University of Otago

PhD Dissertation

Interaction and Emotional Response in Immersive Virtual Reality Learning Environments

---

Jonny M. Collins

Dunedin, New Zealand, February 2019

Thesis supervisors

Prof. Dr. Holger Regenbrecht

Assoc. Prof. Dr. Tobias Langlotz

Abstract

Virtual Reality (VR) is seen as a promising tool for effective education. The flexibility, controllability, and interactive capabilities of VR allow for a range of immersive experiences. This presents an opportunity for educators and researchers to produce engaging tools with the potential to deliver a range of topical content to learners. Several issues become apparent while studying such applications, rooted in the complexities of learning theory. Different perspectives on how education should be practised, coupled with the difficult nature of evaluating learning outcomes, contribute to a complex problem. This makes it difficult to produce a consistent framework that encapsulates pedagogical and Human-Computer Interaction principles for a coherent integration with immersive VR. We contribute to the existing body of research working to address this issue through an investigation of interaction, expertise, and emotional response in immersive VR learning environments.

We revitalize a historic research proposal that introduced the idea of using interactive virtual environments to present learners with novel conceptual problem spaces. Our first contribution reports on an investigation of the relationship between prior expertise and interactive experiences in immersive VR. We uncover difficulties associated with evaluating learning outcomes in VR systems, and the impact of interaction on users with varying expertise. In our second contribution, we refine our investigation to focus on insight learning, or "Aha! moments", their relationship with the users' sense of presence in virtual environments, and the necessary element of engagement. We employ physiological sensors as part of a systematic approach to measurement and analysis of users' emotional responses to immersive VR environments. The final contribution of this work reports on a real-world exploratory study in which an immersive VR learning environment is presented to inmates at a correctional facility. We are able to verify the implementation of our previously established method, and can demonstrate the real-world efficacy of immersive VR learning environments.

This work is of interest to the Human-Computer Interaction and education-technology communities. Implications are particularly relevant for interaction and user experience design in VR environments.

Acknowledgments

Many people have lent their help and support to me throughout my PhD without which would have made the completion of such a task an impossibility.

I am immensely thankful for my supervisors Prof. Holger Regenbrecht, and Assoc. Prof. Tobias Langlotz for their consistent support and guidance as world class supervisors. Their patience, understanding, and trust in me have lead me to this monumental point which would not have been possible without them.

The Otago Human-Computer Interaction lab has been a great source of help and feedback throughout my PhD, and I thank all members across the years for their input. I would like to thank Dr. Steven Mills, Dr. Florian Beyer, and Prof. Joerg Frauendiener for their advice on the mathematical components of my implementation work. Chris Edwards for providing much needed hardware devices. Oliver and Chontira for modelling in figures for my thesis and providing pivotal feedback. Russell Butson for the advisory role he played in the later stages of my work and also to Yekta and Cem for their contributions. To all those that participated in my studies. To the Methodist Mission Southern and Animated Research Limited for their facilitation of parts of this work.

I am grateful to be a part of the Information Science department here at the University of Otago. As a whole it is a supportive environment with many helpful academics. I would particularly like to thank our administration, Heather Cooper and Gail Mercer, for their incredibly consistent and unwavering support over my many years here.

I have received an abundance of support from my family and friends. Thank you so much to all of my family for their many forms of support. To my Father and Jillian, my Mother and Kerry, my Sister, Quintin, and family, and my new family Youmna, Salim, Romana, Bushra, and Masroor, I am infinitely grateful. Chris, Charles, Abdul, Mohammed, Sharlene, Joerg, Matthew, Claudia, Stefanie, Jacob, Jonathan, Stu, Rosa, Vadym, Saurabh, Shehan, and Noel are just some of the great friends that helped and supported me through the ups and downs of my PhD. Particular thanks to Chris Heinrich and Charles Walker for their friendships. I wish Chris the best as he completes his PhD.

Last but not least, a special thanks to my wife, Youmna. I thank you for your love, support, and patience throughout the final stages of my PhD. Your encouragement has gone a long way to keep me driven. I can not express my gratitude for all you do for me.

Contents

1 Introduction **1**

1.1 Motivation 3

1.2 A Historic Research Proposal 4

1.3 Summary of Contributions 6

1.4 Thesis Overview 8

2 Related Work **11**

2.1 Virtual Reality 12

2.2 Immersion and Presence 13

2.2.1 Defining Terminology 14

2.2.2 Spatial Presence 15

2.2.3 Measuring Presence 15

2.3 Education and Learning 17

2.3.1 A Brief History 18

2.3.2 Returning to Arnold 19

2.4 Learning in Virtual Reality 20

2.4.1 Learning in Augmented Reality 21

2.4.2 Virtual Reality for Training 22

2.4.3 Learning on Desktop Virtual Reality 23

2.4.4 Learning in Immersive Virtual Reality 24

2.4.5 Constructivism in Immersive Virtual Reality 25

3 Concept and System **29**

3.1 Arnold's Subject Matter 30

3.2 The Fourth Spatial Dimension 30

3.3 Implementation 33

3.3.1 Interaction Component 33

3.3.2 Operative Component 36

3.3.3 Visualisation Component 38

4 Validating Historic Measures **41**
4.1 Introduction . . . 42
4.2 Experimental Design . . . 42
4.3 Measures . . . 43
4.4 Procedure . . . 44
4.5 Results . . . 46
4.6 Discussion . . . 49

5 Knowledge and Experience: A Thought Experiment Brought to Life **53**
5.1 Introduction . . . 54
5.2 Experimental Design . . . 55
5.3 Apparatus Modifications . . . 55
5.4 Measures . . . 57
5.5 Procedure . . . 58
5.6 Results . . . 58
5.6.1 Expertise . . . 58
5.6.2 Entry Test . . . 59
5.6.3 Hypercube Assessments . . . 59
5.6.4 Ghostcube Matching Task . . . 61
5.6.5 Simulator Sickness Questionnaire . . . 63
5.7 Discussion . . . 64
5.8 Conclusion . . . 65

6 The Search for Insight: Presence and Aha! moments **69**
6.1 Introduction . . . 70
6.2 Emotional Response in VR . . . 72
6.2.1 Sensed Dimension: Emotional Response . . . 72
6.2.2 Reported Dimension: Presence and Aha! moments . . . 73
6.2.3 Observed Dimension: Environment, Performance . . . 74
6.3 Chasing the Eureka! - Usability Study . . . 75
6.3.1 Experimental Design . . . 76
6.3.2 Task and Measurements . . . 76
6.3.3 Procedure . . . 77
6.3.4 Results . . . 77
6.3.5 Discussion . . . 79
6.4 Chasing the Eureka! - Presence and Insight Learning . . . 79
6.4.1 Experimental Design . . . 80
6.4.2 Apparatus and Measurements . . . 81
6.4.3 Procedure . . . 85
6.5 Summary . . . 86

7 Presence, Insight, and Emotional Response: Results 89
7.1 Introduction . . . 90
7.2 Psychological Dimension . . . 91
7.2.1 Presence . . . 91
7.2.2 Aha! Moments . . . 93
7.2.3 Interviews . . . 93
7.2.4 Simulator Sickness Questionnaire (SSQ) . . . 94
7.3 Observational Dimension . . . 95
7.3.1 Ghostcube Solutions . . . 96
7.3.2 Ghostcube Changes . . . 98
7.4 Physiological Dimension . . . 99
7.4.1 Data Preprocessing and Operations . . . 99
7.4.2 Detecting Emotional Response . . . 102
7.5 Interpretations and Discussion . . . 104
7.5.1 The Aha! Hypotheses . . . 106
7.5.2 Emotional Response in Virtual Environments . . . 107
7.5.3 Implications and Future Work . . . 109
7.6 Conclusion . . . 111

8 Immersive VR Education: A Real World Exploratory Study 113
8.1 Introduction . . . 114
8.2 Prototypical Development . . . 116
8.2.1 Environment . . . 116
8.2.2 Virtual Content Integration . . . 117
8.2.3 Navigation and Interaction . . . 118
8.2.4 Additional Virtual Content . . . 119
8.2.5 Educational Content . . . 120
8.3 Commercial Development - Study Apparatus . . . 121
8.3.1 Virtual Environment and Content . . . 122
8.3.2 Educational Tasks . . . 123
8.4 Exploratory Study . . . 126
8.4.1 Study Design . . . 126
8.4.2 Apparatus and Measures . . . 127
8.4.3 Procedure . . . 128
8.5 Discussion . . . 129

9 Discussion & Conclusion 135
9.1 Contributions . . . 135
9.1.1 Interaction and Expertise in Immersive VR . . . 136
9.1.2 Presence, Insight, and Emotional Response in Immersive VR . . . 137
9.1.3 Educational Immersive VR in the Real World . . . 138

9.2 Implications and Generalisability 139
9.3 Limitations 141
9.4 Future Work 142
9.5 Conclusion 144

Bibliography **146**

A Igroup Presence Questionnaire Items **167**

B Simulator Sickness Questionnaire Items **171**

C All Data Dimension Plots - Study Four **173**

D Electrodermal Activity Data - Study Four **199**

List of Figures

1.1 Arnold's original interactive stereoscopic 3D setup for interactions with the fourth spatial dimension 5

3.1 Arnold's original setup followed by our replicated and modernised implementations 31
3.2 Demonstrates the steps through from 0D to 4D. 32
3.3 The principal components and workflow of our implementation 33
3.4 A visualised concept of the hypercube system demonstrates the interactive feedback loop 34
3.5 The two different devices used for interaction with 4D space 35
3.6 A hypercube rendered in two different ways 39

4.1 Possible and impossible hypercubes used for assessing comprehension 44
4.2 Study one procedure outline 45
4.3 Self-report knowledge questionnaire results 47
4.4 Prediction power results for the paper-based hypercube assessment 48
4.5 Overall scores of the paper-based hypercube assessment 49

5.1 The procedure outline for the second study investigating interaction and expertise in VR 56
5.2 The entry test scores between the expert and layperson groups organised by question category 59
5.3 Positive and negative prediction power results from the revised hypercube assessment 60
5.4 Hypercube assessment scores for each group between the first and second exposure 61
5.5 The number of ghostcubes solved for each group between the first and second exposure 62
5.6 Ghostcube solution frequencies per ghostcube task and by proportion of each group 62
5.7 Ghostcube task completion times for the first three target ghostcubes 63

6.1 The stereoscopic 3D and immersive usability study conditions 77
6.2 System Usability Scale (SUS) questionnaire means 78
6.4 Empatica E4 wristbands used for collecting physiological data 84
6.5 Study procedure for the main experiment of the second investigation 86

7.1 Presence results between the two conditions 91
7.2 Presence results analysed by spatial, involvement, and realism factors 92
7.3 An example of how Aha! moments are visualised as events 94
7.4 Ghostcube difficulty assignments based on their rotations. 97
7.5 Solution frequencies for each ghostcube categorised by condition and difficulty 97
7.6 Visualisation of solution and change events 98
7.7 Exemplar plot visualisation of EDA, temperature, accelerometer, and event data 100
7.8 Resulting classifications of physiological data against event-based data . . . 103

8.1 The first Virtual Mechanic conceptualisation as a prototype 115
8.2 Stereoscopic 360 degree panoramic images for the surrounding immersive environment 117
8.3 Virtual content integrated with the 360 degree environment 118
8.4 Displays are used to guide the user in the environment where gaze based interactions are used with an XBOX controller 119
8.5 A virtual brake assembly can be separated into its main components 119
8.6 Passive tasks are available delivering literacy education within the context of car parts 120
8.7 An active task is available where users select letters to make rhyming words 121
8.8 The virtual environment from the commercially developed product 122
8.9 A teleportation metaphor is used to move through the virtual environment . 123
8.10 Users must read tasks from a virtual whiteboard and identify correct workshop hazards 124
8.11 A numeracy task presents barrels which require different amounts of fluid to fill 125
8.12 In a second numeracy task, users must perform simple arithmetic problems to complete certain jobs 125

Chapter 1

Introduction

Contents

1.1 Motivation **3**

1.2 A Historic Research Proposal **4**

1.3 Summary of Contributions **6**

1.4 Thesis Overview **8**

Can *Virtual Reality* (VR) be an effective medium for education delivery? From the outside looking in, the answer appears obvious—of course! When we begin to ask how and why, we realise that the questions are more complex than expected.

VR technology has developed to a stage that it is capable of delivering convincing experiences. Due to the advances of the technology, VR has become a commonly known term in consumer circles. It is not the first time VR has become popular throughout the consumer market as in the 1990s VR went through a similar hype cycle. Since then, the research and development community focused on these technologies have had another two decades to develop hardware, research its application, and streamline its deployment to the stage where VR systems are found in homes. The VR research community have applied systems across a multitude of domains, some of the most popular of which are health [2, 64, 81, 144], training and instruction [4, 67, 186], military [16, 111, 147, 160], automotive manufacturing [105, 119, 181], collaboration [51, 110, 140], and the oil and gas [41, 159] and aerospace industries [155, 180]. It is only natural that the education domain would identify the same potential and adopt VR for its own purposes.

Since the inception of the personal computer, education researchers and developers together with technologists have been building and working in the computer-supported learning (CSL) field. It began with regular desktop applications such as those for supporting conceptual learning, mostly for scientific concepts [13, 92, 156, 191]. These have largely demonstrated effectiveness, especially when computer-supported virtual environments are used in conjunction with conventional real environments [212]. Over time the advancement of technology produced more computing power and therefore allowed for more complex software such as those running 3-Dimensional (3D) environments. These were referred to as Virtual Environments (VE), and the term "virtual reality" was also frequently used to refer to such systems. Desktop VEs demonstrated further potential on top of existing CSL solutions by allowing, for instance, real-time interactive 3D visualisation [129, 177]. The need to distinguish concepts arose once these new technologies became more prevalent, so a definition of VR needed developing [74, 179]. VR became known as a 3D computer-generated, inclusive, and interactive environment.

More recently, reviews of desktop VR-based education have found the technology to be effective in terms of learning outcomes [126]. If computer-supported learning was, and still is, an effective medium for education, and the next generation of CSL in the form of desktop virtual environments is also effective, it only follows that the next generation on will be at least as effective again. As technology has developed, the next step from desktop VEs has become fully immersive VEs, i.e. Virtual Reality Environments (VRE) that a user is "immersed" within by means of a technical medium such as a head-mounted display (HMD). This is the technology that has developed in the past five years to a point that consumers are adopting it for entertainment, and researchers are realising many previously conceived applications of VREs. Purely educational VR systems are yet to be produced on a large scale which is partly due to the lack of research on the efficacy of immersive technology in education. However, companies are beginning to generate smaller scale "education-based" systems such as Altspace VR*, Anatomyou†, and Google Earth‡. Many researchers have identified, as they did with prior CSL advancements, the potential of fully immersive technologies for learning, and are working towards more robust evaluation. The issue lies in the design and development of immersive VR learning (IVRL) systems, and the pedagogical considerations, or lack thereof, required for their successful development.

*https://altvr.com/
†http://anatomyou.com/en/
‡https://www.google.com/earth/explore/products/

1.1 Motivation

It is clear that VR will be adopted in education. What is not clear is how we can develop VREs to maximise its potential for the delivery of education. A further unknown is how generalisable VR development is for different subject matters within the pedagogical domain. Research within the computer and information sciences is working to validate the efficacy of immersive VR systems for use in the education sector. As discussed earlier, computer-supported learning has been shown to be effective in delivering education in multiple topic areas. However, a large part of that work is evaluating desktop VR applications and only recently have the research and commercial sectors committed to creating immersive VR learning solutions.

