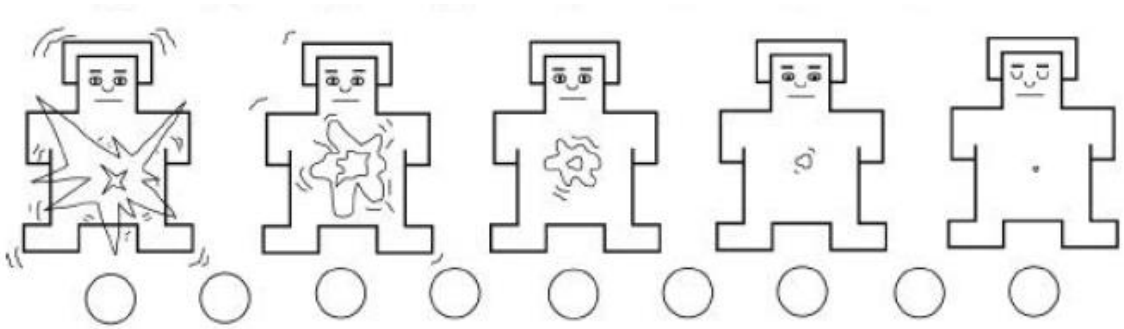
Questionnaire – Part 2

You will be asked to rate your emotions towards to experience in using the Simulator. It will be asked to rate on three separate scales.

Rating Scales

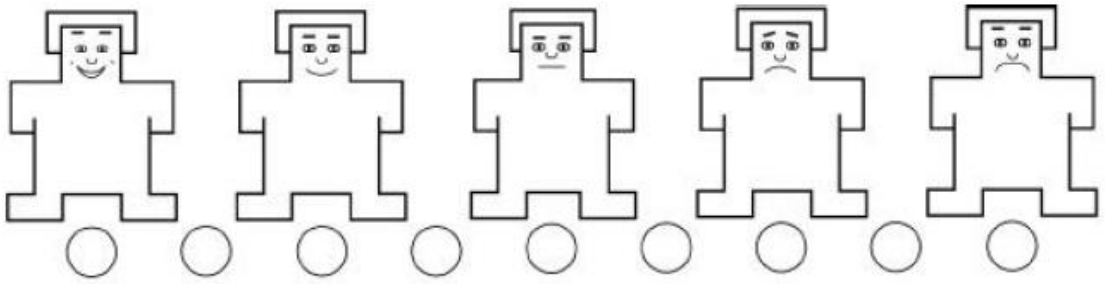
Valence (Pleasant level)

Unpleasant Unsatisfied Neutral Pleased Pleasant



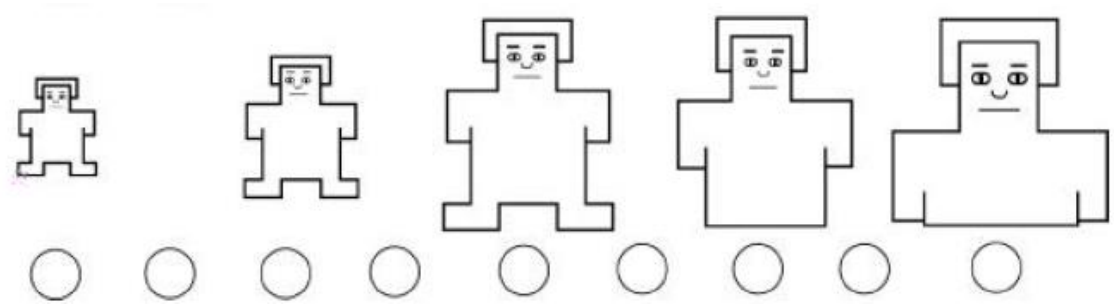
Arousal (Excitement level)

Excited Wide-awake Neutral Dull Calm



Dominance (Emotion Control level)

Dependent Powerlessness Neutral Powerful Independent



Questionnaire – Part 3

The evaluation you are about to perform is a technique that has been developed by Nasa to assess the relative importance of six factors in determining how much workload you experienced while preforming a task that you recently completed. These six factors are defined below on this page.

Read through them to make sure you understand what each factor means. If you have any questions, please ask your administrator.

Workload factors	Definition
Mental Demand Level (low/high)	**How much mental add perceptual activity was required** (for example, thinking, deciding, calculating, remembering, looking, searching, etc)? Was the task easy or demanding, simple or complex, forgiving or exacting?
Physical Demand Level (low/high)	**How much physical activity was required** (for example, pushing, pulling, turning, controlling, activating, etc.)? Was the task easy or demanding, slow or brisk, slack or strenuous, restful or laborious?
Temporal Demand Level (low/high)	**How much time pressure did you feel** due to the rate or pace at which the tasks or task elements occurred? Was the pace slow and leisurely or rapid and frantic?
Performance Level(good/poor)	How successful do you think you were in accomplish the goals of the task set by the experimenter (or yourself)? How satisfied were you with your performance in accomplish these goals?
Effort Level (low/high)	**How hard did you have to work** (mentally and physically) to accomplish your level of performance?
Frustration Level (low/high)	**How insecure, discouraged, irritated, stressed**, and annoyed versus secure, gratified, content, relaxed, and complacent did you feel during the task?

For each pair, choose the factor that was more important to your experience of the workload in the task that you recently performed:		
1	☐ Temporal Demand	☐ Mental Demand
2	☐ Performance	☐ Mental Demand
3	☐ Mental Demand	☐ Effort
4	☐ Temporal Demand	☐ Effort
5	☐ Physical Demand	☐ Performance
6	☐ Performance	☐ Temporal Demand
7	☐ Effort	☐ Physical Demand
8	☐ Mental Demand	☐ Physical Demand
9	☐ Performance	☐ Frustration
10	☐ Effort	☐ Performance
11	☐ Frustration	☐ Effort
12	☐ Frustration	☐ Mental Demand
13	☐ Physical Demand	☐ Temporal Demand
14	☐ Physical Demand	☐ Frustration
15	☐ Temporal Demand	☐ Frustration

You will now be presented with a Series of rating scales.

For each of the six scales, evaluate the task you recently performed by cross on the scale's location that matches your experience. Each line has two endpoint that describe the scale.

Consider your responses carefully in distinguishing among the different task conditions, and consider each individually.

Mental Demand (How mentally demanding was the task?/ How much mental and perceptual activity did you spend for this task?)

Very Low Very High

Physical Demand (How physically demanding was the task?/ How much physical activity did you spend for this task?)

Very Low Very High

Temporal Demand (How hurried or rushed was the pace of the task?/ How much time pressure did you feel in order to complete this task?)

Very Low Very High

Performance (How successful were you in accomplishing what you were asked to do?/ How successful do you think you were in accomplishing the goals of the task?)

Good Poor

Effort (How hard did you have to work to accomplish your level of performance?)

Very Low Very High

Frustration (How insecure, discouraged, irritated, stressed, and annoyed were you during this task?)

Very Low Very High

Questionnaire – Part 4

Pre- and Post- exposure Simulator Sickness Questionnaire

SYMPTOM CHECKLIST

Pre-exposure instructions: please fill in this questionnaire. Circle below if any of the symptoms apply to you now. You will be asked to fill this again after the experiment

1. General discomfort	None	Slight	Moderate	Severe
2. Fatigue	None	Slight	Moderate	Severe
3. Boredom	None	Slight	Moderate	Severe
4. Drowsiness	None	Slight	Moderate	Severe
5. Headache	None	Slight	Moderate	Severe
6. Eyestrain	None	Slight	Moderate	Severe
7. Difficulty focusing	None	Slight	Moderate	Severe
8. Salivation increase	None	Slight	Moderate	Severe
Salivation decrease	None	Slight	Moderate	Severe
9. Sweating	None	Slight	Moderate	Severe
10. Nausea	None	Slight	Moderate	Severe
11. Difficulty concentrating	None	Slight	Moderate	Severe

12. Mental depression	No	Yes (	Slight	Moderate	Severe	)
13. "Fullness of the head"	No	Yes (	Slight	Moderate	Severe	)
14. Blurred vision	No	Yes (	Slight	Moderate	Severe	)
15. Dizziness eyes open	No	Yes (	Slight	Moderate	Severe	)
Dizziness eyes close	No	Yes (	Slight	Moderate	Severe	)
16. Vertigo	No	Yes (	Slight	Moderate	Severe	)
17. Visual flashbacks*	No	Yes (	Slight	Moderate	Severe	)
18. Faintness	No	Yes (	Slight	Moderate	Severe	)
19. Aware of breathing	No	Yes (	Slight	Moderate	Severe	)

20. Stomach awareness	No	Yes (	Slight	Moderate	Severe)
21. Loss of appetite	No	Yes (	Slight	Moderate	Severe)
22. Increased appetite	No	Yes (	Slight	Moderate	Severe)
23. Desire to move bowels	No	Yes (	Slight	Moderate	Severe)
24. Confusion	No	Yes (	Slight	Moderate	Severe)
25. Burping	No	Yes (	Slight	Moderate	Severe)
26. Vomiting	No	Yes (	Slight	Moderate	Severe)
27. Other _____	No	Yes (	Slight	Moderate	Severe)

APPENDIX E. STUDY 3 INFORMATION

a. INFORMED CONSENT FORM

Title of the Study: Evaluation of a Virtual Wheelchair Simulator in Assessing Mobility Skills and Cognitive Abilities in Diverse Populations: A Multicentric Mixed-Methods Pilot Study

Principal Investigator: Debora Pereira Salgado

Introduction:

You are invited to participate in a research study designed to evaluate the effectiveness of a wheelchair skill assessment and training simulator in improving the mobility skills of wheelchair users. Before deciding whether to participate, it is crucial that you understand the study's purpose, procedures, potential risks, and benefits. Please read the following information carefully. If you have any questions or need further clarification, do not hesitate to ask before making your decision.

Study Purpose:

The purpose of this study is to investigate the impact of a virtual wheelchair training simulator and skill assessment application on the power mobility skills and confidence levels of wheelchair users. By participating, you will help advance knowledge in wheelchair training methods and skill assessment and may benefit from improved mobility skills and increased confidence in navigating various environments.

Procedures:

If you agree to participate, you will be asked to:

Demographic Information: You will be asked to provide basic demographic information.

Pre-Test Assessment: Your initial mobility skills will be assessed using standardized tests, such as the Wheelchair Skills Test (WST) and Montreal Cognitive Assessment (MoCA), to understand your current abilities before the training.

Simulator Training Program: You will participate in a structured training program using the wheelchair training simulator, where you will navigate virtual scenarios that replicate real-world environments and challenges.

Guidance and Support: During the simulator sessions, experienced instructors will provide guidance and instructions as you engage with each scenario.