We lack guidance on how immersive learning environments should be developed to maximise the potential of VR. We also lack concrete results in terms of improvements to learning outcomes of learners in immersive systems. The difficult nature of longitudinal studies is one issue holding educational VR research back, although this is not related only to VR research, but educational research in general. A further issue is that not all topical content is catered for in a single VR application. This increases the effort to establish immersive VR as an effective tool for education in the broader sense. A fundamental challenge with this research field stems from the domain of application itself—education and general pedagogical practice is still a topic of study in its own domain with varying schools of thought on how it should be practised. While these issues are not the focus of this thesis in their totality, the result of this work can provide insight contributing to their investigation in the community.

Many of the preliminary results of IVRL evaluations rest upon learners' subjective reports after use of a system emphasising the system's potential as an engaging medium, but are they able to effectively teach? If they are capable of facilitating effective learning, what factors of VREs are most important?

Part of the motivation behind this thesis is theoretical and is focused on the psychology of learning and general pedagogy. There is a constant shift in the "generally accepted" philosophy of learning and while consensus forms over time across a number of schools of thought, domain experts within pedagogy and psychology have yet to settle on a theory of how learning really works. One could argue that for us to generate effective VR systems for learning, it is a requirement to better understand the concept of learning. The goal of this thesis is not to create a new model of learning, although a potential consequence of our work is that we gain insight into learning practices, at least in the scope of IVRE.

Pedagogy in Immersive Virtual Reality Recent research has identified a lack of pedagogical consideration in the research and development of education-based VR systems [58, 91]. While this is true, some previous work on educational VR has identified and outlined particular attributes of VR that are coherent with certain pedagogies. The most common pedagogies to arise from this discussion are *experiential learning* and *constructivism*. It is apparent that fully immersive VREs are able to provide an experience to users that will facilitate learning outcomes, not just in an information recall sense, rather in a way that knowledge gain is internalised. This idea is one of the key aspects of the philosophy of constructivism—learning is the result of an individual's interactions with the world, and knowledge is constructed based on those experiences. Immersive VREs have clear potential for application based on constructivist theories due to their inherent interactive affordances.

One of the first, if not the first, to conceive of using immersive VREs for learning is P. Arnold who was a member of the Biological Computing Laboratory at the University of Illinois. In 1971 Arnold proposed to use a 3D virtual environment to investigate not only learning, but the constructivist foundations of learning, i.e. what are the differences between acting and observing in a learning context [7].

It is this original proposal together with his follow-up proposal [8], and revisited work from Heinz von Foerster [194] that provide the foundation and starting point of this thesis work.

1.2 A Historic Research Proposal

P. Arnold is a proponent of constructivist philosophy which is a learner centred, experience-based philosophy of learning. To support constructivist concepts, Arnold proposed to evaluate the differences between actors and observers in the context of learning. In his proposal, he allowed a learner (actor) to interact with abstract mathematical concepts through manipulations and then view the resulting manipulations on a 3D stereoscopic display. He concurrently placed a second learner (observer) beside the actor where they would only view the actions and resulting manipulations. Arnold hypothesised that only the actor, through interaction and experience, would be able to achieve what he referred to as a "deep non-verbal comprehension", leading to a more correct internal representation of the subject matter. Figure 1.1 demonstrates Arnold's original hand drawn experimental setup for the actor and drawings of stereo pairs for two different 4D shapes.

To be able to robustly evaluate his hypothesis, he selected a subject matter which

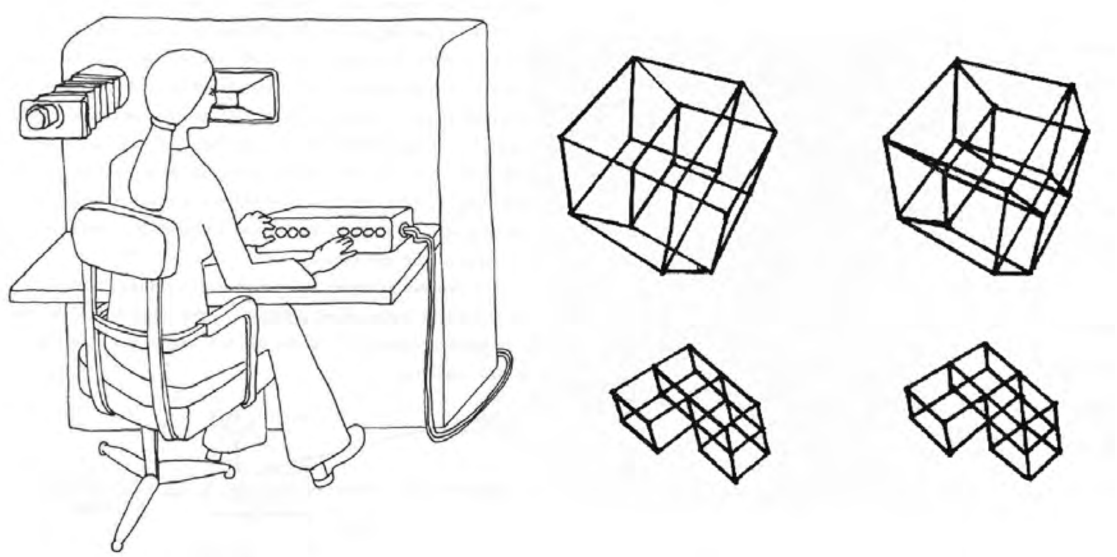


Figure 1.1: The original setup proposed by Arnold has a user sitting at a station with dials for manipulating a 4-Dimensional cube and a stereoscopic viewing platform for visualisation of the resulting user manipulations (left). Arnold also drew stereo pairs of example 4D shapes to demonstrate what a learner would see (right) [7, 8]. Figure permissions granted by the University of Illinois.

has not been experienced by anybody before—the fourth spatial dimension. That is four spatial dimensions that are perpendicular to each other, just as in our 3D space, we have three dimensions perpendicular to each other (x,y, and z). Although it is arguably impossible to imagine four-dimensional (4D) space and constructs within, it is possible to mathematically represent it, and visualise projections of 4D constructs. Arnold proposed to programmatically define 4D shapes such as a 4D cube (hypercube), and mathematically project the shape from 4D to 3D for visualisation. Interaction would be achieved through six analogue dials, each dial contributing to the overall rotation of the 4D shape. The learner (actor) would turn the dials and the rotation would be applied directly to the 4D shape in real-time, the resulting manipulations of which could be visualised in 3D.

Two decades after Arnold's original work [7, 8], Heinz von Foerster, located in the same laboratory, revisited this theoretical proposal adding his own interpretations of what such an experiment would yield [194]. He posited that after enough time using the system, a learner would experience a moment of insight in which they would exclaim "AHA!" signifying their achievement of what Arnold called a deep non-verbal comprehension. Like Arnold, v. Foerster also proposed that only an actor in such a scenario would experience this moment of insight.

This historic study proposal from Arnold together with the revisited discussion by v. Foerster is the foundation and starting point for the work in this thesis. We investigate interactive learning in immersive VR learning environments in two parts, both parts of which stem from Arnold's and v. Foerster's work. Interaction is the focal point of the first branch of investigation in which we evaluate the unique learning opportunities that interaction and experience provide in IVRL environments. The second branch of investigation focuses firstly on moments of insight, i.e. when learners experience, or perceive to experience, sudden and significant learning gains, and secondly what measures can we utilise for assessing those moments in virtual environments. Through these investigations, we bring to life this theoretical discourse which is grounded in the foundations of learning, and we evaluate immersive VR as the primary tool for doing so.

While we utilise many aspects of the originally described work of Arnold and von Foerster, we do not replicate the work in its totality. The focus for this thesis is on interaction and experience in IVRL environments. The fundamental idea of their work is highly relevant and applicable to our own questions which are inspired not only by their work, but by the current state of immersive learning research. Our questions are on the specific role of interaction and how IVRL systems can be developed to maximise the interactive potential of VR. Where Arnold proposed a comparison of actors and observers, we rather see those two roles as unique components of learning which provide their own perspectives and this work aims to focus at first on the interactive component of VREs. Therefore, for a majority of this work, we utilise the most relevant components of Arnold and von Foersters work—the subject matter, the proposed apparatus, and the proposed measurements. The remainder of this chapter outlines the main contributions of this work, followed by an overview of the thesis.

1.3 Summary of Contributions

The opportunity has arrived for immersive VR to be effectively applied in learning and education. The current state of the research space presents the need for more guidance in the research and development of IVRL systems. We as a community appear to be unsure of exactly which questions we should be asking. This is likely due partly to uncertainty in the realm of the pedagogy domain as alluded to earlier, although what we need is a more robust dichotomy of the constituents of learning in IVRL environments. This is the space we contribute to in this thesis.

Between the prior work of Arnold and v. Foerster that partly inspires this thesis, and the current state of the research space, we have drawn two overarching research questions:

RQ1: Does interactive involvement in IVRL environments provide unique learning experiences?

RQ2: How can we (1) lead learners towards, (2) measure, and (3) analyse, interactive learning experiences in IVRL environments?

We address these two research questions through three main contributions.

Contribution One. The first branch of investigation that stems from Arnold's work is the focal point of the first contribution, and addresses primarily our first research question. Through two experiments, we validate the historic elements of Arnold's work and then evaluate interaction in IVRL environments with particular focus on the relationship between interaction and expertise. We are able to draw on the results to provide insight into the role of interaction, not just for application on our chosen subject matter, but with implications for general interaction in immersive learning environments. From this work, we are able to raise additional questions and provide considerations for the various communities working on this form of CSL, namely human-computer interaction, VR, and education technology.

Contribution Two. Two further experiments address our second research question and form our second contribution. This is also the second branch of investigation stemming from Arnold's and von Foerster's work. In this body of work, we investigate Aha! moments (insight), presence, and the measurement and analysis of learning experiences in IVRL environments. Through this work we demonstrate the relationship between users' sense of presence and apparent learning gains. Furthermore, we employ physiological sensors in the measurement and structured analysis of emotional engagement as a part of the learning process. These results have implications not only for insight learning in virtual environments, but our methodology for measuring emotional responses has potential application for the broader VR community, both research and development.

Contribution Three. In a real-world scenario based in a correctional facility, local community education providers deliver context-based literacy and numeracy education to inmates. They identified the potential of immersive VR as a delivery tool, and we developed an initial prototype for this purpose which lead to a commercial development project. In the later stages of this thesis work, we had the opportunity to utilise the commercial prototype to run a semi-structured exploratory style study at the correctional facility with real users. We were able to apply our methodology for measuring emotional responses in virtual environments to inmates, and we provide a discussion on the implications of this exploratory study.

1.4 Thesis Overview

Below is an outline of the thesis structure with a brief summary of what the reader can expect from each section of the work:

Chapter Two: Related Work. This section provides a general overview of the relevant literature and demonstrates part of the motivation behind this thesis. VR concepts are discussed followed by an explanation of the defining concept of presence and its relationship with immersion. A brief theoretical background of education and learning theory is given before presenting prior research with respect to VR for learning.

Chapter Three: Concept and System. We describe the fundamental concepts behind the subject matter used throughout the majority of our work and the motivations behind its selection. We give a thorough description of our principal research apparatus which is implemented through two separate visual and interactive mediums resulting in two systems implementing the same content. These systems are used as the main research apparatuses throughout the remainder of the thesis.

Chapter Four: Validating Historic Measures. We describe the first part of our investigation of interaction and expertise in IVRL environments. This covers our first study ($n = 22$) intended to validate several components we extracted from the original proposal of Arnold, namely the subject matter, research apparatus, and measurements.

Chapter Five: A Thought Experiment Brought to Life. We report on the second part of the investigation of interaction and expertise. A second experiment ($n = 70$) is conducted incorporating the results and lessons learned from the prior validation study. The experiment primarily evaluates the unique value of interaction in IVRL environments, and further demonstrates the value of interaction for different types of learners.

Chapter Six: The Search for Insight: Presence and Aha! moments. The second branch of investigation begins in this section and includes two studies. The first study evaluates interface usability (n = 36). The second (primary) study (n = 24) is designed to investigate the complex relationships between the users' sense of presence, insight learning (Aha! moments), and the physiological measure of emotional response. Engagement is a key factor shared as a dependency between our measured phenomena which we hope to maximise through our structured measurement and analyses methods.

Chapter Seven: Presence, Insight, and Emotional Response. We present the results from the main study of the previous chapter investigating presence, insights and emotional response in IVRL environments. In addition to traditional methods, we applied a structured analysis which we developed based on three separate data dimensions where physiological measures of emotional response are at the heart. We are able to measure and analyse the complex relationships emerging from users' experiences in our immersive interactive learning environment. We discuss the implications of our results on users' experiences in, and the development of, immersive VR learning environments.

Chapter Eight: An Exemplar VR Learning Environment. This chapter reports on the development of a real-world project designed to deliver literacy and numeracy education to high-needs learners. It involves several development phases, and concludes with an exploratory study where we demonstrate our previously established method in practice. We run the study (n = 15) in a correctional facility and report on preliminary results. The results are exploratory in nature, but demonstrate the potential value of IVRL environments and how our method for measuring and analysing emotional responses in VEs can benefit such developments.

Chapter Nine: Discussion and Conclusion. We summarise the contributions and achievements made in this thesis. A critical discussion is presented summarising our findings and exploring the resulting implications for the research and commercial sectors alike. We discuss the underlying limitations of the work, and how aspects of the investigations can be improved upon. Finally, we outline the many opportunities for future work. A concise summary is given as a reflection of the work we have done, contributions made, and implications for us as a research community focusing on applied VR.

Chapter 2

Related Work

Contents

2.1 Virtual Reality **12**
2.2 Immersion and Presence **13**
2.3 Education and Learning **17**
2.4 Learning in Virtual Reality **20**

This chapter presents a background of the foundational concepts of this thesis. The first topic discussed is VR for which a defining characterisation is given. We also discuss presence as a defining characteristic of VR, immersion, and their respective measures. We will then address the concept of learning with a brief discussion of its history, and the relevant philosophies for this work. Finally, literature pertaining to immersive VR learning (IVRL) environments and interaction is discussed, therefore identifying the space in which we work. Parts of this chapter are included in the following work:

1. Collins, J., Regenbrecht, H., and Langlotz, T. (2019). Interaction and expertise in immersive virtual reality learning environments. *In IFIP Conference on Human-Computer Interaction (Submitted).* [38]

2. Collins, J., Regenbrecht, H., Langlotz, T., Can, Y. S., Ersoy, C., and Butson, R. (2019). Measuring Emotional Response in Virtual Environments: A Methodology Exemplified in a Learning Context. *In International Symposium for Mixed and Augmented Reality '19 (Accepted).* [39]

3. Collins, J., Regenbrecht, H., Langlotz, T. (2018). Back to the future: Constructivist learning in virtual reality. *In: Adjunct Proceedings of the IEEE International*

Symposium for Mixed and Augmented Reality 2018 (Poster). [37]

2.1 Virtual Reality

During the 1990s, VR was popularised generating a lot of interest within the research, commercial, and consumer circles, resulting in fictional conceptualisations of VR applications and systems [24]. Despite the surge of VR, it was lacking a consistent definition, though there are several characteristics that can be found in arguably all functional VR systems, and they are: 1) computer generated, 2) three-dimensional (3D), 3) real-time interaction, and 4) the users' sense of presence.