Post-Test Feedback: After the training, you will take part in a feedback session, which includes both a questionnaire and open-ended interview questions. This feedback will help us gather your insights and experiences with the simulator program. We would also like your permission to record this interview to ensure accurate analysis.

Data Collection and Use of Wearables:

During simulator sessions, you may be asked to wear sensors that collect physiological data, and cameras may record movements of your head and eyes to analyse interactions with the virtual environment. All data will be anonymized and used exclusively for research purposes.

Risks and Benefits:

While there are minimal physical risks associated with participating in this study, you may experience mild discomfort or fatigue during the training sessions. However, these risks will be minimized by the presence of trained instructors who will guide you throughout the training process. The potential benefits of participating include improved mobility skills, increased confidence in wheelchair navigation, and contribution to the development of effective training methods for wheelchair users.

Confidentiality:

Your identity and personal information will be kept strictly confidential. Any data collected during this study will be anonymized and stored securely. Only authorized researchers involved in this study will have access to the data. Any data related to your identity and personal information will be discarded at the completion of this research.

Voluntary Participation and Withdrawal:

Participation in this study is entirely voluntary. You have the right to refuse participation or withdraw from the study at any time, without penalty or impact on your medical care. Your decision to participate or withdraw will not affect your current or future relationship with the research team or the institution.

Contact Information:

If you have any questions or concerns regarding this study, you can contact the principal investigator, Debora Pereira Salgado, or IWA’s manager Paul Ryan

Please Answer (YES OR NO) for the following statements:

I am satisfied that I understand the information provided and have had enough time to consider the information.	
I do not suffer from photosensitive epilepsy or any other form of epilepsy.	
I'm not pregnant and/or I am not experiencing any symptoms of pregnancy.	
I have not consumed alcohol beverages for the last 24 hours.	
I slept at least 6 hours on the last 24 hours.	
I do not have physical limitations that prevent me from safely operating the wheelchair with a joystick.	
I do not have cognitive limitations that prevent me from understanding and answering the questionnaire attached to this form and following instructions while using the simulator.	
I do not have pre-existing medical conditions that could be exacerbated or worsened by the use of wheelchair simulator (e.g., uncontrolled seizures, or recent orthopaedic surgeries).	
I do not have a pre-condition for motion sickness or simulator sickness.	
I understand that my participation is voluntary and that I am free to withdraw at any time, without giving any reason, without my legal rights being affected.	
I understand that any data collected during this study will be used for research purpose only and in the strictest confidence.	
I confirm that I have more than 18 years old.	
I have consent from IWA's clinical team to participate the above study.	
I agree to take part in the above study.	
I confirm that I have read the information sheet dated ___/___/__ for the above study and have had the opportunity to ask questions.	

By signing this form, you indicate that you have read and understood the information provided, and voluntarily agree to participate in the study.

Participant's Name: _____

Participant's Signature: _____

Researcher's Name: _____

Researcher's Signature: _____

Date: _____

b. PARTICIPATION INFORMATION SHEET

Title of the Study: Evaluation of a Virtual Wheelchair Simulator in Assessing Mobility Skills and Cognitive Abilities in Diverse Populations: A Multicentric Mixed-Methods Pilot Study

Principal Investigador: Debora Pereira Salgado

Thank you for your interest in participating in our study. Before you make a decision, we would like to provide you with some important information about the study's purpose, procedures, and what your involvement would entail. Please take the time to read this information carefully. If you have any questions or concerns, feel free to contact us for clarification before making your decision.

Study Overview:

The aim of this study is to evaluate the effectiveness of a wheelchair training simulation in enhancing the mobility skills of wheelchair users like yourself. By participating, you will engage in a wheelchair mobility scenarios program that utilizes a virtual simulator to replicate real-life scenarios and challenges faced by wheelchair users.

Participant Involvement:

If you choose to participate, here is what your involvement in the study will entail:

1. Initial Assessment: You will undergo an initial assessment of your current mobility skills and cognitive abilities, including standardized tests such as the Wheelchair Skills Test and Montreal Cognitive Assessment (MoCA), administered by trained professionals.

2. Simulator Sessions: You will use a virtual reality wheelchair simulator, guided by experienced instructors who will support you throughout. You will perform tasks in various simulated scenarios that mirror real-life environments, allowing you to practice and potentially improve your mobility skills in a safe and controlled setting. Your participation may involve up to three sessions, depending on the study requirements.

3. Data Collection: During the study, data will be collected to evaluate the effectiveness of the training simulation. This may include objective measurements of your mobility skills, self-reported assessments of your confidence levels, and feedback regarding your experiences with the simulator.

4. Use of Wearables and Camera: During the simulator session, you may be asked to wear devices that capture physiological data (e.g., wearable sensors) and cameras may be used to record head and eye movements to analyse your interactions with the virtual environment. All recordings will be used strictly for research analysis and will be anonymized.

5. Post-Experience Interview: After the simulator session, we will conduct a post-experience interview to gather further feedback. With your permission, we would like to record this interview to ensure accurate annotation and analysis.

6. Confidentiality and Ethical Considerations: Your participation in this study will be treated with the utmost confidentiality. All data collected during the study will be anonymized and stored securely. Participation is entirely voluntary, and you have the right to withdraw from the study at any time without penalty or impact on your medical care or support services.

7. Potential Benefits and Risks: Potential Benefits: By participating in this study, you may experience improved mobility skills and increased confidence in navigating various environments. Your involvement will also contribute to the advancement of knowledge and the development of more effective training methods for wheelchair users. Risks: While there are minimal physical risks associated with the simulator training, you may experience mild discomfort or fatigue during the sessions. However, the presence of trained instructors will ensure that any risks are minimized, and your well-being is prioritized.

8. Contact Information: If you have any questions or concerns about the study or your participation, please feel free to contact us. We are available to provide additional information and address any queries you may have.

Thank you for considering participating in our study. Your contribution is invaluable, and we greatly appreciate your involvement in advancing wheelchair training methods.

Sincerely,

Debora Pereira Salgado

c. PRE-Experience Section

This section of the questionnaire gathers information about your background and baseline skills relevant to the wheelchair simulator study. Your responses will help us understand your experience with wheelchairs and any pre-existing cognitive or physical limitations that might influence your performance in the simulator.

d. DEMOGRAPHICS INFORMATION

Questions	Information
Gender	
Age	
Duration of wheelchair use (years)	
Wheelchair Type (e.g. Scooter or Power Wheelchair)	
Joystick type (standard or any adaptation)	
Dominant Hand	
Diagnosis	
Additional Information (optional) E.g. Play videogames	

e. WHEELCHAIR SKILLS TEST QUESTIONNAIRE (WST-Q) VERSION FOR POWERED WHEELCHAIRS.

This questionnaire can be downloaded from https://wheelchairskillsprogram.ca/en/skills-manual-forms/.

f. MONTREAL COGNITIVE ASSESSMENT (MOCA) VERSION 8.1 ENGLISH

This questionnaire and training can be download from https://mocacognition.com/

g. DURING EXPERIENCE SECTION

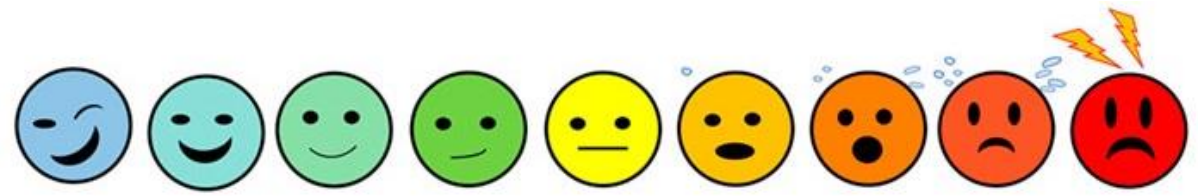
To gain a deeper understanding of your experience using the wheelchair simulator, we'll be asking you to provide real-time feedback on your mental effort and perceived difficulty after completing each task. This information will help us evaluate the simulator's cognitive demands and its ability to replicate real-world challenges.

h. PAAS SCALE

Mental Effort

Very, very low mental effort	Very low mental effort	Low mental effort	Rather low mental effort	Neither low nor high mental effort	Rather high mental effort	High mental effort	Very high mental effort	Very, very high mental effort
1	2	3	4	5	6	7	8	9
○	○	○	○	○	○	○	○	○

Perceived Difficulty



1	2	3	4	5	6	7	8	9
○	○	○	○	○	○	○	○	○

Task	PAAS Scale	
	Mental Effort	Perceived Difficulty
1		
2		
3		
4		
5		
6		
7		
8		
9	NA	NA
10		
11		
12		

i. POWER MOBILITY ROAD TEST ASSESSMENT SHEET

Element/Tasks	Score				Comments
	1	2	3	4	
Approaching furniture					
Start and stopping wheelchair at will					
Crossing doorways without hitting walls					
Turning around a 90 right hand corner					
Turning around a 90 left hand corner					
Driving straight forward					
Driving straight backwards					
Turning 180 degrees					
Starting and stopping wheelchair					
Turning right and left upon command					
Driving straight forward in a narrow corridor not hitting walls					
Manoeuvre between objects					

Score Definition:

4 – Completely Independent: optimal performance, able to perform task in one attempt smoothly and safely.

3 – Completes task hesitantly, requires several tries, requires speed restriction, and/or bumps wall, objects, etc., lightly (without causing harm).

2 – Bumps objects and people in a way that causes harm or could cause harm to driver, other persons or to objects.

1 – Unable to complete task: reason:_____. For example, may require verbal and/or visual cues or physical assistance.

j. POST-EXPERIENCE SECTION

Thank you for participating in the wheelchair simulator study. Your insights are invaluable in helping us understand how effectively the simulator reflects real-world wheelchair skills and provides useful assessments and training. To gain a better understanding of your experience with the wheelchair simulator, we invite you to answer a few brief questions. Your feedback will assist us in enhancing the simulator's user experience and overall effectiveness.

k. NASA-TLX (OVERALL COGNITIVE WORKLOAD)

The evaluation you are about to perform is a technique that has been developed by Nasa to assess the relative importance of six factors in determining how much workload you experienced while preforming a task that you recently completed. These six factors are defined below on this page. Read through them to make sure you understand what each factor means. If you have any questions, please ask your administrator.

Workload factors	Definition
Mental Demand Level (low/high)	**How much mental add perceptual activity was required** (for example, thinking, deciding, calculating, remembering, looking, searching, etc)? Was the task easy or demanding, simple or complex, forgiving or exacting?
Physical Demand Level (low/high)	**How much physical activity was required** (for example, pushing, pulling, turning, controlling, activating, etc.)? Was the task easy or demanding, slow or brisk, slack or strenuous, restful or laborious?
Temporal Demand Level (low/high)	**How much time pressure did you feel** due to the rate or pace at which the tasks or task elements occurred? Was the pace slow and leisurely or rapid and frantic?
Performance Level(good/poor)	How successful do you think you were in accomplish the goals of the task set by the experimenter (or yourself)? How satisfied were you with your performance in accomplish these goals?
Effort Level (low/high)	**How hard did you have to work** (mentally and physically) to accomplish your level of performance?
Frustration Level (low/high)	**How insecure, discouraged, irritated, stressed**, and annoyed versus secure, gratified, content, relaxed, and complacent did you feel during the task?

You will now be presented with a Series of rating scales. For each of the six scales, evaluate the task you recently performed by cross on the scale's location that matches your experience. Each line has two endpoint that describe the scale. Consider your responses carefully in distinguishing among the different task conditions and consider each individually.

Mental Demand (How mentally demanding was the task? How much mental and perceptual activity did you spend for this task?)

Very Low Very High

Physical Demand (How physically demanding was the task? How much physical activity did you spend for this task?)

Very Low Very High

Temporal Demand (How hurried or rushed was the pace of the task? How much time pressure did you feel in order to complete this task?)

Very Low Very High

Performance (How successful were you in accomplishing what you were asked to do? How successful do you think you were in accomplishing the goals of the task?)

Good Poor

Effort (How hard did you have to work to accomplish your level of performance?)

Very Low Very High

Frustration (How insecure, discouraged, irritated, stressed, and annoyed were you during this task?)

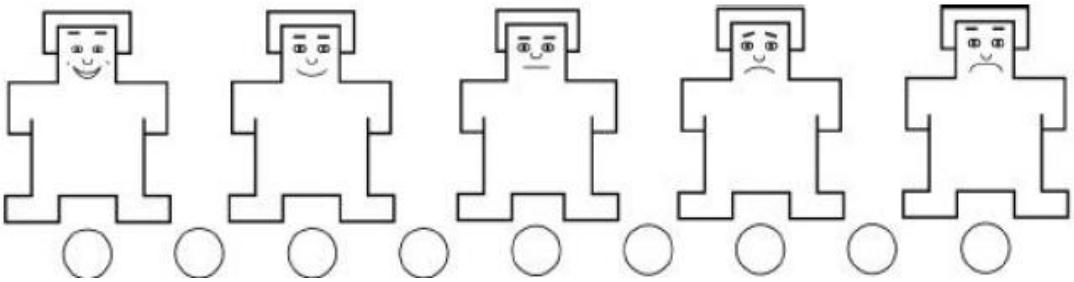
Very Low Very High

1. SELF-ASSESSMENT MANIKIN – SAM (EMOTION)

You will be asked to rate your emotions towards to experience in using the Simulator. It will be asked to rate on three separate scales.

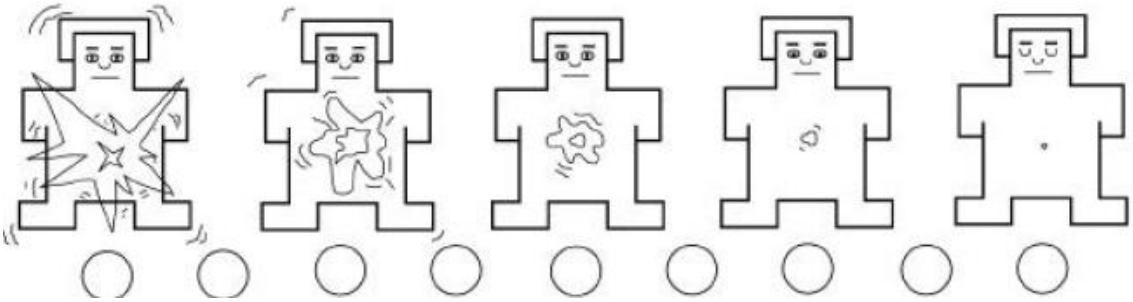
1. Valence (Pleasant level)

Pleasant Satisfied Neutral Unsatisfied Unpleasant



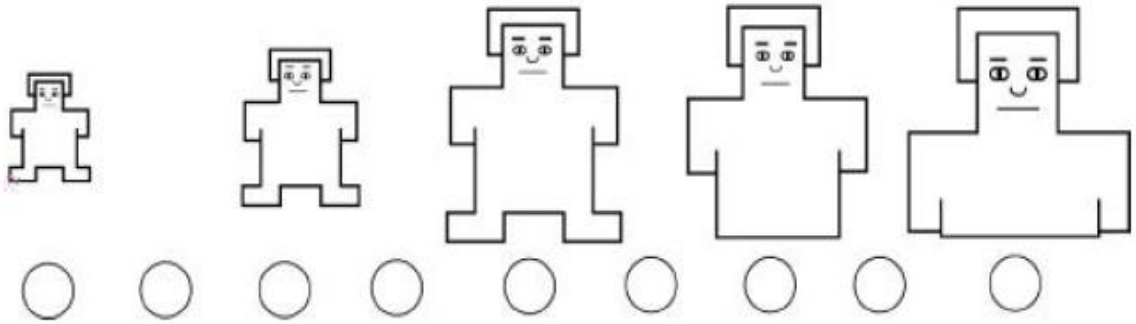
2. Arousal (Excitement level)

Excited Wide-awake Neutral Dull Calm



3. Dominance (Emotion Control level)

Dependent. Powerlessness Neutral Powerful Independent



m. USABILITY, IMMERSION AND ENGAGEMENT

You will see some statements about experiences. Please indicate, whether or not each statement applies to your experience. There are no right or wrong answers, only your opinions count. And please remember: Answer all these questions only referring to this one experience. Read through them to make sure you understand the statement. If you have any questions, please ask your administrator.

1. How well do you believe your performance in the simulator reflects your current power mobility skills (joystick control)?

|_____|_____|_____|_____|_____|

Not at all Moderate Extremely Well

2. Did the tasks seem too easy or too difficult for your abilities?

|_____|_____|_____|_____|_____|

Too easy Moderate Too difficult

3. Do you feel that your power mobility skills improved during the simulator session?

|_____|_____|_____|_____|_____|

No improvement Moderate Significant improvement

4. After using the simulator, do you feel more confident in handling a power wheelchair

|_____|_____|_____|_____|_____|

Not at all Moderate Very Confident

5. How aware were you of the real world surrounding while navigating in the virtual world? (i.e. sounds, room temperature, other people, etc.)?

|_____|_____|_____|_____|_____|

Extreme aware Moderate Not aware at all

6. I had a sense of acting in the virtual space, rather than operating something from outside.

|_____|_____|_____|_____|_____|

Strongly Disagree Neutral Agree Strongly
Disagree Agree

7. How much did your experience in the virtual environment seem consistent with your real-world experience?

|_____|_____|_____|_____|_____|

Not consistent at all Moderate Extremely consistent

8. In the computer generated world, I had a sense of "being there".

|_____|_____|_____|_____|_____|

Not at all Moderate Very much

9. I felt present in the virtual space.

|_____|_____|_____|_____|_____|

Strongly Disagree Neutral Agree Strongly
Disagree Agree

10. I would imagine that people would learn to use the wheelchair training simulator very quickly.

|_____|_____|_____|_____|_____|

Strongly Disagree Neutral Agree Strongly
Disagree Agree

11. I found the system very difficult to use.

|_____|_____|_____|_____|_____|

Strongly Disagree Neutral Agree Strongly
Disagree Agree

12. I needed to learn a lot of things before I could get going with the system.

|_____|_____|_____|_____|_____|

Strongly Disagree Neutral Agree Strongly
Disagree Agree

13. Was the simulator easy to understand and navigate?

|_____|_____|_____|_____|_____|

Too easy Moderate Too difficult

14. Did the interface feel user-friendly and intuitive?

|_____|_____|_____|_____|_____|

Not at all Moderate Extremely Well

15. Were the tasks and instructions clear and easy to follow?

|_____|_____|_____|_____|_____|

Too easy Moderate Too difficult

16. How satisfied are you with your experience using the simulator?

|_____|_____|_____|_____|_____|

Not at all Moderate Extremely Well

17. Would you recommend the simulator as a training tool for new power wheelchair users?

|_____|_____|_____|_____|_____|

Not at all Neutral Extremely Recommend

18. Would you recommend the simulator as assessment tool for new wheelchair users?

|_____|_____|_____|_____|_____|

Not at all Neutral Extremely Recommend

APPENDIX F. LAB-BASED NORMALITY TEST RESULTS

Table 7.3: Normality Test Results from Lab-based study metrics

		Shapiro-Wilk Normality Test Assumption Met		
Category	**Metric**	**Desktop**	**Headset-1**	**Headset-2**
Performance	Number of Joystick Commands		normal	normal
	Number of Collisions		normal	
	Task Completion Time (s)		normal	normal
Usability	Q1. SUS – Too complex			
	Q2. SUS – Easy to use			
	Q3. SUS - Easy to learn			
	Q4. SUS - Difficult to use			
	Q5. SUS - Needed to learn a lot			
Presence	Q6. IPQ - Sense of being there			
	Q7. IPQ - Realness			
	Q8. IPQ - Spatial presence			
	Q9. IPQ - Involved in VE		normal	normal
	Q10. IPQ - Immersion			normal
Usability	Total SUS Score	normal	normal	
Presence	Total IPQ Score	normal	normal	normal
Emotion	Valence (SAM)			normal
	Arousal (SAM)		normal	
	Dominance (SAM)			normal
Cybersickness	General Discomfort (SSQ)	N/A		normal
	Nausea (SSQ)	N/A		
	Oculomotor (SSQ)	N/A	normal	
	Disorientation (SSQ)	N/A	normal	
	Total SSQ Score	N/A	normal	
Cognitive Workload	Mental Demand	normal	normal	normal
	Mental × Weight			
	Weighted Mental Score	normal	normal	normal
	Physical Demand			
	Physical X Weight	normal		
	Weighted Physical Score			
	Temporal Demand	normal		
	Temporal X Weight		normal	normal
	Weighted Temporal Score			
	Performance Demand			