The first two components constitute the computer generated 3D environment and provides the foundation or base of the Virtual Reality system [179]. The real-time interaction component includes the two aspects real-time, and interaction. Interaction refers to all forms of interaction a user could have with an environment. Minimum required interactions would usually refer to visual interaction, i.e. one's ability to look around and have the display update respective of where they look [146]. These are considered as mediums which extend the users sensorimotor systems. Other forms of interaction with the environment are also possible such as positional movement or more direct "physical" interactions with virtual objects, though these are not always essential for a functional VR system [19].

The concept of "real-time" changes based on technological availability and application demands but usually refers to the "event to system response" time, i.e. when a user acts within a system, it should respond quickly enough that a user can operate. This involved both update rates, and latency. Modern VR systems are recommended to run at a 90Hz refresh rate. Acceptable latency measures vary depending on different users across different applications [141]. Latencies ranging from 20ms to 200ms have been proposed as acceptable, but very much depends on the user's task.

The final component of VR is Presence which is, in the context of VR, known as a subjective psychological phenomenon and has an entire body of research dedicated to it. It has been stated that presence could be 'the' defining component of virtual reality [146, 179]. Presence is most commonly defined as "a sense of being in a virtual environment" [175]. It is arguable that if a user does not achieve presence, the goals of the VR system will not be achieved due to a lack of the user's psychological investment in the system. A more in-depth discussion of presence is given later in this section. While consumer circles have only recently invested in VR technology, the research community have been investigating potential uses of VR for decades exploring its application in multiple domain spaces.

A multitude of application spaces have been targeted by the VR research community [57, 66, 71, 131, 143]. For instance, researchers have been able to successfully apply VR technology in the health sector with applications on phobia treatment [5, 128, 146], physical rehabilitation [79, 184], and psychological conditions [2, 14, 150] to name a few. VR has been further applied in the automotive industry [105, 181, 186], military [16, 130, 147, 160], collaboration studies [51, 110, 140], and training and instruction [4, 67, 186]. Evaluation of these systems has yielded tangible evidence of the technology's impact in that it is able to match or improve on current treatment techniques. Advantages afforded by VR include its flexibility and, as of more recently, affordability. Therefore VR may be adopted over some traditional approaches which can be more difficult, expensive, or unsafe to conduct in the real world [148, 154]. The examples given above are indicative of the applicability of VR to a diverse range of domains. Education is also an obvious target for applied VR research, but before we discuss literature concerned with IVRL environments, we will discuss the defining concept of presence, immersion, and how they are measured.

2.2 Immersion and Presence

Presence is referred to in many disciplines however the defining characteristics will often vary depending on the context. Within the field of Virtual Reality alone, there are multiple schools of thought on the phenomenon [118, 165, 175, 205]. Additionally, the term immersion has been used synonymously with the term presence [146] which has lead to an effort within the VR community to differentiate the two terms according to various characteristics [24, 174, 205]. We will proceed to discuss the various schools of thought on presence and immersion and summarise with the definitions we subscribe to throughout the remainder of this thesis.

Before addressing the issue of the synonymous use of presence and immersion as terms, it is worth noting that there are multiple specialised presence constructs. Well known presence specialities include co-presence [171], social presence [133], and spatial presence [146, 161]. This thesis work is not concerned with co-presence or social presence; instead our focus is set on the concept of spatial presence. When one uses the term 'presence' alone in the virtual reality field, they most often refer to spatial presence [118]. Throughout the literature there are also variations in terminology for presence such as 'Telepresence' [75, 165], 'physical presence' [108], or generally the 'Sense of Presence in Virtual Reality' (or virtual environments). For the remainder of this thesis, when we discuss 'presence', we are referring to spatial presence.

2.2.1 Defining Terminology

Presence is best defined in conjunction with immersion as there is a relationship between the two [4, 24], however their meanings require distinguishing from each other. One of the major schools of thought on presence and immersion posited by Witmer et al. states presence, in the context of virtual environments, as "experiencing the computer-generated environment rather than the actual physical locale" and confirms presence to be a subjective experience [205]. Witmer et al. proceeds to describe immersion as "a psychological state characterised by perceiving oneself to be enveloped by, included in, and interacting with an environment that provides a continuous stream of stimuli and experiences." Furthermore, the concept of involvement is added as a necessity for one to achieve a sense of presence, and this is defined as "a psychological state experienced as a consequence of focusing one's energy and attention on a coherent set of stimuli." After establishing these three concepts, Witmer et al. states that for one to achieve presence, one must first have immersion plus involvement. It is worth noting here that all three of these defined phenomena are subjective experiences which is a point we will return to later.

Another of the major schools of thought on presence and immersion is put forth by Slater et al. which states immersion as the extent to which a system provides the "illusion" of another reality [175]. The extent is characterised by four factors: 1) inclusive - the extent to which reality is shut out, 2) extensive - the range of sensory modalities which are accommodated, 3) surrounding - how panoramic the system is, and 4) vivid - fidelity of the medium per modality (i.e. monitor resolution for vision) [176]. It is clear then, given this definition, that this understanding of immersion can be defined objectively. Head-mounted displays (HMD), namely modern ones, are generally inclusive devices as they block out the real world from view. Assuming the HMD includes headphones, we can say that two sensory modalities are catered to (vision and auditory assuming headphones or speakers). Most HMD-based systems are fully tracked and therefore fulfil the surrounding aspect of immersion. Finally, the vivid aspect in this example could be defined by the resolution and field-of-view of the HMD monitors [26]. The quality of the audio component of the system (resolution, spatial audio) is a further example. Slater et al. go on to define presence as "the (psychological) sense of being in the virtual environment." This definition specifies a psychological phenomenon and is therefore subjective by nature.

These definitions aim to clearly distinguish between presence and immersion. From here throughout this thesis, our definitions and terminology are aligned with the latter—those presented by Slater et al., where immersion can be objectively defined, and presence is a subjective experience.

2.2.2 Spatial Presence

Having distinguished presence and immersion, we can more thoroughly explore the presence construct, specifically the aspect we are primarily concerned with for this work, spatial presence. Regenbrecht et al. reviewed existing definitions of presence, and while agreeing with them in essence, found them to be imprecise [146]. One such definition states presence as “the perceptual illusion of nonmediation” [116] and is said to occur “in two distinct ways: (a) The medium can appear to be invisible or transparent ... and (b) the medium can appear to be transformed into something other than a medium, a social entity”. Many definitions of presence consider it to be a “distinct feature of a user’s spatial cognition” [116, 165, 179]. One characteristic identified through many discussions of presence is the spatial element—the concept of being “transported” or physically being somewhere else. This is the essence of what spatial presence is and is most commonly expressed as the sense of “being there” [118, 146].

2.2.3 Measuring Presence

Presence is a subjective phenomenon and is therefore inherently difficult to measure. Both objective and subjective approaches have been proposed [88, 118, 146, 165]

Objective Measures. Two common forms of objective measurement are behavioural and physiological measures [162]. Researchers propose that, in a mediated environment (VE), a user will respond to stimuli as they would in the real world so long as the environment allows the user to feel present [59], although, as Welch points out [197], there is little solid evidence behind this belief. Held et al. present an example of behavioural measures in which they measure by observing reflex movements as reactions to fast-moving virtual objects [75]. Freeman et al. also utilised behavioural measures in a study where the visual output was manipulated in order to induce a postural change in the user [59]. Hodges et al. recorded subjects’ physical and verbal responses to their virtual flying system as indirect evidence of users’ sense of presence [78]. Subjects’ postural responses to visual stimuli were recorded in a study comparing monoscopic and stereoscopic viewing mediums to screen size where a presentation of a rally car race was the subject matter [84]. Malbos et al. successfully evaluated a Behaviour Presence Test they had developed by presenting participants with threatening situations [120]. It has been acknowledged by many presence researchers that while useful, these observational forms of measurement often require additional instrumentation of, or implementation features within, a VR application [146]. Slater states that "special features have to be introduced into an application that may be tenuously connected to the application itself" [170] which implies that these measures can not be applied ubiquitously.

Physiological measures are a further objective method for measuring presence. Heart rate, galvanic skin response, and skin temperature are among the most common of the physiological measures applied in VR research [124]. Prior research has established these measures to be highly correlated with the user's sense of presence in virtual environments [123]. Felnhofer et al. investigated the relationship between anxiety and presence in which they utilised the physiological measure of heart rate [55]. Further work from Felnhofer et al. attempted to find a relationship between electrodermal activity (EDA) (otherwise known as galvanic skin response) and presence, but only found EDA to be a poor indicator [56]. Heart rate and skin conductance were measured in a study evaluating presence and fear of spiders which also found no significant correlations [134].

Although physiological measures are being employed at an increasing rate as the technology improves and becomes more ubiquitous, subjective forms of presence measures remain to be the most common [85, 190].

Subjective Measures. The most common method for measuring the sense of presence is by self-report questionnaire [117]. As we discussed earlier there are different schools of thought on an exact definition of presence which changes the formation of a questionnaire to measure the phenomenon. Early questionnaire style measures were used by Slater et al. to evaluate various factor influences on the level of reported presence [173, 174]. Hendrix et al. used self-report measures of presence to evaluate the effect of spatialised auditory cues in VEs [76]. Pictoral realism and interactivity were also evaluated for its impact on the user sense of presence with positive effects found [198]. Welch et al. also used a post-experiment interview to support measures of presence. The work of Kim et al. applied a self-report measure of presence in a traditional media context after having participants view media on a television with inconclusive results [100].

A "presence questionnaire" was produced together with an "immersive tendencies questionnaire" (ITQ) by Witmer et al. according to their definition of presence and immersion [205]. The presence questionnaire measures a user's sense of presence in VEs, and the ITQ measures "differences in the tendencies of individuals to experience presence". The two were found to have reliable internal consistency. An 18-item self-report questionnaire was applied in the work of Barfield et al. to evaluate the effect of update rate and interactive medium on presence [11]. A factor analysis was conducted by Schubert et al. on the sense of presence which found three primary factors being spatial, involvement, and realism, where spatial and involvement yielding highest weights. This factor analysis resulted in the 14-item Igroup Presence Questionnaire (IPQ) [145, 161]. A cross-media presence questionnaire was developed by Lessiter et al. called the ITC-Sense of presence inventory (ITC-SOPI) which was developed based on presenting a range of media types [112]. Slater et al. presented an alternative method of a self-report presence measure where they asked users to report on transitions between real and virtual contexts as opposed to interrupting their current process to ask them about presence [172]. They found the measure to be significantly associated with body movement in the virtual environment where the users' task was to reach out and touch virtual chess pieces.

There has been push back against the proliferation of questionnaires measuring presence. In later work, Slater ran a study where he made up a psychological phenomenon called "the colorfulness of the experience" and found that through questionnaires, he could associate this made up attribute with task completion [170]. This demonstrated a degree of meaninglessness in these forms of analyses, showing a potential weakness in questionnaires as a valid form of measurement. This should be considered when evaluating the users' sense of presence in virtual environments.

Presence is arguably the defining element of VR [146, 179] and must be considered in the design and development of any immersive VR solution. The user sense of presence is in fact a necessity for the success of many VR systems [157]. As discussed earlier, the psychotherapeutic domain is a common space for applied VR research. For instance phobia [27, 64, 78, 149], post-traumatic stress disorder [50, 150], anxiety [137], and pain distraction [80, 81] treatments are common examples which demonstrate the effectiveness of VR, but also the importance of the users' sense of presence. If a user is not psychologically engaged, i.e. involved, in that virtual environment, those treatments will not be as effective [78, 146]. It is easy to imagine the same is true of VR solutions for educational purposes, i.e. if a learner is not engaged, the delivery of the education will be less effective.

2.3 Education and Learning

Before discussing the literature concerning VR and learning, a summary of education and learning will be provided. This will give a brief history of the evolution of educational practices therefore framing the current state of the philosophies and theories of learning practice.

2.3.1 A Brief History

Education practice has gone through multiple different implementations over the past century. We provide a summary of the major concepts of learning from a comprehensive overview provided by Hanna et al. [70].

Behaviourism was the most prominent learning theory of the early 20th century and was based on drill and repeat practices of a task. Learning was the process of changing a learners behaviour based on associations between environmental stimuli and observable responses of the individual. A further aspect of behaviourism is positive reinforcement and it was posited that if a teacher provides reinforcement such as “Good, job!”, that the learner’s “correct” behaviour would be solidified. Then by exercise and repetition, the behaviours become stronger, hence learning [169, 188].

In the mid 20th century, there was a shift from behaviourism to psychology, particularly cognitive psychology [65]. Learning was aligned with a knowledge-acquisition metaphor, rather than being seen as a series of “responses to stimuli”. This altered the preferred method for the delivery of education to focus on lecturing and reading textbooks. Cognitive psychology places teachers as knowledge providers and learners as knowledge receivers who process information and store it [168].

Constructivism emerged and became popular in the mid-late 1900s and remains one of the most prominent concepts implemented upon. In contrast to the cognitive psychological take on learning, it sees the learning process as knowledge construction rather than knowledge acquisition. Knowledge is built as a result of the learners’ interactions with world stimuli and previously built knowledge structures are reorganised and rebuilt as new interactions occur. Once again, the metaphor for learning changed from knowledge-acquisition to knowledge-construction [122]. There are immediate variations of constructivism such as radical constructivism and moderate constructivism which begin to incorporate ontological philosophies into pedagogy [195]. For instance, radical constructivists posit one’s knowledge construction as entirely idiosyncratic and does not necessarily reflect an objective world. Under moderate constructivism, learners’ various constructions eventually arrive at some representation of external reality.

The common theme among any constructivist theory is its learner-centred nature. Constructivism has served as a foundation for further, more recent approaches to learning [70], often adding context as a main theme (e.g. social, cultural, or environmental context), but remain grounded in the original learner-centred approach. Socio-constructivism is an example which claims social interaction as an especially important context influencing the knowledge construction process [196]. Knowledge is still considered as a consequence of activity and world stimuli together, and not something external to be acquired.

Constructivism is the guiding principle for this work where interpretations from P. Arnold [7], and later von Foerster [194] spark our investigation of immersive VR for constructivist approaches to learning.

2.3.2 Returning to Arnold

We discussed the research proposal of Arnold in section 1.2 in which his aim was to utilise VR environments to support constructivist practices. At the time of Arnold's proposal, constructivism was in the process of popularisation and much of the work of major constructivist thinkers such as Jean Piaget was still emerging [138]. After almost five decades, there have been many successful attempts to verify the concept of constructivism such as in the work of De Corte et al. where they found children applied techniques in solving mathematical problems that they had not been previously taught [45]. While an apparent consensus is forming on the reality of constructivism, there are still multiple schools of thought on how one should implement such theories and concepts in practice [70].

One of the reasons there are alternative interpretations is due to the complexity of constructivism in that it sits between pedagogy and epistemology [195]. It defines not only the learning process, but also the nature and form of knowledge itself. Such an abstract concept makes it difficult for one to implement in the practice of education. This has caused pedagogical thinkers to dissect constructivism, and try to extract pragmatic components that can be more easily integrated into the practice of education. Experiential learning is an example of a more pragmatic theory to emerge from constructivism [102]. It is defined as the process of learning through experience or "learning through reflection on doing" and is a derivative of constructivism [47, 102].