	Performance X Weight			normal
	Weighted Performance Score		normal	normal
	Effort Demand		normal	normal
	Effort X Weight			normal
	Weighted Effort Score			normal
	Frustration Demand			
	Frustration X Weight			
	Weighted Frustration Score			
	Raw TLX Score	normal	normal	normal
	Weighted TLX Score	normal	normal	normal
EDA	Mean EDA - Baseline			
	Mean EDA - Test			
	Mean EDA - 1st Collision			
	Mean EDA - Difference (Test - Baseline)			
	Mean EDA - Difference (1st Collision - Baseline)			
	Mean EDA % Change (Test vs. Baseline)		normal	
	Mean EDA % Change (1st Collision vs. Baseline)		normal	
HR	Mean HR - Baseline	normal	normal	normal
	Mean HR - Test	normal	normal	normal
	Mean HR - 1st Collision	normal		normal
	Mean HR - Difference (Test - Baseline)	normal		normal
	Mean HR - Difference (1st Collision - Baseline)	normal		normal
	Mean HR % Change (Test vs. Baseline)	normal		normal
	Mean HR % Change (1st Collision vs. Baseline)	normal l		normal
IBI	Mean IBI - Baseline	normal	normal	normal
	Mean IBI - Test	normal	normal	normal
	Mean IBI - Difference (Test - Baseline)			
	Mean IBI % Change (Test vs. Baseline)			
	Mean IBI - 1st Collision	normal	normal	normal

	Mean IBI - Difference (1st Collision - Baseline)		normal	normal
	Mean IBI % Change (1st Collision vs. Baseline)		normal	normal
	SDNN IBI Baseline			normal
	SDNN IBI Test	normal		normal
	SDNN IBI Difference (Test- Baseline)		normal	
	RMSSD IBI Baseline		normal	
	RMSSD IBI Test	normal		
	RMSSD IBI Difference (Test - Baseline)	normal		
Head Movements	Pitch Range (rad)		normal	normal
	Yaw Range (rad)	normal		
	Roll Range (rad)		normal	normal
	Mean Angular Velocity X (rad/s)			
	Mean Angular Velocity Y (rad/s)			
	Mean Angular Velocity Z (rad/s)			normal

APPENDIX G. STUDY 1 & 2 STATISTICS TABLES

This appendix presents the descriptive and inferential statistics from two lab-based studies:

Study 1: Between-group comparison based on display type, both using a high-jerk motion profile:

- Desktop (non-immersive) vs. Headset-1 (immersive)

Study 2: Two pairwise comparisons involving the Headset-2 condition (low-jerk motion profile):

- Headset-2 vs. Desktop
- Headset-2 vs. Headset-1

a. COGNITIVE WORKLOAD (NASA-TLX)

Table 7.4: Study 1 Mann–Whitney U test results for NASA-TLX subscale comparisons between desktop and headset-1 groups.

	Mean Rank		Test Statistics		
Metric	Desktop	Headset-1	U	Z	p-value
Mental Demand	22.15	19.38	176.500	-0.729	0.466
Mental × Weight	22.75	18.53	162.000	-1.148	0.251
Weighted Mental Score	22.38	19.06	171.000	-0.874	0.382
Physical Demand	20.25	22.06	186.000	-0.477	0.633
Physical X Weight	21.48	20.32	192.500	-0.311	0.755
Weighted Physical Score	21.46	20.35	193.000	-0.292	0.770
Temporal Demand	21.00	21.00	204.000	0.000	1.000
Temporal X Weight	20.52	21.68	192.500	-0.310	0.757
Weighted Temporal Score	19.79	22.71	175.000	-0.775	0.438
Performance Demand	**13.92**	**31.00**	**34.000**	**-4.511**	**0.000***
Performance X Weight	18.92	23.94	154.000	-1.380	0.168
Weighted Performance Score	**14.35**	**30.38**	**44.500**	**-4.224**	**0.000***
Effort Demand	20.58	21.59	194.000	-0.265	0.791
Effort X Weight	22.21	19.29	175.000	-0.809	0.418
Weighted Effort Score	21.31	20.56	196.500	-0.199	0.843
Frustration Demand	20.58	21.59	194.000	-0.268	0.789
Frustration X Weight	19.56	23.03	169.500	-0.979	0.328
Weighted Frustration Score	19.17	23.59	160.000	-1.238	0.216
*. The difference is significant at the 0.05 level.					

Table 7.5: Study 1 independent samples t-test results comparing overall raw and weighted NASA-TLX scores between Desktop and Headset-1 groups.

Metric	Desktop M (SD)	Headset-1 M (SD)	t(df)	p-value	Levene's F (p)	Mean Diff.	95% CI
Raw TLX Score	31.32 (12.13)	41.94 (12.72)	-2.71 (39)	0.01*	0.02 (.897)	-10.62	[-18.55, -2.68]
Weighted TLX Score	39.24 (16.88)	56.03 (13.39)	-3.41 (39)	0.002*	1.76 (.193)	-16.79	[-26.76, -6.83]
*. The mean difference is significant at the 0.05 level.							

Table 7.6: Study 2 Kruskal–Wallis Test Results for NASA-TLX Subscales Across Simulator Conditions

Metric	Group	Mean (SD)	Median [IQR]	H	p-value	Effect Size (η^2)	Post hoc
mental	Desktop	52.92 (25.62)	60.00 [30.00 – 70.00]	0.85	0.654	0.02	ns
	Headset-1	47.12 (27.47)	47.50 [27.50 – 68.75]	0.85	0.654	0.02	ns
	Headset-2	54.06 (27.67)	65.00 [30.00 – 75.00]	0.85	0.654	0.02	ns
mental weight	Desktop	3.54 (1.32)	4.00 [3.00 – 5.00]	3.63	0.163	0.03	ns
	Headset-1	3.12 (1.17)	3.00 [2.75 – 4.00]	3.63	0.163	0.03	ns
	Headset-2	3.81 (1.38)	4.00 [3.50 – 5.00]	3.63	0.163	0.03	ns
mental weighted	Desktop	203.12 (135.69)	195.00 [75.00 – 310.00]	2.13	0.345	0.00	ns
	Headset-1	160.76 (128.34)	135.00 [75.62 – 230.00]	2.13	0.345	0.00	ns
	Headset-2	231.72 (150.24)	250.00 [92.50 – 350.00]	2.13	0.345	0.00	ns
physical	Desktop	25.21 (22.58)	20.00 [10.00 – 30.00]	2.45	0.294	0.01	ns
	Headset-1	27.12 (21.27)	27.50 [11.25 – 32.50]	2.45	0.294	0.01	ns
	Headset-2	20.00 (22.51)	7.50 [5.00 – 27.50]	2.45	0.294	0.01	ns
physical weight	Desktop	1.92 (1.21)	2.00 [1.00 – 3.00]	1.68	0.431	0.01	ns
	Headset-1	1.82 (1.59)	2.00 [0.00 – 3.25]	1.68	0.431	0.01	ns
	Headset-2	1.44 (1.55)	1.00 [0.00 – 2.00]	1.68	0.431	0.01	ns
physical weighted	Desktop	63.12 (106.35)	25.00 [12.50 – 55.00]	2.54	0.281	0.01	ns
	Headset-1	67.88 (91.20)	30.00 [0.00 – 105.62]	2.54	0.281	0.01	ns
	Headset-2	55.78 (107.70)	6.25 [0.00 – 41.25]	2.54	0.281	0.01	ns
temporal	Desktop	24.38 (18.43)	25.00 [7.50 – 35.00]	0.21	0.902	0.03	ns
	Headset-1	30.65 (31.57)	12.50 [2.50 – 56.25]	0.21	0.902	0.03	ns
	Headset-2	21.88 (19.74)	17.50 [5.00 – 35.00]	0.21	0.902	0.03	ns
temporal weight	Desktop	2.08 (1.82)	2.00 [0.50 – 3.50]	0.13	0.939	0.03	ns
	Headset-1	2.18 (1.59)	2.00 [1.00 – 3.00]	0.13	0.939	0.03	ns
	Headset-2	1.94 (1.18)	2.00 [1.00 – 3.00]	0.13	0.939	0.03	ns
temporal weighted	Desktop	59.79 (73.21)	27.50 [0.00 – 90.00]	0.61	0.736	0.03	ns
	Headset-1	96.91 (127.84)	17.50 [6.25 – 161.88]	0.61	0.736	0.03	ns
	Headset-2	49.22 (57.68)	21.25 [6.25 – 90.00]	0.61	0.736	0.03	ns
performance	Desktop	28.33 (20.41)	22.50 [15.00 – 40.00]	28.29	0.000	0.49	ns
	Headset-1	79.71 (22.34)	82.50 [76.25 – 93.75]	28.29	0.000	0.49	ns
	Headset-2	77.19 (22.98)	82.50 [67.50 – 95.00]	28.29	0.000	0.49	*
performance weight	Desktop	3.54 (1.32)	4.00 [2.50 – 5.00]	2.68	0.262	0.01	ns
	Headset-1	4.06 (1.25)	5.00 [3.00 – 5.00]	2.68	0.262	0.01	ns
	Headset-2	3.56 (1.09)	4.00 [3.00 – 4.00]	2.68	0.262	0.01	ns
Performance	Desktop	101.88 (80.47)	77.50 [42.50 – 142.50]	25.52	0.000	0.44	ns

weighted	Headset-1	329.56 (144.17)	387.50 [235.62 – 437.50]	25.52	0.000	0.44	ns
	Headset-2	276.09 (111.25)	283.75 [228.75 – 343.75]	25.52	0.000	0.44	*
effort	Desktop	37.50 (25.88)	30.00 [17.50 – 55.00]	0.30	0.861	0.03	ns
	Headset-1	39.18 (26.92)	47.50 [17.50 – 60.00]	0.30	0.861	0.03	ns
	Headset-2	40.31 (21.83)	42.50 [22.50 – 57.50]	0.30	0.861	0.03	ns
effort weight	Desktop	2.83 (1.17)	2.50 [2.00 – 4.00]	1.39	0.499	0.01	ns
	Headset-1	2.53 (1.07)	2.00 [2.00 – 3.00]	1.39	0.499	0.01	ns
	Headset-2	2.94 (1.06)	3.00 [2.00 – 4.00]	1.39	0.499	0.01	ns
effort weighted	Desktop	118.54 (118.55)	85.00 [40.00 – 135.00]	0.68	0.712	0.02	ns
	Headset-1	112.32 (104.43)	95.00 [31.25 – 190.00]	0.68	0.712	0.02	ns
	Headset-2	116.09 (66.55)	115.00 [61.25 – 176.25]	0.68	0.712	0.02	ns
frustration	Desktop	19.58 (24.22)	10.00 [5.00 – 22.50]	0.67	0.715	0.02	ns
	Headset-1	27.85 (32.42)	12.50 [2.50 – 45.62]	0.67	0.715	0.02	ns
	Headset-2	23.75 (28.37)	5.00 [2.50 – 50.00]	0.67	0.715	0.02	ns
frustration weight	Desktop	0.96 (1.37)	0.00 [0.00 – 2.00]	1.17	0.557	0.02	ns
	Headset-1	1.29 (1.57)	1.00 [0.00 – 2.00]	1.17	0.557	0.02	ns
	Headset-2	1.31 (1.54)	1.00 [0.00 – 2.00]	1.17	0.557	0.02	ns
frustration weighted	Desktop	42.08 (89.30)	0.00 [0.00 – 27.50]	1.57	0.456	0.01	ns
	Headset-1	73.00 (147.51)	7.50 [0.00 – 70.00]	1.57	0.456	0.01	ns
	Headset-2	40.78 (80.08)	3.75 [0.00 – 48.75]	1.57	0.456	0.01	ns
raw TLX	Desktop	31.32 (12.13)	30.84 [23.33 – 40.83]	8.18	0.017	0.11	ns
	Headset-1	41.94 (12.72)	45.00 [36.46 – 50.94]	8.18	0.017	0.11	ns
	Headset-2	39.53 (12.12)	40.84 [32.08 – 45.41]	8.18	0.017	0.11	ns
weighted TLX	Desktop	39.24 (16.88)	37.50 [29.17 – 49.50]	11.78	0.003	0.18	ns
	Headset-1	56.03 (13.39)	55.50 [51.71 – 65.92]	11.78	0.003	0.18	ns
	Headset-2	51.31 (13.13)	54.83 [42.17 – 60.84]	11.78	0.003	0.18	ns