One of the complexities of experiential interpretations of constructivism is that experience comes in many forms. Arnold made a division within his proposal of acting and observing resulting in two different types of experience: (1) interactive experience, and (2) passive experience. In this thesis, we are primarily focused on the interactive form of experience and what it can provide for learning in the context of immersive VR environments. We bring constructivism into focus through our investigations of learning in virtual environments serving as a reminder of its importance in the discussion of learning and education in general.

Having summarised the fundamental philosophy underlying the motivation of our work, a summary of the literature in the space of educational-VR will be presented.

2.4 Learning in Virtual Reality

Researchers have considered VR technology for educational applications since the 1990s. The technology was of a low fidelity in the early stages of the research and as one would expect, many usability issues were raised. Despite the usability issues, VR proved to be an engaging tool for learners, and in some cases there were indications of learning gains [46, 202]. Youngblut provides a summary of the early efforts of applying VR in schools [210] reiterating many of the usability issues experienced by learners, but also initiating a comprehensive discussion on where and how VR might (and might not) be useful in education.

We also need to return to the issue of terminology. In subsection 2.2.1, we clarified our definitions of presence and immersion. When we discuss the immersiveness of a system, we are referring to the objective measure of immersion, i.e. the extent of the technical components comprising a VR system. Furthermore, when we refer to immersive VR systems or immersive VR learning (IVRL) environments, we explicitly refer to "fully immersive" systems such as those that fulfil Slater's components of immersion (inclusive, extensive, surrounding, and vivid) to a high degree. For instance, we consider HMD-based VR systems as immersive systems because they shut out the real world, they cater to at least the visual modality, and modern systems are fully surrounding (360 degree, 6 DOF), have high tracking precision, and high resolutions. A cave automatic virtual environment (CAVE) is a space surrounded with rear-projection screens for the walls and floor onto which projectors throw stereo images at 120Hz providing a surrounding 3D virtual environment [42]. CAVEs are also considered immersive due to their surrounding and inclusive environments [60]. We consider VR environments delivered on monitors or mobile devices as "desktop VR".

Once more, there are different schools of thought on what an immersive system really is. Therefore, the literature is full of references to "immersive environments" and "immersive VR" when they happen to fall into our category of desktop VR. Augmented Reality (AR) is a concept which lies between VR and reality combining the two. For instance, a mobile device using its camera to capture the real environment, and overlaying virtual content within that real environment is generating an AR environment. AR has a research field of its own, and there are references to immersive AR learning environments within that literature too [17, 98, 107, 206, 211]. For completeness, we also cover AR learning research with a brief overview.

The following related work will be divided into four sections beginning with research pertaining to education-based AR applications. Desktop and immersive VR learning systems will then be covered. Finally, we cover education-based VR research where the designers, developers, or researchers have used or considered pedagogical guidelines, namely constructivism.

2.4.1 Learning in Augmented Reality

AR technology is not a focus of our work though it is relevant and, should it prove effective, has potential to be present within classrooms and home environments in an educational sense [18]. Furthermore, AR shares the primary factor of virtuality in common with VR and it is likely there will also be shared approaches towards the design and implementation of immersive VR in educational environments. AR technologies still face many fundamental issues in their interactive capabilities, and in particular, their ability to provide a visually coherent experience [36]. A significant and active community are working to address these issues. Recent work has discussed the future of AR as a pervasive experience tailored to individuals [69] which demonstrates AR to have potential in education. There are many examples of education-based AR research and, like VR, explores application in various topics [17, 18, 23, 30, 206].

Construct3D is an application built for mathematics and geometry education at high school and university level [93, 94] in which learners wear a stereoscopic HMD and hold a two handed interaction device for manipulating 3D geometries. Studies were exploratory but showed promising results in terms of learners' engagement and encouragement. Eleftheria et al. combine gamification and AR to create an application aimed at teaching scientific concepts on a tablet [54] though the application was not formally evaluated in terms of learning outcomes. Kerawalla et al. implemented an AR system for teaching primary school science and included an evaluation but found students were less engaged when using the AR device [97]. An issue that was identified was the teachers' inability to interact and observe the actions that were taking place. A monitor-based AR system was implemented to concurrently train psychomotor and cognitive skill sets in the context of a medical treatment [103]. An informal evaluation with experts and medical students found the application to provide advantages for students over traditional approaches. Further examples of implementations are found in the areas of language [115], engineering [121], and for classroom management [167].

While many of these applications, and AR in general, show promise for education, they provide a different medium and a differing environment than immersive VR. The majority of these AR examples, with the exception of the work from Kaufmann et al. (who employed a see-through HMD) utilise tablets, monitors, or other AR devices such as optical see-through headsets. VR provides an entirely inclusive environment not constrained in its development by any aspect of the real-world, at least in terms of conception. Furthermore, it provides unique interactive capabilities within those environments. We continue below with literature concerning VR-based training systems.

2.4.2 Virtual Reality for Training

VR applications targeting the education domain have been studied at an increasing rate over the past two decades. The literature can be divided into two categories: 1) training systems, and 2) education systems. Training systems are those designed to train users on specific real world tasks. Within the health care domain VR has been used for retraining stroke patients' upper limb motor skills [144], operating room skill training [62, 164], disaster preparedness training [6], and for community health work [178]. One of the most prominent applications of VR technology is in military for field training [16, 130] and military medical education [52].

Further applications are found in safety- [207], navigation- [20], industry assembly task- [22, 67], and pre-flight- [182] training, to name a few. In many cases these works have been able to establish robust skill transfer to the real world [109]. The second category within the literature are those targeting education in the purer sense, rather than skill training. We distinguish between educational and training systems due to the differing pedagogical aims.

Training systems often have a specific desired outcome which can be measured using existing robust measurements (i.e. the completion time of assembly tasks is representative of whether the system is effective). In an educational context where for example a student is learning conceptual subject matter, learners are often assessed by other means such as subject matter examinations (tests) after exposure in a system [114, 166]. Consequently, there is a body of research investigating how VR can be applied in education and while consensus is beginning to form regarding the general efficacy of VR for education, exactly how it should be implemented or how effective it is has yet to be shown. This is reflected in the range of results produced by the research community [3, 77, 106]. As discussed earlier, there are two categories of educational VR implementations: 1) desktop VR, and 2) immersive VR.

2.4.3 Learning on Desktop Virtual Reality

Freina et al. present a recent summary of the current state of educational VR [60]. They identified the primary target for educational VR, particularly immersive VR, as adult training or higher level education applications. A recent meta analysis of learning outcomes in K-12 and higher education when applying (desktop-) VR-based instructions analysed 69 studies [126] showing overall that systems were effective in terms of gains in learning outcomes. They identify three different forms of assessment for determining learning outcomes: knowledge-based, ability-based, and skill-based, although the meta study found no significant differences between the three.

Mathematics' sub-topics such as spatial thinking and geometry are common focuses of educational VR researchers because of the visualisation and interaction benefits afforded by VR environments. Song et al. presented a VR modelling language system for teaching geometry at middle school level resulting in a positive effect on students' learning [177]. Geometry education was also targeted by Yeh [209] in which a VR system was informally evaluated with 6 children, though no learning outcomes were reported on. VR is applied in biology education for teaching about the structure of the human eye [166]. A formal evaluation found that students' self-report of engagement with the subject matter improved, and also that the VR group's score significantly improved post-exposure on a content assessment. The educational effectiveness of VR is also evaluated in the context of anatomy education [132] with significant results supporting the VR delivery method. Two groups of students (between-subject) were assessed on a 15 item anatomy quiz. An english learning achievement test is used to assess learners' improvements after exposure to a VR system [208]. Students' long-term learning was shown to improve significantly together with students' motivation. Sun et al. presented a system for teaching arts curriculum material which students engaged with, but did not appear to improve learning gains compared to current methods [183]. An earth motion simulator is presented on a desktop VR system and is evaluated with 21 sixth grade (approx. eleven year old) students with significant effects on students pre- and post-test scores [32].

Research has shown desktop VR can be implemented as an effective medium for education delivery. Immersive VR systems are able to provide the same advantages as desktop

VR in terms of flexibility of environment and complex visualisations, however they do so in a more involved way where the user is a part of the virtual environment. Additionally, they provide more integrated interaction metaphors and techniques due to that transportation and inclusion in the environment. We proceed by summarising general IVRL systems followed by a discussion of systems that have explicitly integrated, considered, or discussed pedagogical groundings in their design and development.

2.4.4 Learning in Immersive Virtual Reality

A common topic areas targeted by IVRL research are in the science domain such as physics and chemistry. A stereoscopic system incorporating haptic feedback was implemented for teaching astro-physics concepts [34]. It produced a positive effect on students' achievement as well as their motivation, autonomy, and encouragement. The work of Izatt et al. presents a CAVE system called Neutrino-KAVE to demonstrate neutrino physics to students [86] but unfortunately this system was not evaluated with users. Limniou et al. also used a CAVE approach to teach college students chemistry concepts with students' understanding improving after use of the immersive system [114] based on five open-ended achievement questions. ScienceSpace was an early HMD-based immersive VR project for teaching complex and abstract scientific concepts at secondary and college level with inconclusive results [46]. Cheng et al. show HMD-based VR to be effective for teaching cultural interactions in the context of language learning [31] but could not provide any conclusions regarding language learning.

Another earlier example called the VR Roving Vehicle Project is presented by Winn, and is comprised of VR systems allowing students to create their own virtual worlds. Students' attitudes toward various subject matter were shown to improve [201]. A CAVE system was evaluated in a study investigating the benefits of immersion for abstract information visualisation [142] and found the fully immersive application to be more useful in terms of both viewing the datasets and performing specific tasks. Johnson et al. report on The Round Earth Project [89], though only pilot studies were run. Preliminary observations are presented of a study using a collaborative CAVE-based VR system for teaching children about ecosystems [90]. Again, childrens' attitudes toward the subject matter improved based on observation. The authors point out that many studies evaluating learning gains of VR systems utilise paper and pencil assessments and propose that alternative measures should be applied to get more robust measures of the learning achievements in those systems.

So far there have been repeated instances of researchers identifying the need for more extensive research in the area of educational VR. For instance, Mikropoulos et al. identified the lack of immersive educational VR environments in a 10 year review [127]. They also discuss the need for more longitudinal studies, the importance of investigation of presence and learning [43], and state that little can be concluded regarding the retention (construction) of knowledge acquired in educational VR systems. For desktop and immersive systems alike, there is a need for a structured process of building learning into VR environments and furthermore, there is a need for revised assessment techniques. Questions are often raised in the current literature with respect to: 1) which topics immersive VR might be appropriate for and where is it effective? 2) what does effective mean in terms of immersive VR learning environments? And 3) how do we measure learning outcomes effectively when learning is conducted within IVRL environments. [3, 58, 77, 91, 106].

We do not address all of these questions in our work, though this does demonstrate the size of the research space we are working within. We focus on the interactive experience component of constructivist learning, and investigate what it can provide for IVRL environments. This informs the process of building learning and pedagogy into educational VR environments contributing to the structured process the community is in much need of. Below we cover the literature concerning VR learning projects that have explicitly integrated pedagogical foundations into the design of their systems.

2.4.5 Constructivism in Immersive Virtual Reality

While many educational VR systems have shown effectiveness in varying degrees, the results remain to be inconclusive, particularly for fully immersive VR systems. Researchers have identified a lack of pedagogical consideration in immersive VR learning developments [58]. VR learning projects that do mention pedagogy often mention and include constructivism [91]. This is likely due to the affordances provided by VR technology, especially immersive VR. In a review of almost two decades of educational VR systems, Dalgarno et al. identified the advantages 3D VEs provide for learning [43]. Among the identified affordances are "spatial knowledge representation, greater opportunities for experiential learning, increased motivation/engagement, [and] contextualisation of learning", which are in fact consistent with constructivist pedagogy. Johnston et al. also investigated current applications in the market place that target education [91] and found that the majority of them exhibit experiential principles.

Other work has also placed focus on the aspects of constructivist learning that are afforded by VR technologies, and how aspects of learning can be practically implemented in VR systems. For instance, Chee et al. present a collaborative, simulation-based desktop VR system designed on both experiential and constructivist/socio-constructivist principles though the system was not evaluated with users [29]. Peruzza et al. focus on the learner centred aspects of constructivist principles as a guide for the implementation of a modular desktop VR system demonstrated with physics material which, similarly, was not evaluated [125]. Hauptman presents "Virtual Spaces 1.0" for teaching spatial thinking [73] and found the system to significantly improve the non-verbal aspects of spatial thinking. Hauptman also used constructivist principles to inform the design of the system by integrating feedback and self-regulated (reflective) mechanisms.

Recent work [82] uses web-based desktop VR technologies for implementation of educational systems and proposes constructivist learning strategies that can be applied when developing VR learning environments. They use two case studies to evaluate VR learning environments concluding with a discussion of their insights. They discovered issues such as environment fidelity (compared to real world), difficulty to implement, and 3D user interface usability issues. Recent fully immersive technologies and state of the art computational systems are able to address many of these issues. Schwienhorst provides a detailed discussion on the concept of learner autonomy and VR in the context of computer-assisted language learning [163]. Three primary elements are identified as being naturally facilitated by VR: 1) awareness, 2) interaction and collaboration, and 3) experimental, learner-centred environments. The second element here is particularly relevant for this work. Similarly, Winterbottom identified constructivist "practical values" such as atomic simplicity, multiplicity, practical exploration, control, and reflective process, which might inform the design of immersive VR learning environments [203, 204].

The most relevant related work for us is that of Roussou et al. where user interaction in immersive virtual learning environments is investigated [151, 152]. The conclusion of the work suggests that interactivity facilitates childrens' problem-solving abilities, but conceptual formations and changes were not significantly effected. Rather, passive environments were shown to be more effective for conceptual changes. More recently, work from Roussou et al. evaluates learning in interactive VEs against passive VEs [153]. The subject matter used was arithmetical fraction problems. They concluded that, based on quantitative data, "participants in the VEs seemed to have a greater overall gain than those that did not perform the activity in VR" although they go on to say it was not clear if the interactivity was in fact the defining reason for the gain. The qualitative results of the study indicated the passive environment provided sustained conceptual change which supports their earlier findings. Roussou et al. discuss the timely need for research into interactive VEs for education given the recent proliferation of VR technology.

This thesis addresses three main gaps in the literature. The first is in the space of interactive experiences in IVRL environments. We have discussed the work of Roussou et al. who identified this as an area requiring research. Our work follows this path in our first investigation where we report on the general importance of interaction using an abstract subject matter - the 4th spatial dimension - while also investigating the effect on people with expert knowledge compared to laypeople. The findings highlight the importance of the interaction component and answer questions regarding their feasibility for users of different expertise which is important for the future development of computer-supported learning environments and the interactions within those environments.

The second gap we address is in the space of presence in IVRL environments and the techniques used to measure experiences in such a context. We reported on some of the work investigating the phenomenon of presence, however there is an apparent lack of enquiry into presence and immersive, interactive, learning environments. Our second investigation explores the potential impact of presence in such environments and also reports on new measurements that can contribute to the design and development of systems which aim to facilitate learning processes.

The final gap addressed in this thesis is more broadly with respect to immersive VR as a tool for learning. This is the space with the most prior work however it was noted that there is a lack of pedagogical consideration in the design of immersive VR applications, and IVRL systems have only recently become available on a larger scale. This thesis touches on this general topic, and in particular, the third component of this work discusses a real world IVRL project providing both contextual and general insights into the employment of immersive VR systems in complex learning environments.

In the following chapter, we describe the subject matter and research apparatus we use throughout our studies for our investigation of learning in immersive virtual environments.

Chapter 3

Concept and System

Contents

3.1 Arnold's Subject Matter **30**
3.2 The Fourth Spatial Dimension **30**
3.3 Implementation **33**

For any investigation of learning to take place, novel subject matter needs to be presented to learners. The mathematical concept of the fourth spatial dimension provides the required novelty, and through the development of a research apparatus, we provide learners with an interactive experience of constructs within abstract mathematical space. In this chapter, we describe the mathematical subject matter used throughout this thesis for our investigation of learning in VR environments. We also detail the principal research apparatus we have implemented to facilitate the investigation. The apparatus is realised through two systems, each of which provides learners with an experience of the fourth spatial dimension such that they might gain comprehension—that is Arnold's "deep non-verbal comprehension" or von Foerster's "begreifen". We begin by describing the primary motivation behind selection of the subject matter, and the apparatus design. Parts of this chapter are included in the following work:

1. Collins, J., Regenbrecht, H., and Langlotz, T. (2019). Interaction and expertise in immersive virtual reality learning environments. *In IFIP Conference on Human-Computer Interaction (Submitted).* [38]

2. Collins, J., Regenbrecht, H., Langlotz, T., Can, Y. S., Ersoy, C., and Butson, R. (2019). Measuring Emotional Response in Virtual Environments: A Methodology Exemplified in a Learning Context. *In ISMAR '19 (Submitted).* [39]

3. Collins, J., Regenbrecht, H., Langlotz, T. (2018). Back to the future: Constructivist learning in virtual reality. *In: Adjunct Proceedings of the IEEE International Symposium for Mixed and Augmented Reality 2018 (Poster).* [37]

3.1 Arnold's Subject Matter

When investigating learning in empirical studies, we need to consider the effect of participants' existing knowledge to mitigate any unwanted effect. A potential solution is to use children as participants in educational studies and to present them with content they would otherwise be learning anyway (i.e. arithmetic or fractions). This is a valid approach, although there is still the issue that children have already been exposed at different stages to those concepts due to their prevalence in everyday life. To find common content to deliver to adults for educational purposes is arguably even more complex due to adults' further exposure to many of the commonly selected topics, and the diverse range of exposure adults carry with them. Arnold's solution was to use the abstract concept of four-dimensional (4D) space as subject matter [7]. The assumption is that not many people have a firm spatial or mathematical understanding of 4D space making it an ideal subject matter for learning. He proposed an apparatus with a stereoscopic VR system as its core component. The system should be capable of visualising a 3D projection of a hypercube (a cube in 4D) and further allow for direct manipulation of the hypercube, in the fourth dimension, in real-time. In his original proposal the hypercube, or more precisely its' rotation in 4D, is controlled with six dials and is then rendered on a stereoscopic CRT monitor (see Figure 3.1 (left)) [8]. Although Arnold's work was revisited at a later date [194], the actual apparatus and study appear never to have been realised.

The subject matter and apparatus proposed by Arnold inspire the developments used for the main body of this work. We implement the interactive hypercube apparatus in two ways. The first is closely aligned with Arnold's originally proposed system (see Figure 3.1 (middle)) and the second is developed on the same principal components, but utilises fully immersive technology for the interactive and visual mediums (see Figure 3.1 (right)).

3.2 The Fourth Spatial Dimension

Most of us are capable of understanding one- (1D), two- (2D), and 3D concepts (see Figure 3.2). We are able to imagine three dimensions which exist perpendicular to each other, however, to attempt imagining a fourth dimension which maintains perpendicularity

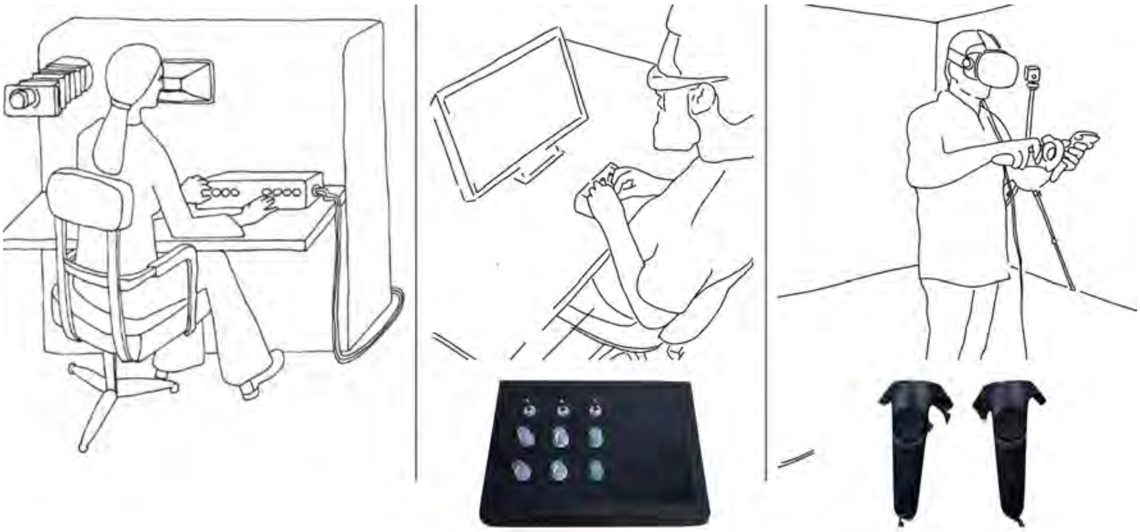


Figure 3.1: Left: the original setup proposed by Arnold has a user sitting at a station with dials for manipulating a 4-Dimensional cube and a stereoscopic viewing platform for visualisation of the resulting user manipulations [7, 8]. Middle: our implementation of Arnold's original proposal. Right: our implementation of the original principal but with modern immersive technology.

with the previous three seems near impossible. Therefore, a vertex in 4D-space possesses not just x, y, and z coordinates, but an additional fourth coordinate.

Shapes in four dimensional space possess the same characteristics as their 3D counterparts. Arnold makes mention of different potential 4D shapes. For instance, he describes a pentahedroid, tesseract (hypercube), and a handed block (from a soma cube) as potential shapes to be implemented. We opt to implement a hypercube as our primary target shape. The concept of 4D space is already complicated, and we want to facilitate one's "begreifen" of the subject matter. It is unnecessary to further complicate the subject matter by presenting more complex shapes. A hypercube is a 4D cube and therefore shares the same characteristics; all edges are the same length, all internal angles are 90 degrees, but where a regular 3D cube consists of 8 vertices, 12 edges, and 6 faces, a hypercube consists of 16 vertices, 32 edges, and 24 faces.

Once the hypercube and 4D space are defined we can consider 4D transformations. Common transformations include scaling, translation, and rotation. We select rotation as the primary form of manipulation to be applied to the 4D cube. This is due to Arnold's original description in which one turns (rotates) dials resulting in rotational manipulations. This maintains a coherent interaction metaphor based on rotation. A common technique for rotation operations in 3D space is that of Euler angles, i.e. rotating an object about

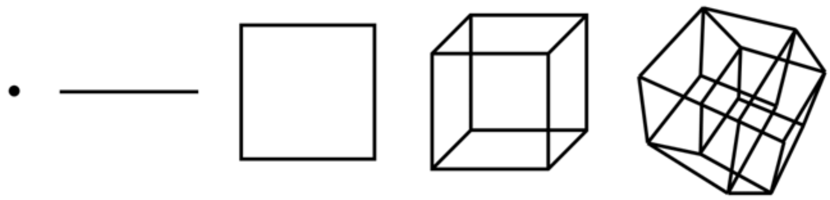


Figure 3.2: Demonstrates the steps through from 0D to 4D.

each of its axes by a chosen degree of rotation. While this technique is easy to understand and apply rotation operations with, it suffers from a problem commonly known as gimble lock. This occurs when two of the three axes of an object are rotated in such a way that they become parallel with each other, effectively limiting the object to rotation in only two degrees of freedom. An alternative to using Euler angles in 3D space is quaternions, or more specifically, 'unit length quaternions' but we use the term quaternions throughout this chapter. Quaternions are an extension of imaginary numbers which can also be used to define rotations in three dimensions. They encode an axis-angle representation in four numbers (x, y, z, w) which can be applied to a 3D point.

As object representations change between the third and fourth dimensions, so too does rotation. The gimble lock problem however exists when rotating 4D objects using Euler angles just as it does in the third dimension. As specified earlier, quaternions are only applicable to 3D space however there are 4D equivalents called octonions which are similar to quaternions in principal (except they represent a rotation in 4D space instead of 3D), but contain four additional imaginary components (resulting in seven + one real number). Arthur Cayley discovered 4D rotations can be decomposed into a double quaternion representation* [28]. The more recent work of Thomas et al. [135] presents an approach to this decomposition based on characteristics of 4X4 matrices which we can take advantage of in our implementation (details discussed later in subsection 3.3.2). Therefore, to effectively rotate an object in 4D space requires input devices amounting to two quaternions.

To enable the actual user experience of the subject matter described above, we implement the principal apparatus, and realise the interactive experience of 4D space through two systems. The resulting apparatuses are utilised throughout this thesis for the investigation of learning in VR environments. In the remainder of this chapter, we describe the implementation of our principal systems in detail.

*wikipedia.org/wiki/Cayley-Dickson_construction

3.3 Implementation

Three main components form our principal system: 1) the interactive component, 2) the operative component, and 3) the visualisation component. The operative component of our system is responsible for receiving input from (1), performing the required operations, and outputting to (3), and it is built within the Unity3D (https://unity3d.com/) framework. Figure 3.3 presents the workflow and system configuration.

First we generate the hypercube by defining the 16 vertices that comprise the shape (as discussed above). Operations such as rotation, projection, translation, and scaling are also defined. We will describe these operations and their respective implementations according to the workflow of the system, beginning with the interaction component.

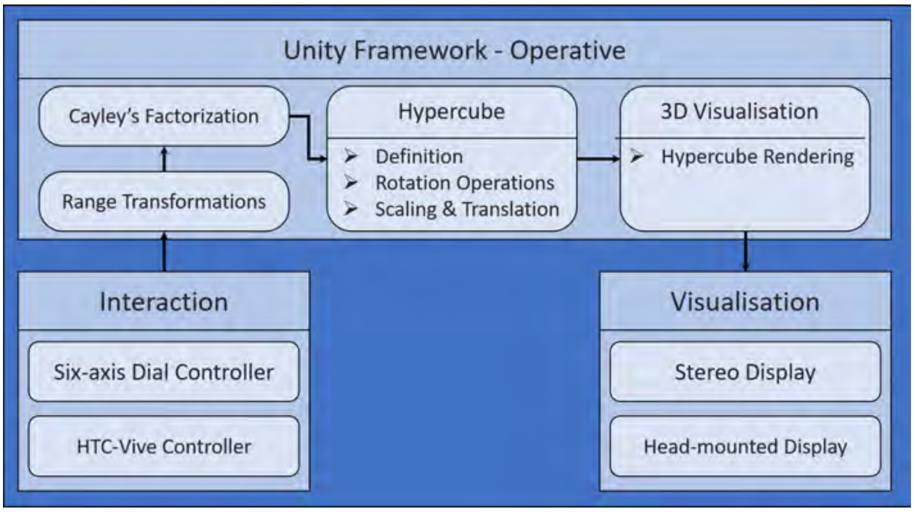


Figure 3.3: The primary components of our system are shown indicating the workflow. The interaction devices provide input to our system's operative component (Unity3D) where operations are applied, which is in turn rendered to the visualisation component for the user.

3.3.1 Interaction Component

We have implemented two different forms of interaction which enable an interactive experience of a hypercube. In this system, users will be able to rotate the 4D cube using: (1) a six dial input device (as proposed in previous works of Arnold [7, 8] and von Foerster [194]), and (2) two HTC Vive† controllers. Hence, this forms one part of the interactive experience, being able to interact with a hypercube by mathematically rotating it (see

†https://www.vive.com/

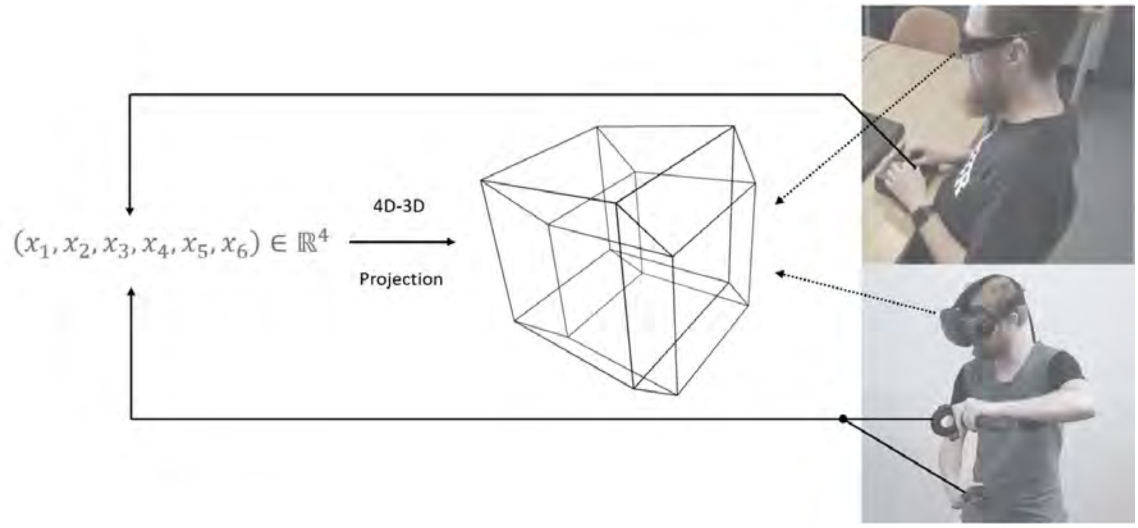


Figure 3.4: Concept of the hypercube system: An operator (right) is controlling the rotations of the 4D mathematical cube either by turning dials on a board, or rotating the hands in real space (2x3DOF). The manipulated hypercube rotation is then projected into 3D space to be observed by the operator either in an HMD, or with 3D glasses on a screen. This implements a closed, interactive feedback loop.

Figure 3.4). Each of the interaction mediums is described below.