Table 7.7: Study 2 Pairwise Comparisons Using Mann–Whitney U Tests for NASA-TLX Subscales

		Headset-2	Compared Group	Test Statistics		
Metric	Group Comparison	Mean (SD)	Mean (SD)	U	Z	p-value
Mental Demand	Headset-2 vs. Desktop	54.06 (27.67)	52.912 (25.62)	175.00	-0.47	0.652
	Headset-2 vs. Headset-1		47.12 (27.47)	115.00	-0.76	0.465
Mental x Weight	Headset-2 vs. Desktop	3.81 (1.38)	3.54 (1.32)	164.00	-0.80	0.452
	Headset-2 vs. Headset-1		3.12 (1.17)	83.50	-1.96	0.058
Weighted Mental Score	Headset-2 vs. Desktop	231.72 (150.24)	203.13 (135.69)	164.00	-0.77	0.452
	Headset-2 vs. Headset-1		160.76 (128.34)	97.00	-1.41	0.168
Physical Demand	Headset-2 vs. Desktop	20.00 (22.51)	25.21 (22.58)	143.00	-1.36	0.183
	Headset-2 vs. Headset-1		27.12 (21.27)	99.50	-1.33	0.191
Physical x Weight	Headset-2 vs. Desktop	1.44 (1.55)	1.92 (1.21)	143.50	-1.38	0.183
	Headset-2 vs. Headset-1		1.82 (1.59)	116.50	-0.72	0.488
Weighted Physical Score	Headset-2 vs. Desktop	55.78 (107.70)	63.13 (106.35)	129.50	-1.73	0.084
	Headset-2 vs. Headset-1		67.88 (91.20)	113.00	-0.84	0.423
Temporal Demand	Headset-2 vs. Desktop	21.88 (19.74)	24.38 (18.43)	176.00	-0.44	0.672
	Headset-2 vs. Headset-1		30.65 (31.57)	126.50	-0.35	0.736
Temporal x Weight	Headset-2 vs. Desktop	1.94 (1.18)	2.08 (1.82)	190.50	-0.04	0.967
	Headset-2 vs. Headset-1		2.18 (1.59)	127.50	-0.31	0.763
Weighted Temporal Score	Headset-2 vs. Desktop	49.22 (57.68)	59.79 (73.21)	184.50	-0.21	0.838
	Headset-2 vs. Headset-1		96.91 (127.84)	122.00	-0.51	0.631
Performance Demand	Headset-2 vs. Desktop	77.19 (22.98)	28.33 (20.41)	34.00	-4.38	0.000*
	Headset-2 vs. Headset-1		79.71 (22.34)	127.50	-0.31	0.763
Performance x Weight	Headset-2 vs. Desktop	3.56 (1.09)	3.54 (1.32)	187.50	-0.13	0.902
	Headset-2 vs. Headset-1		1.82 (1.59)	96.00	-1.50	0.157
Weighted Performance	Headset-2 vs. Desktop	276.09 (111.25)	101.88 (80.47)	43.50	-4.10	0.000*
	Headset-2 vs. Headset-1		329.56 (144.17)	96.50	-1.42	0.157
Effort Demand	Headset-2 vs. Desktop	40.31 (21.83)	37.50 (25.88)	172.00	-0.55	0.594
	Headset-2 vs. Headset-1		39.18 (26.92)	129.50	-0.24	0.817
Effort x Weight	Headset-2 vs. Desktop	2.94 (1.06)	2.83 (1.17)	179.00	-0.37	0.733
	Headset-2 vs. Headset-1		2.53 (1.07)	104.50	-1.19	0.260
Weighted Effort Score	Headset-2 vs. Desktop	116.09 (66.55)	118.54 (118.55)	163.00	-0.80	0.436
	Headset-2 vs. Headset-1		112.32 (104.43)	119.00	-0.61	0.557
Frustration Demand	Headset-2 vs. Desktop	23.75 (28.37)	19.58 (24.22)	163.00	-0.81	0.436
	Headset-2 vs. Headset-1		27.85 (32.42)	120.00	-0.59	0.581
	Headset-2 vs. Desktop	1.31 (1.54)	0.96 (1.37)	164.00	-0.84	0.452

Frustration x Weight	Headset-2 vs. Headset-1		1.29 (1.57)	133.50	-0.09	0.929
Weighted Frustration Score	Headset-2 vs. Desktop	40.78 (80.08)	42.08 (89.30)	169.00	-0.69	0.539
	Headset-2 vs. Headset-1		73.00 (147.51)	122.50	-0.50	0.631
Raw TLX Score	Headset-2 vs. Desktop	39.53 (12.12)	31.32 (12.13)	122.50	-1.92	0.054
	Headset-2 vs. Headset-1		41.94 (12.72)	114.50	-0.78	0.444
Weighted TLX Score	Headset-2 vs. Desktop	51.31 (13.13)	39.24 (16.88)	107.00	-2.35	0.018*
	Headset-2 vs. Headset-1		56.03 (13.39)	113.00	-0.83	0.423
*. The difference is significant at the 0.05 level.						

Table 7.8: Study 2 One-Way ANOVA Results for Normally Distributed NASA-TLX Scores

	df	Mean Square	F	Sig.
Mental Demand	2.00	239.86	0.34	0.717
Weighted Mental Score	2.00	21188.16	1.12	0.335
Raw TLX Score	2.00	644.43	4.26	0.019l
Weighted TLX Score	2.00	1553.42	6.99	0.002*
*. The difference is significant at the 0.05 level.				

Table 7.9: Study 2 Bonferroni-Adjusted Post Hoc Comparisons for Significant ANOVA Effects

Dependent Variable			Mean Difference	Std. Error	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
Mental Demand	Headset-2	Desktop	1.15	8.64	1.000	-20.19	22.48
		Headset-1	6.94	9.32	1.000	-16.08	29.97
weighted mental score	Headset-2	Desktop	28.59	44.48	1.000	-81.30	138.49
		Headset-1	70.95	48.00	0.436	-47.65	189.56
Raw TLX Score	Headset-2	Desktop	8.21	3.97	0.130	-1.60	18.02
		Headset-1	-2.41	4.29	1.000	-13.00	8.18
Weighted TLX Score	Headset-2	Desktop	12.08	4.81	0.045*	0.18	23.97
		Headset-1	-4.72	5.19	1.000	-17.55	8.12
*. The mean difference is significant at the 0.05 level.							

b. PHYSIOLOGICAL RESPONSE (HR, HRV AND EDA)

Table 7.10: Study 1 Mann–Whitney U test results comparing HR, EDA, and HRV (IBI) metrics.

Metric	Group	Mean (SD)	Median [IQR]	U	Z	p-value	Effect size (r)
Mean HR - Baseline	Desktop	75.14 (8.99)	76.50 [68.92 – 80.90]	162.00	-0.92	0.371	0.14
	Headset-1	78.22 (9.79)	80.05 [68.57 – 85.09]				
Mean HR - Test	Desktop	74.43 (9.30)	74.81 [68.19 – 81.21]	105.00	-2.48	0.013*	0.39
	Headset-1	83.75 (13.41)	86.27 [73.62 – 91.18]				
Mean HR - 1st Collision	Desktop	76.84 (10.21)	77.79 [69.07 – 82.56]	111.00	-2.08	0.037*	0.33
	Headset-1	87.43 (16.76)	82.03 [79.53 – 91.60]				
Mean HR - Difference (Test - Baseline)	Desktop	-0.71 (5.36)	-1.47 [-3.52 – 3.17]	110.00	-2.34	0.019*	0.37
	Headset-1	5.53 (10.32)	3.50 [-1.36 – 7.72]				
Mean HR - Difference (1st Collision - Baseline)	Desktop	1.70 (7.11)	0.18 [-3.10 – 7.21]	129.00	-1.57	0.121	0.25
	Headset-1	8.14 (15.40)	3.56 [-1.62 – 12.80]				
Mean HR % Change (Test vs. Baseline)	Desktop	-0.75 (7.13)	-2.07 [-5.39 – 4.50]	108.00	-2.39	0.016*	0.37
	Headset-1	7.26 (13.19)	4.09 [-1.81 – 11.04]				
Mean HR % Change (1st Collision vs. Baseline)	Desktop	2.56 (10.11)	0.22 [-3.96 – 9.22]	126.00	-1.66	0.101	0.26
	Headset-1	10.62 (19.37)	4.55 [-2.02 – 16.84]				
Mean IBI - Baseline	Desktop	0.83 (0.12)	0.79 [0.75 – 0.90]	137.00	-1.60	0.113	0.25
	Headset-1	0.77 (0.11)	0.74 [0.68 – 0.84]				
Mean IBI - Test	Desktop	0.82 (0.11)	0.81 [0.76 – 0.87]	138.00	-1.12	0.271	0.18
	Headset-1	0.79 (0.10)	0.76 [0.71 – 0.84]				
Mean IBI - 1st Collision	Desktop	0.81 (0.12)	0.78 [0.74 – 0.90]	33.00	-1.40	0.177	0.22
	Headset-1	0.74 (0.08)	0.73 [0.68 – 0.80]				
Mean IBI - Difference (Test - Baseline)	Desktop	0.01 (0.02)	0.00 [-0.00 – 0.01]	126.00	-1.48	0.145	0.23
	Headset-1	0.01 (0.05)	-0.00 [-0.00 – 0.00]				
Mean IBI - Difference (1st Collision - Baseline)	Desktop	-0.06 (0.21)	-0.01 [-0.03 – 0.01]	69.00	-2.75	0.005*	0.43
	Headset-1	-0.50 (0.40)	-0.70 [-0.81 – -0.03]				
Mean IBI % Change (Test vs. Baseline)	Desktop	0.73 (2.84)	0.36 [-0.65 – 0.62]	129.00	-1.39	0.171	0.22
	Headset-1	1.30 (8.37)	-0.41 [-0.65 – 0.12]				
Mean IBI % Change (1st Collision vs. Baseline)	Desktop	-1.69 (9.14)	-0.90 [-3.82 – 1.16]	54.00	0.00	1.000	0.00
	Headset-1	-1.40 (4.35)	-0.18 [-6.59 – 1.10]				
SDNN IBI Baseline	Desktop	0.07 (0.03)	0.06 [0.06 – 0.10]	173.00	-0.31	0.767	0.05
	Headset-1	0.08 (0.05)	0.07 [0.05 – 0.09]				
SDNN IBI Test	Desktop	0.07 (0.03)	0.06 [0.05 – 0.08]	177.00	-0.20	0.855	0.03
	Headset-1	0.08 (0.03)	0.08 [0.06 – 0.09]				
	Desktop	-0.00 (0.03)	-0.00 [-0.02 – 0.01]	119.00	-1.86	0.065	