The six dial device (shown in Figure 3.5 (left)) is driven using a Freetronics Leostick[‡] which takes six analogue potentiometers as input. The three toggle switches on the device are not used. The Leostick is programmed in C using Arduino software, and is configured to connect to a PC as a joystick device. Arnold carefully selected analogue potentiometers (rheostats) as input devices for rotational manipulation because at any given position, the dial maps to an actual rotation (i.e. when turned half way, it maps to a rotation of 180 degrees). Digital dials could be used, but the user would experience difficulty knowing the rotation of the hypercube in the respective axis. The potentiometer inputs are converted to 8 bit signals (0 - 255) as axis outputs. Unity3D's input management allows us to receive the raw input from the Leostick as joystick axes (due to its configuration) which is then passed to a range transformation operation which maps the raw potentiometer outputs to Euler angles. Values are shifted from [0,255] to [0,360] using a standard range transformation:

$$y' = \frac{x - A}{B - A} \times (C - D) + D \tag{3.1}$$

where y′ is the transformed value, x is the input value (from a dial), A is the minimum

[‡]www.freetronics.com.au

input range, B is the maximum input range, C is the maximum target range, and D is the minimum target range. The potentiometers are quite sensitive and result in jitter once applied to the hypercube. We implement a sliding window of 30 input frames to smooth out the potentiometer inputs. This results in the slight appearance of latency, however the update rate remains steady, and manipulations are easily conducted. Once this operation is performed for each of the six dial axes received, we can form two 3D rotations. Unity3D contains its own quaternion library which provides a function that takes three Euler angles as arguments and returns a quaternion that is representative of such a Euler rotation. Given our six input dials that are transformed into Euler angle values, we are able to use Unity's quaternion library to generate two separate quaternions ready to be combined into a four dimensional rotation. Before moving on to rotation handling we will discuss the second interactive medium, the HTC Vive controllers.

Figure 3.5: The six dial device as an interface device analogous with Arnold's previous description (left), and the HTC Vive controllers (right), both used as input devices for the rotation of the hypercube in 4D space.

The HTC Vive, made by Valve in collaboration with HTC, provides us with a fully immersive interaction technique. The Vive comes with a head-mounted display (HMD), two wireless controllers (see Figure 3.5 (right)), and two base stations. The HMD and two controllers are tracked in an area that is approximately 4m x 3m. Plugins are provided for integration with various development environments including Unity. An OpenVR package is available free for download and provides a prefab[§] which represents the tracking space, the HMD, and the two controllers in the 3D environment. When the system is running the user can see the two controllers in the VR space. All objects in a Unity scene hierarchy have their orientation represented by quaternions. That means, when the Vive controllers are being tracked, we are able to access the absolute orientation of both controllers in space

[§]Prefabs in Unity3D are designed for reuse and are instances of a pre-configured objects often comprised of models, scripts, and other required components.

which therefore provides us with two separate quaternions. We do not require the range transformation of independent axes (as described above for the six dial device) because we can access the controller quaternions directly. Once we have the quaternions, they are ready for rotation operations.

3.3.2 Operative Component

We now come back to Arthur Caley's decomposition of 4D rotations that was mentioned earlier. The overall idea is to combine two input quaternions to form a 4D rotation. More specifically, the generalisation states that a 4D rotation can be decomposed into left and right isoclinic rotations and can be represented by rotation matrices of the form:

$$R^L = \begin{pmatrix} l_0 & -l_3 & l_2 & -l_1 \\ l_3 & l_0 & -l_1 & -l_2 \\ -l_2 & l_1 & l_0 & -l_3 \\ l_1 & l_2 & l_3 & l_0 \end{pmatrix} \tag{3.2}$$

and

$$R^R = \begin{pmatrix} r_0 & -r_3 & r_2 & r_1 \\ r_3 & r_0 & -r_1 & r_2 \\ -r_2 & r_1 & r_0 & r_3 \\ -r_1 & -r_2 & -r_3 & r_0 \end{pmatrix} \tag{3.3}$$

These isoclinic rotation matrices hold a commutative property such that a 4D rotation matrix, R, can be expressed as:

$$R = R^L R^R = R^R R^L \tag{3.4}$$

where:

$$R^L = l_0 I + l_1 A_1 + l_2 A_2 + l_3 A_3 \tag{3.5}$$

and

$$R^R = r_0 I + r_1 B_1 + r_2 B_2 + r_3 B_3 \tag{3.6}$$

I is the 4x4 identity matrix and our left (A) and right (B) bases for the respective isoclinic rotations have the property such that:

$$A_1^2 + A_2^2 + A_3^2 = A_1A_2A_3 = -I, \tag{3.7}$$

and

$$B_1^2 + B_2^2 + B_3^2 = B_1B_2B_3 = -I, \tag{3.8}$$

Therefore, if we have two quaternions, Q^{L} and Q^{R}, of the form:

$$Q = w + xi + yj + zk \tag{3.9}$$

for which the imaginary components hold the property:

$$i^2 = j^2 = k^2 = ijk = -1 \tag{3.10}$$

we can multiply their components according to Equation 3.5 and Equation 3.6 respectively, resulting in two (left- and right-isolinic) rotation matrices which can be multiplied together as in Equation 3.4. The resulting 4D rotation matrix (R) is then ready to be applied to our 4D vertices. A vertex in $\mathbb{R}^4$ is defined as:

$$v = \begin{pmatrix} x \\ y \\ z \\ w \end{pmatrix} \tag{3.11}$$

Then we apply the final rotation matrix R to each of our 16 hypercube vertices:

$$v' = Rv \tag{3.12}$$

resulting in an orientation of the hypercube based on the two quaternions acquired from input devices. Once the rotation operation is applied in 4D space, the hypercube is projected into 3D space. We use a projection matrix:

$$Proj = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & pw & 1 \end{pmatrix} \tag{3.13}$$

where pw is the extent of perspective to be applied. When pw is zero the projection will be orthographic. Of these two common forms of projection we use an orthographic projection throughout our studies. Once the projection matrix is applied to the hypercube vertices, the 3D projection operation is applied. Given a 4D vertex as in Equation 3.11, in order to project from 4D to 3D, we scale the x, y, and z point coordinates by the fourth coordinate (w):

$$P = \begin{pmatrix} \frac{x}{w} & \frac{y}{w} & \frac{z}{w} & 1 \end{pmatrix}^T \tag{3.14}$$

Unity3D provides a structure (Vector3) for representing 3D points which we use to hold the projected x, y, and z values from the hypercube vertices giving us 16 Vector3 points. Once the hypercube points are projected from 4D space, we apply any scaling and translation operations that we might need to aid moving or resizing the hypercube. Translation is therefore:

$$T_{i,j,k}(x, y, z) = (x \pm i, y \pm j, z \pm k) \tag{3.15}$$

for each of the hypercube's 16 vertices post-projection. Scaling operations are also applied to the projected hypercube as expected using:

$$S_k(x, y, z) = (xk, yk, zk) \tag{3.16}$$

With the input received from the users actions, the operations computed and applied to the construct, the hypercube is ready for visualisation.

3.3.3 Visualisation Component

Unity provides line rendering techniques which allow us to input two vertices (per line). Given the 16 projected vertices we attained using the process described above, we can draw the 24 edges by connecting the appropriate 3D vertices. We create 24 "edges" which contain two indices representing the vertices (defined earlier) to be connected. We connect the vertices in the same manner as one would do with a 3D cube, though there are an additional eight vertices. This means that each vertex has four edges connecting to it, not three. In order to aid the visualisation we also added the option of having light-grey semi-transparent faces (see Figure 3.6).

Just as we provide two interaction mediums (six-dial and Vive controller approaches), we also provide two visual mediums: (1) stereoscopic display and (2) HTC Vive HMD. If the six dial input device is being used, we visualise on a 3D monitor (120Hz refresh

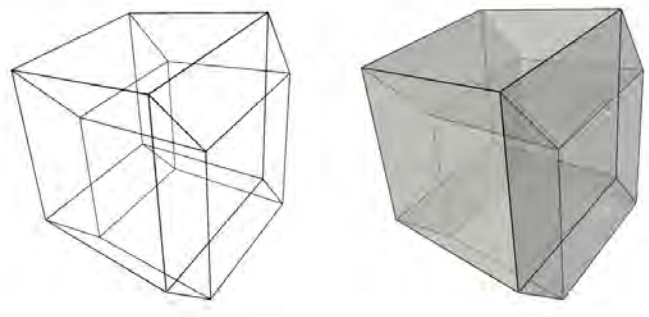


Figure 3.6: Hypercube rendered with edges only (left) and rendered with semi-transparent faces (right).

rate) as a stereo-image (2X2D images, 60Hz refresh rate per eye) which the user can view by wearing stereoscopic glasses. Unity3D has a built in stereoscopic library which handles splitting the scene into stereo views for rendering. The stereoscopic implementation runs on an Alienware M17X (i7-2860, NVIDIA GTX580M, 24GB RAM). Our resulting implementation runs at 120Hz, with no noticeable latency with the exception of the sliding window frame buffer for the input.

A head-mounted display (HMD) allows for the immersive viewing of a stereoscopic image by having two small monitors with tailored magnifying lenses in front of the user's eyes. Usually, the user does not see anything else from the real environment. The position and orientation of the HMD is tracked in real space so the respective views into a virtual environment are rendered so the user gets the impression they are looking around a virtual object and within a virtual environment. We are using this immersive technology to give operators a 3D viewing experience of our projected hypercube. This provides an advantage over the stereoscopic display approach in that the user is able to move around the hypercube and therefore achieve different views. Our fully immersive implementation runs on a desktop PC (i7-6700, NVIDIA GTX970, 8GB RAM), rendering to the HTC Vive HMD with a total resolution of 2160x1200, running at a 90Hz refresh rate. The final system runs at over 100 frames per second with no noticeable latency.

We have described two different methods of experiencing the fourth spatial dimension through direct interaction. There are notable differences between the experiences. The six-dial device provides more refined control of the hypercube rotations as rotational components of the hypercube can easily be altered independently. Conversely, the immersive system has a user holding two controllers which, when an arbitrary rotation is made of one controller, will usually result in changes to multiple dimensions of rotation. Further differences are found in the visualisation. We project the hypercube manually from 4D-3D space as one of the functions in our operational component. In the immersive system, the hypercube is visualised in 3D space, through the HMD, so the hypercube has only undergone one projection from 4D to 3D. In the stereo system however, the hypercube is projected firstly from 4D to 3D, and then an additional time from 3D to 2D, resulting in the final projection from 4D-3D-2X2D. Additionally, in the stereo system, the point-of-view is fixed to the position of the virtual camera in the Unity environment. In the immersive system, users can freely move and look around in six degrees of freedom (DOF). It is worth noting the apparatus description of Arnold with particular focus on his system specifications. He described the visualisation stating "The images range over a square approximately nine by nine inches on a cathode ray tube (vector scope)". He went on to describe the update rate: "the cycle time of checking the dial setting, recomputing the angle, and displaying is from 1 to 20 times a second, and thus a fairly smooth real-time response ... is possible" [8].

A Preliminary Implementation. Prior to the full development of the systems described above, we implemented a pilot version of our system. Before immersive VR controllers were around, and before we had a six-dial input device, we utilised two Nexus 4 mobile smart phones as input controllers. We used the built-in gyroscope sensor from each phone, and sent the sensor data over wireless network via a socket connection to Unity. Gyroscope readings were unreliable and resulted in much jitter and erratic readings which we could mitigate with a sliding window implementation. We visualised the projected hypercube in an Oculus Rift DK2. This gave us an idea of how the system could function and allowed us to test the visualisation together with part of the operational component. We were able to run informal pilot tests of the system at a bi-annual international science festival that runs at the University of Otago. The main users at that time were children. We were able to utilise some of the feedback from users to refine the system for the further implementation described above.

We proceed to utilise these apparatuses throughout this thesis work to facilitate our investigation of learning in VR environments. In the following chapter, we begin to report on our first branch of investigation—interaction and expertise in immersive VR learning environments.

Chapter 4

Validating Historic Measures

Contents

4.1 Introduction **42**
4.2 Experimental Design **42**
4.3 Measures **43**
4.4 Procedure **44**
4.5 Results **46**
4.6 Discussion **49**

As described in chapter 1, there are two main branches of investigation. This chapter is the beginning of the first investigation of the interplay between interaction and knowledge in IVRL environments. Several aspects of Arnold's proposal demonstrate potential for the investigations conducted within this work. Given that Arnold's and von Foerster's works [7, 8, 194] were just proposals and were not actually formally evaluated, we need to validate the aspects of their work that we use for our investigations. The purpose of the first experiment reported in this chapter, is to validate the subject matter, the operation of the research apparatus described in the previous chapter, and the measures proposed by Arnold. Parts of this chapter are included in the following work:

1. Collins, J., Regenbrecht, H., and Langlotz, T. (2019). The Uniqueness of Experience for Learning in Immersive Virtual Reality. *In International Journal of Human-Computer Studies (Submitted).* [38]

2. Collins, J., Regenbrecht, H., Langlotz, T. (2018). Back to the future: Constructivist learning in virtual reality. *In: Adjunct Proceedings of the IEEE International Symposium for Mixed and Augmented Reality 2018 (Poster).* [37]

4.1 Introduction

Throughout previous sections we have elaborated on the concept of the fourth spatial dimension and its qualification as subject matter for our experimentation. It is not a concept that has likely been taught to, or learned by the average person and it is certainly not a concept we can conceive of just by interacting with the reality around us. However, this also makes it more difficult to measure one's comprehension of such a concept. When Arnold originally proposed this method for investigating learning differences between observers and actors, he needed to devise a method for evaluating learners' comprehension. He suggested presenting participants with a set of tasks in order to determine whether they had acquired either a "partial or complete mastery of the situation" [8], though the task descriptions lacked detail. We use these broad descriptions to guide the design and implementation of our measures.

Arnold claimed there would be indications that a learner has achieved the "deep non-verbal comprehension"* [8] of the subject matter. In particular, he stated that one's achievement of comprehension would result in objective effects such as a learner having the ability to perform tasks in 4-space, and subjective effects such as having the ability to spot inconsistencies in figures. We have implemented two of Arnold's measures for validation: (1) a subjective style measure in which we present possible and impossible hypercubes where the learner must distinguish representations, and (2) an objective task-based measure where a target hypercube is presented and learners must manipulate a second hypercube to match the exact form of the target. More detail is provided on these measures below in section 4.3.

Before we can begin to address our first research question, we need to verify our study system and the measures we plan to use throughout this thesis. Therefore, we have one hypothesis for this experiment: (H1) Our implementation of Arnold's measures will result in a valid measure of comprehension.

4.2 Experimental Design

The experiment is a single factor between-participant design with two conditions. Our two systems comprise our between-participant conditions: (1) the six-dial interface while viewing on a 3D stereoscopic display (3D glasses) - *desktop experience*, and (2) the Vive

*This form of comprehension, or learning, was later termed in German as 'begreifen' by von Foerster [194] controller interface while viewing in an HMD - *immersive experience*. Participants' conditions are pre-randomised.

Participants. We recruited from the university student and staff populations with the only inclusion criterion being age (between 18 and 65).

4.3 Measures

Here we describe the types of data collected and the tools used.

Demographics Questionnaire. A demographics questionnaire collects data including age, gender, ethnicity, vision, and prior VR experience.

Self-assessment Knowledge Questionnaire. We designed a short knowledge questionnaire presenting two questions intended as a self-assessment of the participants' current understanding of 4D space and the Hypercube (or Tesseract). The first question is "I understand the concept of the fourth spatial dimension", and the second is "I know what a Hypercube is." Both questions are answered on a Likert-like scale where the range is from -3 ("Not at all") to 3 ("Very Much").

Hypercube Assessment Questionnaire. This is the first of Arnold's original measures we selected and we call this the *hypercube assessment questionnaire*. The hypothesis stated by Arnold was that if a participant is able to competently distinguish correct and incorrect hypercubes from each other, this is one indication that they have likely formed a 'correct' internal representation of a hypercube. Therefore, as a measure, we snapshot images of hypercubes rotated in various different ways which comprise a set of correct hypercubes. We then snapshot various different rotated hypercubes with different obscurities (rendering impossible hypercubes). Overall, 36 total snapshots form the hypercube assessment on which participants should tick only hypercubes they believe are correct/possible (see Figure 4.1 for example correct/incorrect hypercube forms).

Ghostcube Matching Task. The second of Arnold's measures places a learner in control of one hypercube where their task is to manipulate it so that it matches to a second statically rotated hypercube. We call this the *ghostcube matching task*. This gives us an additional requirement of our system described in chapter 3. With that implementation,

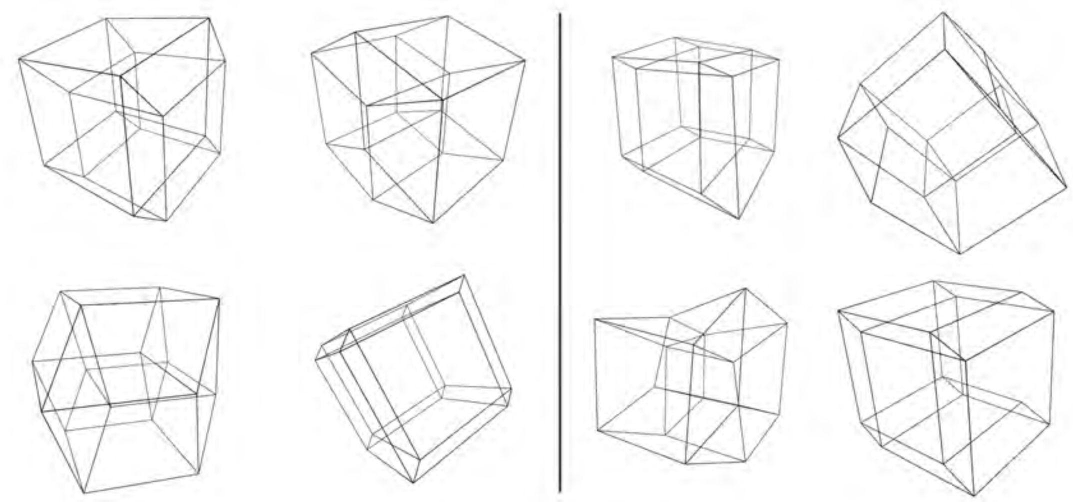


Figure 4.1: Example hypercube forms presented to the participants where their challenge is to distinguish between correct and incorrect. On the left are examples of correct hypercubes, and on the right are incorrectly rendered hypercubes. Hypercubes are assigned randomly to the pages of the paper-based hypercube assessment questionnaire.

we project a static hypercube with a preset rotation (we call this the ghostcube). Next to the ghostcube, we project a second hypercube which can be manipulated using one of the interactive mediums described above. The task for the user is to rotate their hypercube so that it matches the form of the ghostcube. The user manipulates their hypercube until they believe they have the right form then they must hold that form for 2 seconds at which time, if the form is correct, the manipulation will halt and a success indicator is given.

The comparison algorithm compares all hypercube vertex positions against each other, i.e. there are multiple possible solutions for the task. The independent rotations of the users' hypercube axes do not have to match the respective rotations of the ghostcube axes for a solution to be considered correct.

4.4 Procedure

Upon arrival participants are greeted, introduced to the study, given a consent form and if agreed presented with the demographics questionnaire. Before beginning the study, participants are given the knowledge questionnaire in order to help determine their current knowledge of our subject matter - 4D space and the hypercube (Figure 4.2-Q1). We then present to each participant a short explanatory video (Figure 4.2-A) introducing par-

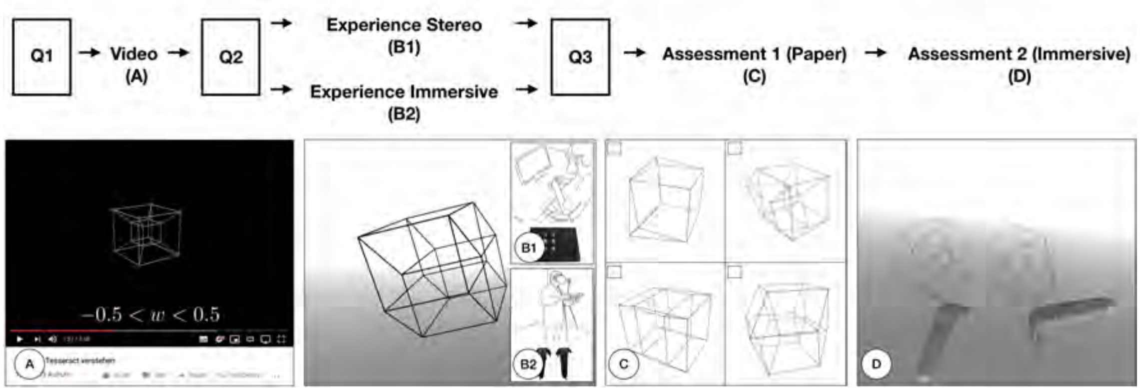


Figure 4.2: This figure demonstrates the experimental procedure for Study 1. Participants watched a short video, diverged into their conditions, completed the hypercube assessment, then all participants attempted the ghostcube matching task in the immersive system. Q1, Q2, and Q3 are the instances in which knowledge questions were asked to participants.

ticipants to the context of the subject matter ensuring all participants are of a similar knowledge level. The self-assessment knowledge questionnaire is once again presented after having viewed the video (Figure 4.2-Q2). At this stage participants diverge into their pre-randomised condition in which they experience the hypercube on their assigned system, either desktop experience (Figure 4.2-B1) or immersive experience (Figure 4.2-B2), for a five minute period. Participants are then asked, for the third and final time, to fill out the self-assessment knowledge questionnaire (Figure 4.2-Q3) followed by the hypercube assessment questionnaire described earlier (Figure 4.2-C). The final task for participants before being released is the ghostcube matching task (Figure 4.2-D). All participants perform this task in the immersive system (HTC Vive) to allow us to investigate how experience on one system affects performance on another system. There is an upper time limit of 10 minutes to complete the task. Upon completion it automatically stops and stores their participant ID, and the completion time. Participants are asked to spend at least two minutes trying to complete the task after which they are allowed to give up if they so choose. Finally, participants are thanked, compensated for their time, and released.

Upon closing the final experiment, we decided to improve the validity testing of our paper-based hypercube assessment. We recruited an additional 11 participants for 10 minutes to fill out the hypercube assessment without having any experience at all. They were only asked to distinguish correct and incorrect 4D cubes (hypercubes) from each other. This would serve as a baseline result for our hypercube assessment.

4.5 Results

There were 22 participants (15 male, 7 female) in total with a mean age of 25.9. All participants completed the experiment in full with only one participant giving up in the final task. Different forms of analyses are used for each measure beginning with the self-assessment knowledge questionnaire results.

Shapiro-Wilk tests are used throughout the analysis for distribution testing of the data. Where data is not normal (non-parametric), we apply a Wilcoxon Signed Rank test for analysing significance of means differences on paired data, and a Wilcoxon Rank Sum test (equivalent to a Mann-Whitney U test) for unpaired data. Where data is normally distributed (parametric), t-tests are applied for analysing significance of paired or unpaired data.

Self-report Knowledge Assessment Questionnaire. Questionnaire data was of a non-parametric distribution (Shapiro-Wilk). Wilcoxon Signed-rank tests show a significant increase in reported ratings between the first and second instances for both questions (Q1: first instance median = 3.0, second instance median = 5.0, $p < 0.01$; Q2: first instance median = 1.0, second instance median = 6.0, $p < 0.01$). No significance was shown between the second and third instances (Q1: third instance median = 6.0; Q2: third instance median = 6.0; $p > 0.05$). Results for the self-report questionnaire are visualised in Figure 4.3.

Significance tests for independent groups revealed the stereo group increased significantly between the first and second instances for both questions ($p < 0.05$) but not between second and third instances (Stereo Q1: first instance median = 3.0, second instance median = 5.0, third instance median = 6.0; Stereo Q2: first instance median = 4.0, second instance median = 6.0, third instance median = 6.0). These results are supported by a Friedman's ANOVA (First Question: $X^2(2) = 13.15, p < 0.01$, Second Question: $X^2(2) = 16.424, p < 0.01$). The immersive group reported significant increases across the first and second instances for both questions ($p < 0.05$) and also increases across the second and third instances ($p < 0.05$) for Q2 only (Immersive Q1: first instance median = 2.0, second instance median = 5.0, third instance median = 6.0; Immersive Q2: first instance median = 1.0, second instance median = 5.0, third instance median = 6.0). For the rest, no significance was found ($p > 0.05$). A Friedman's ANOVA confirmed these results (First Question: $X^2(2) = 17.684, p < 0.01$, Second Question: $X^2(2) = 17.897, p < 0.01$).

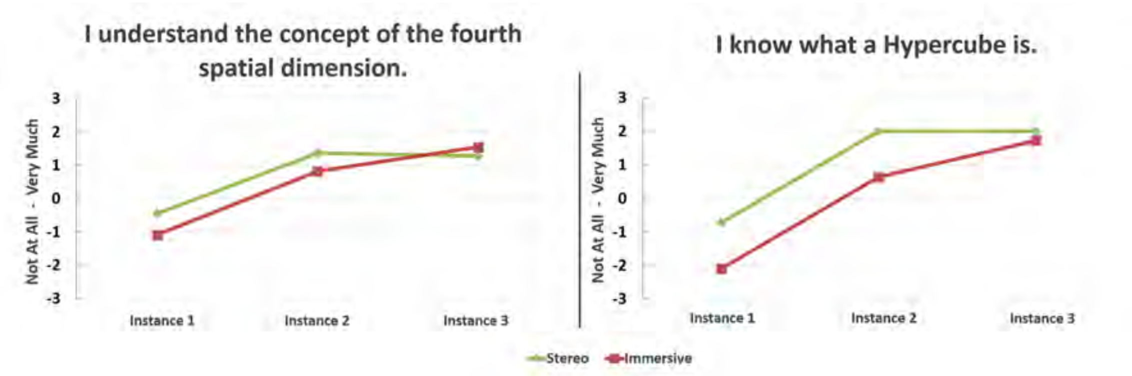


Figure 4.3: Mean Knowledge Questionnaire results for question one and two where the three instances correlate to the three times participants filled out the questionnaire during the experiment. Significant increase in reported knowledge is found between the first and second instance for both questions. Significant increase between the second and third instance only found for the immersive group for the second question "I know what a Hypercube is."

Hypercube Assessment Questionnaire. The assessment data is analysed in terms of Positive Prediction Power (PPP) and Negative Prediction Power (NPP) which is stated to be a measure of accuracy [185]. This accounts for false positive and false negative values reported by participants. PPP is computed using:

$$PPP = H/(H + FA) \tag{4.1}$$

where H represents the number of correctly ticked hypercubes and FA represents the number of false alarms - incorrect hypercubes that participants ticked. A participant that ticks all the correct cubes, and none of the incorrect cubes receives a PPP value of 1. NPP is computed using:

$$NPP = CR/(CR + M) \tag{4.2}$$

where CR represents correct rejections - incorrect hypercubes participants correctly left unticked - and M represents misses - correct hypercubes that participants incorrectly left unticked. A participant that leaves all incorrect hypercubes unticked and leaves no correct hypercubes unticked receives a NPP value of 1. We analyse and later interpret these values in conjunction with the overall scores of participants. It is possible for a participant to have a perfect PPP or NPP score, but have a very low overall score. For instance, in a hypothetical scenario a participant correctly ticks two out of a total of eight possible hypercubes, and does not tick any of the four impossible hypercubes. Their PPP score would be $2/(2+0) = 1.0$ (100%), but their overall score would be only 50%.

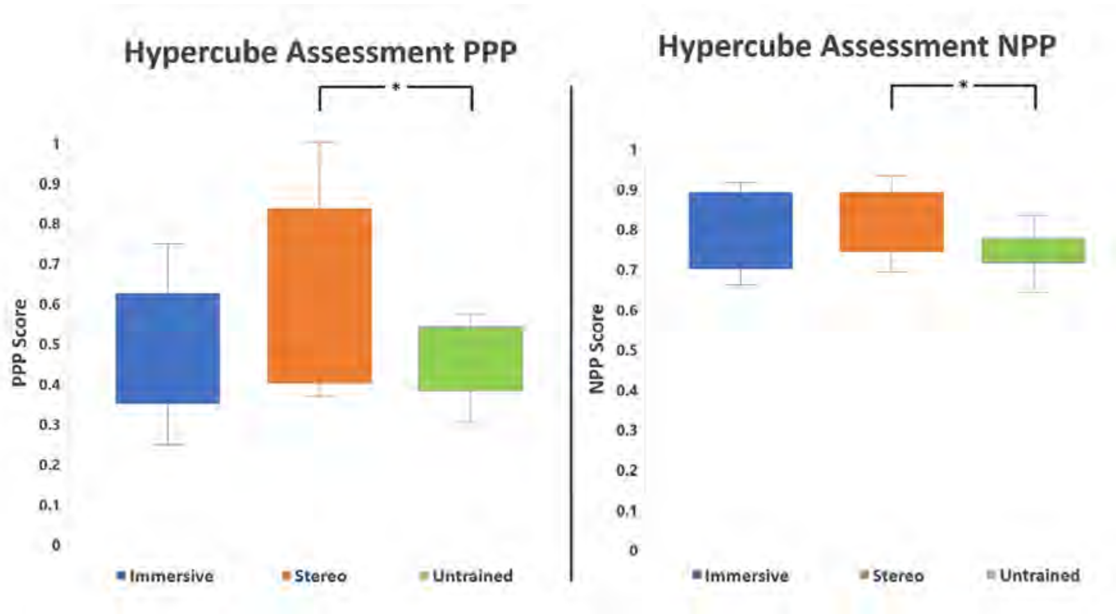


Figure 4.4: Positive and Negative Prediction Power (PPP/NPP) scores for the first and second hypercube assessments (left and right respectively). Significance is found between the stereo (desktop) and untrained groups but no significance is found in other group comparisons. ** denotes statistical significance where $p < 0.05$.*

A Shapiro-Wilk test revealed a parametric distribution. A t-test revealed no significant differences in the PPP/NPP results of the stereo (*M=0.65/0.80, S.D=0.22/0.08*) and immersive (*M=0.52/0.79, S.D=0.15/0.09*) groups ($p > 0.05$).

The additional 11 participants recruited after the initial part of the study were asked to tick pictures they believed were possible/correct 4D cubes. This forms, in essence, an additional participant group for this assessment only. The reason we did this is to further inform us of the sensitivity of the assessment as a measure. The untrained group scored lower mean scores than both original groups for both PPP/NPP (*M=0.46/0.72, S.D=0.09/0.05*). An unpaired t-test revealed significance between the desktop (stereo) and untrained group ($p < 0.05$) for both PPP and NPP. No significance was found between the immersive and untrained groups ($p > 0.05$). We applied an ANOVA which confirmed a significant effect of our conditions on participants' PPP performance on the assessment ($F(2,30) = 3.53$, $p < 0.05$, $\omega = 0.37$), but not for NPP ($F(2,30) = 2.71$, $p > 0.05$, $\omega = 0.28$).