SDNN IBI Difference (Test - Baseline)	Headset-1	-0.00 (0.05)	0.00 [-0.02 – 0.02]				0.29
RMSSD IBI Baseline	Desktop	0.07 (0.03)	0.07 [0.05 – 0.08]	136.00	-1.37	0.177	0.21
	Headset-1	0.09 (0.09)	0.07 [0.05 – 0.08]				
RMSSD IBI Test	Desktop	0.08 (0.03)	0.07 [0.05 – 0.09]	155.00	-0.83	0.420	0.13
	Headset-1	0.08 (0.02)	0.09 [0.06 – 0.10]				
RMSSD IBI Difference (Test - Baseline)	Desktop	0.01 (0.03)	0.01 [-0.01 – 0.03]	178.00	-0.17	0.877	0.03
	Headset-1	-0.01 (0.08)	0.00 [-0.00 – 0.03]				
Mean EDA - Baseline	Desktop	0.65 (0.93)	0.27 [0.13 – 0.79]	167.00	-0.78	0.448	0.12
	Headset-1	0.45 (0.53)	0.17 [0.09 – 0.73]				
Mean EDA - Test	Desktop	0.85 (1.24)	0.29 [0.16 – 1.28]	186.00	-0.26	0.808	0.04
	Headset-1	0.90 (1.27)	0.49 [0.17 – 0.76]				
Mean EDA - 1st Collision	Desktop	0.70 (0.86)	0.30 [0.17 – 0.89]	174.00	-0.29	0.789	0.04
	Headset-1	0.92 (1.36)	0.47 [0.16 – 0.61]				
Mean EDA - Difference (Test - Baseline)	Desktop	0.20 (1.01)	0.03 [-0.01 – 0.08]	150.00	-1.24	0.221	0.19
	Headset-1	0.45 (0.97)	0.07 [-0.04 – 0.62]				
Mean EDA - Difference (1st Collision - Baseline)	Desktop	0.04 (0.57)	0.02 [-0.02 – 0.05]	133.00	-1.46	0.151	0.23
	Headset-1	0.45 (1.10)	0.06 [-0.02 – 0.46]				
Mean EDA % Change (Test vs. Baseline)	Desktop	53.68 (150.09)	25.24 [-1.71 – 48.95]	138.00	-1.57	0.120	0.25
	Headset-1	138.01 (172.04)	116.35 [-10.07 – 237.45]				
Mean EDA % Change (1st Collision vs. Baseline)	Desktop	31.44 (81.51)	17.85 [-6.49 – 32.61]	127.00	-1.63	0.107	0.25
	Headset-1	122.70 (167.14)	60.28 [-4.43 – 217.72]				

Table 7.11: Study 2 Kruskal-Wallis Results for Physiological Metrics

	Mean Ranks			Test Statistics		
Metric	Desktop	Headset-1	Headset-2	Kruskal-Wallis H	df	Asymo. Sig,
Mean HR - Baseline	24.65	29.59	32.88	2.507	2	0.285
Mean HR - Test	22.96	35.65	28.88	5.930	2	0.052
Mean HR - 1st Collision	22.98	33.50	24.12	4.978	2	0.083
Mean HR - Difference (Test - Baseline)	26.35	37.76	21.75	8.627	2	0.013*
Mean HR - Difference (1st Collision - Baseline)	26.00	33.13	19.23	6.074	2	0.048*
Mean HR % Change (Test vs. Baseline)	26.22	37.76	21.94	8.527	2	0.014*
Mean HR % Change (1st Collision vs. Baseline)	25.87	33.38	19.15	6.387	2	0.041*
Mean IBI - Baseline	31.70	23.65	29.06	2.407	2	0.300
Mean IBI - Test	29.59	24.19	27.94	1.110	2	0.574
Mean IBI - 1st Collision	18.56	12.17	18.80	2.094	2	0.351
Mean IBI - Difference (Test - Baseline)	31.23	22.94	26.94	2.601	2	0.272
Mean IBI - Difference (1st Collision - Baseline)	32.32	18.88	25.63	7.115	2	0.029*
Mean IBI % Change (Test vs. Baseline)	31.09	23.19	26.88	2.374	2	0.305
Mean IBI % Change (1st Collision vs. Baseline)	17.00	17.67	18.30	0.112	2	0.946
SDNN IBI Baseline	28.96	30.38	24.25	1.310	2	0.519
SDNN IBI Test	27.35	28.81	28.13	0.080	2	0.961
SDNN IBI Difference (Test- Baseline)	23.91	33.81	28.06	3.603	2	0.165
RMSSD IBI Baseline	24.87	32.56	27.94	2.176	2	0.337
RMSSD IBI Test	24.83	29.19	31.38	1.701	2	0.427
RMSSD IBI Difference (Test - Baseline)	27.22	28.25	28.88	0.107	2	0.948
Mean EDA - Baseline	27.22	23.47	35.69	4.866	2	0.088
Mean EDA - Test	24.17	25.06	38.38	8.241	2	0.016*
Mean EDA - 1st Collision	23.04	24.25	35.38	6.017	2	0.049*
Mean EDA - Difference (Test - Baseline)	21.65	27.94	38.94	10.627	2	0.005*
Mean EDA - Difference (1st Collision - Baseline)	21.17	28.00	34.08	6.247	2	0.044
Mean EDA % Change (Test vs. Baseline)	22.39	31.12	34.50	5.830	2	0.054
Mean EDA % Change (1st Collision vs. Baseline)	21.09	29.69	32.15	5.452	2	0.065

Table 7.12: Study 2 Mann–Whitney U test results comparing HR, EDA, and HRV (IBI) metrics

		Headset-2	Compared Group	Test Statistics		
	Group Comparison	Mean (SD)	Mean (SD)	U	Z	p-value
Mean HR - Baseline	Headset-2 vs. Desktop	82.21 (12.36)	75.14 (8.99)	129.00	-1.57	0.121
	Headset-2 vs. Headset-1		78.22 (9.79)	121.00	-0.54	0.606
Mean HR - Test	Headset-2 vs. Desktop	79.40 (11.95)	74.43 (9.30)	147.00	-1.06	0.301
	Headset-2 vs. Headset-1		83.75 (13.41)	105.00	-1.12	0.276
Mean HR - 1st Collision	Headset-2 vs. Desktop	77.61 (9.40)	76.84 (10.21)	141.50	-0.26	0.795
	Headset-2 vs. Headset-1		87.43 (16.76)	65.00	-1.71	0.092
Mean HR - Difference (Test - Baseline)	Headset-2 vs. Desktop	-2.81 (7.92)	-0.71 (5.36)	148.00	-1.03	0.315
	Headset-2 vs. Headset-1		5.53 (10.32)	64.00	-2.59	0.009*
Mean HR - Difference (1st Collision - Baseline)	Headset-2 vs. Desktop	-4.19 (9.52)	1.70 (7.11)	106.00	-1.43	0.159
	Headset-2 vs. Headset-1		8.14 (15.40)	53.00	-2.24	0.025*
Mean HR % Change (Test vs. Baseline)	Headset-2 vs. Desktop	-3.03 (8.65)	-0.75 (7.13)	149.00	-1.00	0.329
	Headset-2 vs. Headset-1		7.26 (13.19)	66.00	-2.52	0.011*
Mean HR % Change (1st Collision vs. Baseline)	Headset-2 vs. Desktop	-4.32 (10.59)	2.56 (10.11)	106.00	-1.43	0.159
	Headset-2 vs. Headset-1		10.62 (19.37)	52.00	-2.28	0.022*
Mean IBI - Baseline	Headset-2 vs. Desktop	0.80 (0.13)	0.83 (0.12)	169.00	-0.43	0.682
	Headset-2 vs. Headset-1		0.77 (0.11)	112.00	-0.86	0.402
Mean IBI - Test	Headset-2 vs. Desktop	0.81 (0.12)	0.82 (0.11)	168.00	-0.24	0.827
	Headset-2 vs. Headset-1		0.79 (0.10)	113.00	-0.57	0.590
Mean IBI - 1st Collision	Headset-2 vs. Desktop	0.82 (0.14)	0.81 (0.12)	88.00	-0.10	0.944
	Headset-2 vs. Headset-1		0.74 (0.08)	19.00	-1.19	0.263
Mean IBI - Difference (Test - Baseline)	Headset-2 vs. Desktop	0.00 (0.04)	0.01 (0.02)	144.00	-0.95	0.356
	Headset-2 vs. Headset-1		0.01 (0.05)	105.00	-0.87	0.402
Mean IBI - Difference (1st Collision - Baseline)	Headset-2 vs. Desktop	-0.29 (0.39)	-0.06 (0.21)	115.00	-1.23	0.230
	Headset-2 vs. Headset-1		-0.50 (0.40)	97.00	-1.17	0.254
Mean IBI % Change (Test vs. Baseline)	Headset-2 vs. Desktop	0.53 (5.27)	0.73 (2.84)	144.00	-0.95	0.356
	Headset-2 vs. Headset-1		1.30 (8.37)	106.00	-0.83	0.423
Mean IBI % Change (1st Collision vs. Baseline)	Headset-2 vs. Desktop	-0.73 (3.14)	-1.69 (9.14)	81.00	-0.43	0.689
	Headset-2 vs. Headset-1		-1.40 (4.35)	29.00	-0.11	0.958
SDNN IBI Baseline	Headset-2 vs. Desktop	0.06 (0.02)	0.07 (0.03)	151.00	-0.94	0.358
	Headset-2 vs. Headset-1		0.08 (0.05)	101.00	-1.02	0.323
SDNN IBI Test	Headset-2 vs. Desktop	0.07 (0.03)	0.07 (0.03)	176.00	-0.23	0.832
	Headset-2 vs. Headset-1		0.09 (0.09)	122.00	-0.23	0.838
SDNN IBI Difference (Test- Baseline)	Headset-2 vs. Desktop	0.08 (0.07)	0.07 (0.03)	155.00	-0.83	0.420
	Headset-2 vs. Headset-1		0.08 (0.03)	100.00	-1.06	0.305

RMSSD IBI Baseline	Headset-2 vs. Desktop	0.10 (0.11)	0.08 (0.03)	160.00	-0.69	0.507
	Headset-2 vs. Headset-1		0.08 (0.02)	103.00	-0.94	0.361
RMSSD IBI Test	Headset-2 vs. Desktop	0.02 (0.07)	0.00 (0.03)	140.00	-1.26	0.217
	Headset-2 vs. Headset-1		0.00 (0.05)	118.00	-0.38	0.724
RMSSD IBI Difference (Test - Baseline)	Headset-2 vs. Desktop	0.03 (0.10)	0.01 (0.03)	172.00	-0.34	0.746
	Headset-2 vs. Headset-1		-0.01 (0.08)	126.00	-0.08	0.956
Mean EDA - Baseline	Headset-2 vs. Desktop	2.30 (4.22)	0.65 (0.93)	126.00	-1.66	0.101
	Headset-2 vs. Headset-1		0.45 (0.53)	79.00	-2.05	0.041*
Mean EDA - Test	Headset-2 vs. Desktop	5.00 (6.94)	0.85 (1.24)	94.00	-2.57	0.009*
	Headset-2 vs. Headset-1		0.90 (1.27)	68.00	-2.45	0.014*
Mean EDA - 1st Collision	Headset-2 vs. Desktop	5.13 (7.75)	0.70 (0.86)	80.00	-2.29	0.022*
	Headset-2 vs. Headset-1		0.92 (1.36)	58.00	-2.02	0.045*
Mean EDA - Difference (Test - Baseline)	Headset-2 vs. Desktop	2.71 (4.75)	0.20 (1.01)	72.00	-3.20	0.001*
	Headset-2 vs. Headset-1		0.45 (0.97)	81.00	-1.98	0.049*
Mean EDA - Difference (1st Collision - Baseline)	Headset-2 vs. Desktop	2.49 (5.40)	0.04 (0.57)	78.00	-2.35	0.018*
	Headset-2 vs. Headset-1		0.45 (1.10)	77.00	-1.18	0.249
Mean EDA % Change (Test vs. Baseline)	Headset-2 vs. Desktop	238.14 (357.75)	53.68 (150.09)	101.00	-2.37	0.017*
	Headset-2 vs. Headset-1		138.01 (172.04)	123.00	-0.47	0.657
Mean EDA % Change (1st Collision vs. Baseline)	Headset-2 vs. Desktop	220.55 (363.59)	31.44 (81.51)	82.00	-2.22	0.026*
	Headset-2 vs. Headset-1		122.70 (167.14)	98.00	-0.26	0.812

APPENDIX H. FIELD-BASED STATISTICS REPORT

This appendix presents the descriptive and inferential statistics from the field-based study:

- Users (power wheelchair users) vs. Control (able-body participants)

a. HEART RATE

Table 7.13: Study 3 Group-wise comparisons of HR at task level

Task	Metric	Group	Mean(SD)	Median [IQR]	U	Z	p-value	Effect size (r)
	HR Baseline	User	91.99 (13.68)	87.59 [81.56 - 106.71]	35.00	-0.44	0.696	0.10
		Control	95.39 (8.06)	97.1 [95.00 - 98.65]				
1	HR Task	User	78.13 (8.88)	77.94 [75.31 - 84.78]	39.00	-0.09	0.965	0.02
		Control	80.77 (12.44)	78.1 [71.70 - 87.25]				
	HR Change	User	-13.86 (8.83)	-15.97 [-20.61 - -9.21]	20.00	-1.78	0.083	0.42
		Control	-14.61 (14.03)	-18.8 [-22.20 - -9.32]				
	HR SD	User	2.11 (3.53)	0.84 [0.52 - 1.75]	28.00	-1.07	0.315	0.25
		Control	1.21 (1.07)	1.24 [0.29 - 1.89]				
2	HR Task	User	87.49 (11.27)	84.83 [78.91 - 91.99]	31.00	-0.80	0.460	0.19
		Control	78.84 (10.73)	77.39 [70.58 - 81.17]				
	HR Change	User	-4.5 (11.54)	-0.31 [-8.00 - 3.09]	37.00	-0.27	0.829	0.06
		Control	-16.54 (11.89)	-18.42 [-23.58 - -11.39]				
	HR SD	**User**	**3.38 (2.73)**	**2.89 [1.25 - 5.23]**	**15.00**	**-2.22**	**0.027***	**0.52**
		Control	**2.24 (1.57)**	**1.92 [1.13 - 2.63]**				
3	**HR Task**	**User**	**94.67 (14.77)**	**94.83 [79.47 - 109.60]**	**14.00**	**-2.31**	**0.021***	**0.54**
		Control	**89.06 (14.67)**	**91.37 [76.87 - 92.94]**				
	HR Change	User	2.67 (9.95)	2.48 [-3.70 - 7.74]	23.00	-1.51	0.146	0.36
		Control	-6.33 (13.03)	-7.19 [-12.19 - -1.75]				
	HR SD	**User**	**0.54 (0.5)**	**0.31 [0.18 - 0.94]**	**9.00**	**-2.75**	**0.004***	**0.65**
		Control	**0.47 (0.47)**	**0.28 [0.09 - 0.71]**				
4	HR Task	User	90.03 (15.71)	85.26 [78.06 - 104.62]	26.00	-1.24	0.237	0.29
		Control	83.3 (13.04)	80.89 [75.13 - 88.41]				
	HR Change	**User**	**-1.96 (11.2)**	**-5.8 [-9.25 - 5.22]**	**14.00**	**-2.31**	**0.021***	**0.54**
		Control	**-12.09 (12.6)**	**-13.63 [-21.05 - -5.89]**				
	HR SD	User	1.62 (1.23)	1.25 [0.69 - 2.52]	34.00	-0.53	0.633	0.13
		Control	1.36 (1.28)	0.72 [0.44 - 2.33]				
5	HR Task	User	85.29 (20.48)	79.23 [68.97 - 102.67]	18.00	-1.95	0.055	0.46
		Control	79.32 (11.88)	75.94 [70.52 - 84.20]				

	HR Change	**User**	**-6.7 (15.51)**	**-9.66 [-16.13 - 1.15]**	**17.00**	**-2.04**	**0.043***	**0.48**
		Control	**-16.07 (12.27)**	**-18.59 [-27.88 - -6.83]**				
	HR SD	User	1.32 (0.94)	1.14 [0.57 - 2.01]	20.00	-1.78	0.083	0.42
		Control	0.8 (0.6)	0.73 [0.47 - 0.89]				
6	HR Task	User	92.93 (14.68)	90.57 [83.36 - 102.75]	25.00	-1.33	0.203	0.31
		Control	77.78 (10.62)	76.61 [72.65 - 83.04]				
	HR Change	**User**	**0.94 (11.37)**	**1.8 [-5.72 - 6.08]**	**10.00**	**-2.67**	**0.006**	**0.63**
		Control	**-17.61 (13.17)**	**-18.8 [-27.77 - -7.56]**				
	HR SD	**User**	**0.42 (0.32)**	**0.39 [0.12 - 0.66]**	**3.00**	**-3.29**	**0.000***	**0.77**
		Control	**0.36 (0.36)**	**0.19 [0.08 - 0.62]**				
7	**HR Task**	**User**	**95.01 (17.74)**	**91.69 [81.88 - 106.67]**	**11.00**	**-2.58**	**0.009***	**0.61**
		Control	**75.03 (9.07)**	**77.56 [64.50 - 81.16]**				
	HR Change	**User**	**3.02 (12.83)**	**1.67 [-5.48 - 14.72]**	**7.00**	**-2.93**	**0.002***	**0.69**
		Control	**-20.36 (8.93)**	**-20.33 [-22.62 - -12.74]**				
	HR SD	**User**	**1.03 (1.59)**	**0.58 [0.20 - 0.80]**	**16.00**	**-2.13**	**0.034***	**0.50**
		Control	**0.26 (0.23)**	**0.19 [0.08 - 0.39]**				
8	**HR Task**	**User**	**92.86 (18.65)**	**90.32 [78.84 - 111.00]**	**13.00**	**-2.40**	**0.016***	**0.57**
		Control	**80.68 (11.73)**	**78.01 [73.98 - 84.11]**				
	HR Change	User	0.86 (11.28)	4.42 [-9.61 - 6.89]	38.00	-0.18	0.897	0.04
		Control	-14.71 (18.8)	-18.66 [-23.98 - -13.52]				
	HR SD	User	2.46 (1.92)	2.15 [0.75 - 3.87]	30.00	-0.89	0.408	0.21
		Control	4.16 (4.2)	1.71 [1.01 - 6.43]				
10	HR Task	User	98.62 (17.67)	97.62 [81.28 - 114.32]	35.00	-0.44	0.696	0.10
		Control	77.23 (11.98)	75.01 [71.62 - 80.29]				
	HR Change	User	6.63 (15.38)	3.71 [-0.68 - 10.77]	31.00	-0.80	0.460	0.19
		Control	-18.15 (13.55)	-21.22 [-25.79 - -14.16]				
	HR SD	User	1.07 (1.06)	0.75 [0.62 - 1.06]	24.00	-1.42	0.173	0.34
		Control	0.69 (0.39)	0.58 [0.41 - 1.00]				
11	HR Task	User	91.3 (17.31)	87.97 [78.91 - 108.46]	35.00	-0.44	0.696	0.10
		Control	81.12 (10.64)	81.72 [71.79 - 86.71]				
	HR Change	User	-0.69 (10.83)	1.32 [-6.90 - 3.54]	19.00	-1.87	0.068	0.44
		Control	-14.27 (10.46)	-14.77 [-24.39 - -10.52]				
	HR SD	User	0.74 (0.96)	0.24 [0.17 - 1.38]	27.00	-1.16	0.274	0.27
		Control	0.55 (0.62)	0.3 [0.12 - 0.77]				
12	HR Task	User	88.42 (12.76)	86.72 [78.43 - 98.03]	31.00	-0.80	0.460	0.19
		Control	75.21 (8.47)	74.73 [70.58 - 79.09]				
	HR Change	User	-3.57 (14.37)	-2.73 [-9.30 - 6.92]	39.00	-0.09	0.965	0.02
		Control	-20.17 (7.72)	-20.35 [-24.05 - -16.79]				
	HR SD	User	3.42 (2.31)	3.19 [1.27 - 5.21]	39.00	-0.09	0.965	0.02

		Control	3.74 (3.38)	2.97 [1.41 - 4.60]				
*. The mean difference is significant at the 0.05 level.								

b. WRIST MOTION (ACCELERATION AND JERK)

Table 7.14: Study 3 Study 3 Group-wise comparisons of Jerk and Acceleration Magnitude at session level

Metric	Group	Mean (SD)	Median [IQR]	U	Z	p-value	Effect size (r)
accMag	User	62.73 (1.08)	62.04 [61.96 - 63.70]	25	-1.33	0.203	0.31
	Control	63.53 (1.06)	63.75 [62.41 - 64.53]				
jerk_rms	User	12.96 (6.14)	11.42 [8.69 - 14.72]	35	-0.44	0.696	0.1
	Control	16.83 (14.93)	13.18 [5.20 - 19.51]				
jerk_mean	User	0.02 (0.11)	-0.03 [-0.04 - 0.07]	36	-0.36	0.762	0.08
	Control	-0.01 (0.07)	-0.01 [-0.04 - 0.05]				
jerk_std	User	12.97 (6.14)	11.42 [8.70 - 14.72]	35	-0.44	0.696	0.1
	Control	16.85 (14.95)	13.19 [5.20 - 19.53]				

Table 7.15: Study 3 Group-wise comparisons of Jerk and Acceleration Magnitude at task level

Task	Metric	Group	Mean (SD)	Median [IQR]	U	Z	p-value	Effect size (r)
1.00	accMag	User	62.72 (1.15)	62.04 [61.92 - 63.75]	26.00	-1.24	0.237	0.29
		Control	63.48 (1.15)	63.66 [62.43 - 64.57]				
	jerk_rms	**User**	**7.23 (7.92)**	**4.51 [3.43 - 6.06]**	**16.00**	**-2.13**	**0.034***	**0.50**
		Control	**9.20 (16.25)**	**1.87 [1.09 - 2.43]**				
	jerk_mean	User	0.04 (0.26)	-0.03 [-0.07 - 0.02]	24.00	-1.42	0.173	0.34
		Control	0.01 (0.03)	0.01 [-0.01 - 0.03]				
	jerk_std	User	7.24 (7.92)	4.52 [3.44 - 6.06]	16.00	-2.13	0.034*	0.50
		Control	9.21 (16.26)	1.87 [1.09 - 2.43]				
2.00	accMag	User	62.79 (1.10)	62.16 [61.95 - 63.84]	28.00	-1.07	0.315	0.25
		Control	63.47 (1.05)	63.70 [62.38 - 64.44]				
	jerk_rms	User	19.28 (18.58)	10.57 [6.91 - 28.41]	39.00	-0.09	0.965	0.02
		Control	19.57 (17.10)	17.04 [6.57 - 27.80]				
	jerk_mean	User	-0.00 (0.01)	-0.00 [-0.01 - 0.00]	22.00	-1.60	0.122	0.38
		Control	0.07 (0.18)	0.00 [-0.00 - 0.06]				
	jerk_std	User	19.28 (18.58)	10.58 [6.91 - 28.41]	39.00	-0.09	0.965	0.02
		Control	19.57 (17.10)	17.04 [6.57 - 27.81]				
3.00	accMag	User	62.58 (1.11)	62.06 [61.81 - 63.36]	21.00	-1.69	0.101	0.40

		Control	63.50 (1.06)	63.62 [62.57 - 64.36]				
	jerk_rms	User	7.77 (6.78)	5.02 [2.83 - 12.49]	32.00	-0.71	0.515	0.17
		Control	16.39 (19.06)	8.68 [3.35 - 20.52]				
	jerk_mean	User	-0.13 (0.22)	-0.02 [-0.33 - 0.03]	30.00	-0.89	0.408	0.21
		Control	-0.04 (0.57)	-0.00 [-0.09 - 0.08]				
	jerk_std	User	7.78 (6.79)	5.03 [2.82 - 12.52]	32.00	-0.71	0.515	0.17
		Control	16.44 (19.12)	8.68 [3.36 - 20.55]				
4.00	accMag	User	62.86 (1.11)	62.19 [62.06 - 63.93]	28.00	-1.07	0.315	0.25
		Control	63.63 (1.16)	64.06 [62.42 - 64.50]				
	jerk_rms	User	17.97 (14.02)	11.93 [7.80 - 27.89]	29.00	-0.98	0.360	0.23
		Control	13.82 (12.45)	9.91 [2.50 - 26.06]				
	jerk_mean	User	0.04 (0.12)	-0.01 [-0.03 - 0.10]	37.00	-0.27	0.829	0.06
		Control	-0.00 (0.06)	-0.00 [-0.05 - 0.01]				
	jerk_std	User	17.98 (14.03)	11.94 [7.80 - 27.90]	29.00	-0.98	0.360	0.23
		Control	13.83 (12.46)	9.92 [2.50 - 26.08]				
5.00	accMag	User	62.72 (1.10)	62.00 [61.95 - 63.74]	26.00	-1.24	0.237	0.29
		Control	63.44 (1.14)	63.70 [62.50 - 64.51]				
	jerk_rms	User	10.42 (7.02)	8.54 [5.94 - 12.31]	35.00	-0.44	0.696	0.10
		Control	17.51 (21.08)	5.57 [3.34 - 27.89]				
	jerk_mean	User	-0.01 (0.25)	-0.01 [-0.12 - 0.01]	34.00	-0.53	0.633	0.13
		Control	0.00 (0.05)	-0.01 [-0.01 - 0.04]				
	jerk_std	User	10.43 (7.02)	8.55 [5.95 - 12.32]	35.00	-0.44	0.696	0.10
		Control	17.53 (21.09)	5.58 [3.35 - 27.92]				
6.00	accMag	User	62.78 (1.17)	62.06 [61.89 - 63.96]	31.00	-0.80	0.460	0.19
		Control	63.39 (1.10)	63.69 [62.25 - 64.27]				
	jerk_rms	User	8.14 (6.13)	5.74 [4.69 - 9.03]	28.00	-1.07	0.315	0.25
		Control	12.34 (20.68)	1.94 [0.60 - 15.97]				
	jerk_mean	User	0.17 (0.42)	0.05 [-0.04 - 0.40]	29.00	-0.98	0.360	0.23
		Control	0.02 (0.07)	0.01 [-0.01 - 0.05]				
	jerk_std	User	8.15 (6.15)	5.75 [4.70 - 9.04]	28.00	-1.07	0.315	0.25
		Control	12.37 (20.74)	1.94 [0.60 - 16.00]				
7.00	accMag	User	62.74 (1.09)	62.18 [61.91 - 63.61]	22.00	-1.60	0.122	0.38
		Control	63.62 (1.11)	63.81 [62.59 - 64.51]				
	jerk_rms	User	12.46 (9.96)	8.69 [3.94 - 22.21]	26.00	-1.24	0.237	0.29
		Control	14.87 (22.24)	2.55 [1.62 - 21.41]				
	jerk_mean	User	-0.02 (0.11)	-0.02 [-0.07 - 0.01]	40.00	0.00	1.000	0.00
		Control	-0.05 (0.16)	-0.00 [-0.04 - 0.01]				
	jerk_std	User	12.47 (9.96)	8.70 [3.95 - 22.22]	26.00	-1.24	0.237	0.29
		Control	14.90 (22.29)	2.55 [1.63 - 21.45]				

8.00	accMag	User	62.82 (0.99)	62.31 [62.07 - 63.68]	26.00	-1.24	0.237	0.29
		Control	63.48 (1.05)	63.62 [62.47 - 64.44]				
	jerk_rms	User	23.94 (11.25)	22.05 [14.72 - 31.17]	26.00	-1.24	0.237	0.29
		Control	16.78 (13.26)	14.22 [7.80 - 21.22]				
	jerk_mean	User	0.02 (0.05)	0.01 [-0.00 - 0.07]	25.00	-1.33	0.203	0.31
		Control	-0.01 (0.02)	0.00 [-0.01 - 0.00]				
	jerk_std	User	23.95 (11.25)	22.06 [14.72 - 31.18]	26.00	-1.24	0.237	0.29
		Control	16.78 (13.26)	14.23 [7.80 - 21.22]				
10.00	accMag	User	62.54 (1.11)	61.92 [61.88 - 63.44]	15.00	-1.95	0.055	0.46
		Control	63.50 (0.98)	63.74 [62.32 - 64.29]				
	jerk_rms	User	11.60 (6.75)	9.43 [7.27 - 16.11]	35.00	0.00	1.000	0.00
		Control	23.34 (28.97)	14.69 [2.64 - 31.99]				
	jerk_mean	User	-0.03 (0.05)	-0.02 [-0.08 - 0.00]	32.00	-0.29	0.813	0.07
		Control	-0.03 (0.11)	-0.02 [-0.04 - 0.01]				
	jerk_std	User	11.61 (6.75)	9.44 [7.28 - 16.12]	35.00	0.00	1.000	0.00
		Control	23.35 (29.00)	14.70 [2.65 - 32.01]				
11.00	accMag	User	62.57 (1.15)	61.96 [61.86 - 63.45]	14.00	-2.05	0.043	0.48
		Control	63.78 (1.32)	64.05 [62.50 - 64.49]				
	jerk_rms	User	7.64 (3.81)	8.17 [4.52 - 10.71]	28.00	-0.68	0.536	0.16
		Control	21.98 (40.88)	3.57 [1.88 - 21.72]				
	jerk_mean	User	0.10 (0.42)	-0.04 [-0.08 - -0.01]	33.00	-0.20	0.887	0.05
		Control	-0.05 (0.24)	-0.02 [-0.05 - -0.00]				
	jerk_std	User	7.64 (3.80)	8.18 [4.53 - 10.70]	28.00	-0.68	0.536	0.16
		Control	22.03 (41.00)	3.58 [1.89 - 21.77]				
12.00	accMag	User	62.70 (1.14)	61.99 [61.94 - 63.46]	18.00	-1.66	0.109	0.39
		Control	63.56 (1.05)	63.95 [62.30 - 64.37]				
	jerk_rms	User	17.14 (8.88)	15.77 [11.94 - 19.64]	33.00	-0.20	0.887	0.05
		Control	19.34 (14.81)	18.10 [7.55 - 22.43]				
	jerk_mean	User	-0.00 (0.03)	0.00 [-0.01 - 0.01]	30.00	-0.49	0.669	0.12
		Control	0.00 (0.01)	0.00 [-0.00 - 0.01]				
	jerk_std	User	17.14 (8.88)	15.78 [11.94 - 19.64]	33.00	-0.20	0.887	0.05
		Control	19.34 (14.82)	18.10 [7.56 - 22.43]				

*. The mean difference is significant at the 0.05 level.