The overall assessment scores (scored out of 36) reflect the PPP/NPP results. Both the immersive group *(M=24.73, S.D=3.22)* and the desktop stereo group *(M=25.91, S.D=4.44)* had higher overall scores than the additional untrained group *(M=22.1, S.D=2.43)*. A Shapiro-Wilk test found a parametric distribution and an unpaired t-test revealed significant differences between both the immersive and untrained, and stereo and untrained groups *($p < 0.05$)* (see figure Figure 4.5).

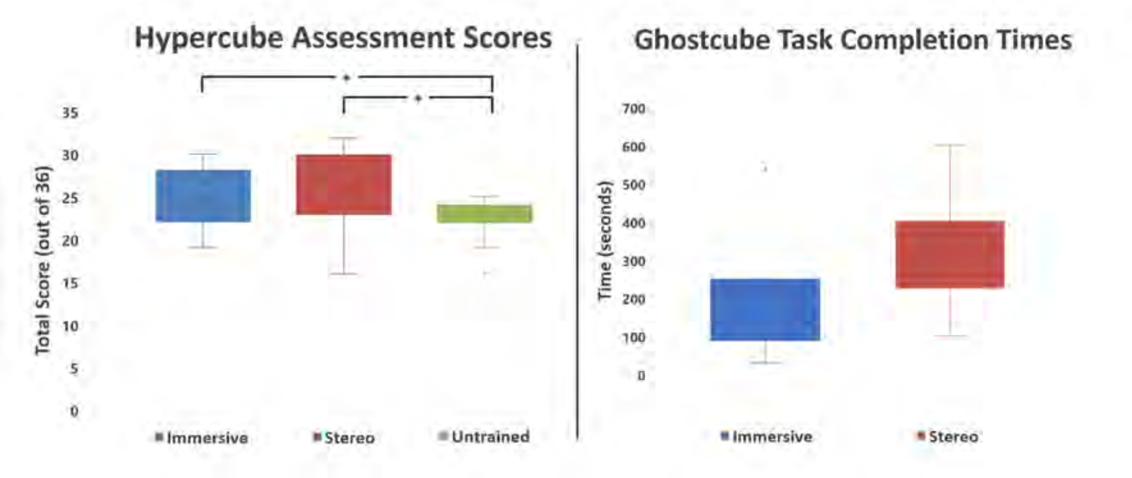


Figure 4.5: Hypercube assessment overall scores (out of 36). Significant differences were found between the immersive and untrained, and the stereo and untrained groups. Stereo scored slightly higher than immersive, but not significantly. No significant differences were found between immersive and desktop groups' ghostcube completion times (right). ** denotes statistical significance where $p < 0.05$.*

Ghostcube Matching Task. All but two participants completed the ghostcube matching task. Upon completion, participants' completion times are recorded in seconds. Shapiro-Wilk tests revealed a non-parametric distribution. A Mann-Whitney U test revealed no significant differences in the completion times between the groups (Stereo median time = 281.0, immersive median time = 220.0, $p > 0.05$). Figure 4.5 (right) presents the ghostcube results. A Mann-Whitney U test could not reveal any statistical significance ($p > 0.05$)

4.6 Discussion

The main purpose of this study was to validate the subject matter, implementation, and measurements proposed by Arnold. Overall, we are able to conclude that Arnold's measures work, although they are quite insensitive. We are able to apply our insights from the study to improve the measures for our next study.

For the self-report knowledge questionnaire, we found an expected increase in self-perceived understanding of both 4D-space and the hypercube construct for all users. Both groups reported significant increases between the beginning of the study and after watching the video. The immersive group also reported a significant increase on the third instance after the exposure for the statement “I know what a hypercube is”. We attribute this to the immersiveness of the system which provides participants with higher visualisation fidelity and perhaps more importantly, a more “intuitive” or “embodied” interaction metaphor as compared to the desktop system. This measure provided us with broad insight which is useful to give an indication of a participant’s perception of their own comprehension.

The hypercube assessment was one of our primary validation targets. While the desktop group scored an observably higher positive prediction power (see Figure 4.4 (left)), no significant differences were found between the two groups for either of positive or negative prediction power. The desktop group did however yield significantly higher accuracy than the added untrained group, but no significance was found between the untrained and the immersive groups. The trend was similar for the overall scores except both stereo and immersive scored significantly higher than the untrained group. This demonstrates that the tool is measuring comprehension, albeit with a low sensitivity which indicates that it requires improvement. These results suggest the experience of both groups was beneficial and that the assessment is measuring that effect.

While both the stereo and immersive conditions achieved significantly higher scores than the untrained group, it was only the stereo implementation that provided its users with significantly higher prediction accuracy as is reflected in the PPP/NPP results against the untrained group. The hypercube questionnaire was presented to participants on paper (2D), and the $2\frac{1}{2}$D (2 X 2D) nature of the desktop visual medium means the stereo group are seeing a hypercube representation rendered closer to what they are presented on the paper questionnaire. This shows that participants’ ability to differentiate hypercubes was significantly influenced by the presentation of the material, and their means of experience.

We had all participants attempt the ghostcube matching task in the immersive system to assess how one group’s experience would impact their performance or apparent comprehension on a different medium. The expected outcome from this was that the immersive group would be significantly faster at completing the task (given their experience condition), but while they had a slightly lower mean completion time, no significance was found. We observed this measure to be a valuable tool for measuring comprehension, but like the hypercube questionnaire, it requires reworking to generate a more sensitive tool.

In summary, we have found evidence to support our hypothesis that our implementation of Arnold's measures is indicative of subject matter comprehension. Although the measures require tweaking for improvement, the result has been a successful validation of measures. Participants were also able to successfully operate our systems for the purposes of the experiment providing validation of the implementation and the subject matter.

Limitations. We presented the video at the beginning of the experiment to provide context to participants, and also to ensure they were on approximately the same level of knowledge entering the study. In hindsight, we believe the video may have biased participants' self-report knowledge and potentially the answers on the hypercube questionnaire too due to certain misleading (but not incorrect) content we overlooked in the video.

Participants were often novice VR users, and when they began to use the system it took them some time to figure out how to look around and especially move around in the virtual environment. This could have potentially had an adverse effect on the benefits afforded by the fully immersive VR system therefore affecting the assessment results.

The time participants were given for this study was minimal. The "experience" phase of the system was only five minutes long which, while perhaps long enough to provide indications for our validation purposes, is not long enough to observe any significant gains in comprehension.

Finally, the questions we asked on the self-report knowledge questionnaire could have been more comprehensive. One participant expressed their misunderstanding of the questions asking (paraphrased) "if I understand the theory of a hypercube construct, but I have no idea how it looks or behaves, does that count as 'knowing'?". We could have asked more elaborate questions in addition.

Lessons Learned. The hypercube assessment questionnaire should be moved from a paper-based to VR-based presentation and more hypercubes should be presented of both correct and incorrect states. We also think that a time limit for each hypercube that is presented should be in place because if a learner either already possesses, or has attained the level of competency we are searching for, they should not require a long observation time to determine correctness.

The ghostcube matching task should be modified to add more hypercubes of varying difficulties. The current function required users to hold their rotation in place, then the algorithm would perform the vertex comparison and after two seconds of hold time, if the rotation was correct, a notification sound would occur and the hypercubes turned green.

To improve usability, the user should be presented with an indicator during their two second hold time to inform them of whether their rotation is correct. It became apparent the users needed more feedback, otherwise they did not know the current status of the system.

One of the most notable insights is that more time should be given to participants for the whole study. Time limits can be tailored to certain tasks, but an "experience phase" would provide interactive exposure that is not goal directed. Arnold and von Foerster proposed different ideas of how long one should need to spend manipulating the subject matter. Arnold suggested anywhere upwards of 25 hours [8] might be necessary for achievement of a deep non-verbal comprehension, and von Foerster, more optimistically, stated somewhere between 30 and 40 minutes [194]. These are long exposure times, even when delivered by modern VR technology and is potentially one of the primary factors behind their inability to conduct their intended studies in full. We should try however to provide more time for participants' experience sessions.

Having participants report on their own perceptions of their understandings is a valuable tool. Our self-assessment knowledge questionnaire did not provide enough data or allow participants to express themselves properly. An open ended interview form would be more beneficial for allowing participants to explain their thoughts and feelings regarding their experience.

Conclusion. We have gained useful insights with respect to the operation of our system, the subject matter, and in particular, the measurements proposed by Arnold. Users in each condition were able to operate their respective systems with the aim of achieving comprehension of our subject matter. We found the measures we implemented were not as sensitive to comprehension as we required although through improvements based on our observations, we can use them to further investigate our research questions. We proceed with our investigation in a new experiment with a design grounded in the philosophical thought experiment known as the knowledge argument [87]. We utilise our validated implementation, revised measurements, and lessons learned.

Chapter 5

Knowledge and Experience: A Thought Experiment Brought to Life

Contents

5.1 Introduction **54**
5.2 Experimental Design **55**
5.3 Apparatus Modifications **55**
5.4 Measures **57**
5.5 Procedure **58**
5.6 Results **58**
5.7 Discussion **64**
5.8 Conclusion **65**

In this chapter we use a thought experiment called the knowledge argument, otherwise known as Mary's Room, to guide our investigation of interaction and expertise in immersive VR environments. The lessons learned and insights gained from the previous chapter also inform the current experiment in which we utilise our improved research apparatus, and measurements. We have shown through our experiment that even for those with full theoretical knowledge of subject matter, interactive experience can provide a gain in comprehension. Furthermore, we have shown that the interactive experience facilitates the growth of comprehension in expert participants more so than in layperson participants. These results inform the discussion on learning in immersive environments, particularly the value of interaction, and interface design. Parts of this chapter are included in the following submitted work:

1. Collins, J., Regenbrecht, H., and Langlotz, T. (2019). The Uniqueness of Experience for Learning in Immersive Virtual Reality. *In International Journal of Human-Computer Studies (Submitted).* [38]

5.1 Introduction

The knowledge argument, often referred to as the Mary's room thought experiment, explores the idea that certain knowledge exists such that it is only attainable through conscious experience. It does so by proposing a scenario in which a woman named Mary lives in a black and white room with only black and white belongings for her entire life without ever seeing any colour. During her time in this black and white world she learns everything there is to know about colour. The biological, chemical, and physical aspects are all learned by Mary. Once she has learned all there is to know about colour, suddenly a colourful red apple appears in her black and white world and she experiences colour for the first time in her life. The question is, does Mary attain any new knowledge having had this experience? Does Mary gain any new form of comprehension, or in constructivist terms, does Mary's internal representation of colour alter to become more complete?

We are able to integrate the knowledge argument into our work to investigate these questions. We can use the same scenario as is described in Mary's room, only we change several variables. We propose to take a participant that possesses mathematical expertise such that they theoretically understand all there is to know about 4D space and the essence of a hypercube. This participant would be the equivalent of Mary once she has learned of all the aspects of colour. Then, just as Mary was exposed to a red apple, we would allow our experts to experience 4D space through our system and concurrently observe it. We can then ask the same question - did our mathematical expert attain any new understanding or comprehension (or for von Foerster, "begreifen") of 4D space through their experience?

We propose to extend the idea and provide further depth to our investigation by exposing theoretical subject laypeople in addition to theoretical subject experts. We then have, upon conclusion of the study, participants fitting into two categories: (1) those who have theoretical understanding plus an interactive experience of the subject matter, and (2) those who have only interactive experience without that initial understanding. This provides us with a more detailed investigation of the relationship between, and nature of, interaction and expertise.

5.2 Experimental Design

Our experiment is a single factor between-participant design with two conditions where the independent variable is participant expertise. The conditions are: 1) *theoretical subject expertise*, and 2) *no theoretical subject knowledge*. The process for each stream of participants is identical; we assess them before giving them an experience of 4D space, and then reassess. With this study, we are able to directly address our first overarching research question outlined in chapter 1 (RQ1): Does interactive involvement in IVRL environments provide unique learning experiences? From this research question and our study approach, we draw the following hypotheses:

H1. Subject experts will perform more effectively than subject laypeople on post-exposure assessments.

H2. Subject experts will perform more effectively than subject laypeople on both pre- and post-exposure assessments.

H3. Subject experts will report having increased comprehension of the subject matter.

Participants. Given our conditions are based on subject knowledge, we needed to recruit from specific populations. Our subject experts are recruited primarily from the mathematics or physics departments and our laypeople are recruited randomly from various non-mathematical disciplines. Inclusion criterion was the same as in the first study—age (between 18 and 65).

5.3 Apparatus Modifications

As described above, our independent variable is participant expertise. Our aim is to investigate interaction in immersive VR learning environments so we will only be using the fully immersive hypercube system from the first study. The stereo system is not included here as the added condition is not required for our investigation. We needed to modify our immersive system in several ways based on our lessons learned and the identified improvements for Arnold's measures.

Training Scene. To ensure our participants understand the possible actions that are available to them, we have created a "VR training scene" where participants will be presented with a regular 3D cube and will be asked to point at each side of the cube with a

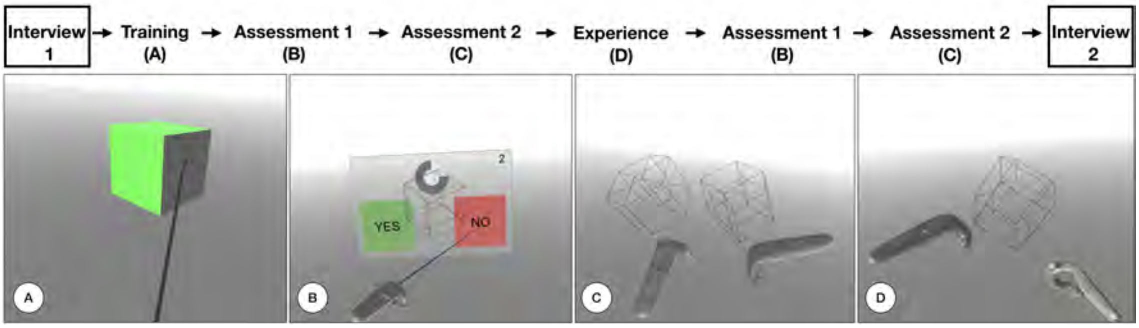


Figure 5.1: This figure demonstrates the procedure for the second study. The images (from left to right) depict the training scene, the revised hypercube assessment, the ghostcube task, and the experience task. The procedure order is shown above the images.

controller (see Figure 5.1A). This serves two purposes. Firstly, it prompts participants to look at objects from different angles hence demonstrating they are able to move around in virtual space. Secondly, it familiarises them with the interface they will use in further parts of the study.

Hypercube Assessment. In the first study we identified the need to shift the hypercube assessment from paper-based to VR-based. The VR-based assessment generates 50 hypercube forms, half are possible/correct hypercube forms, the other half are obscured rendering them impossible/incorrect hypercube forms. One at a time, hypercubes are randomly selected and presented to a user who then points at either a green 'yes' or a red 'no' button depending on whether they believe the presented hypercube is correct or incorrect respectively (see Figure 5.1B). We also implement a 10 second timer for each hypercube that users can see in the form of a circle that slowly fills up. If the timer expires, an "NA" is recorded for the answer. The timer solves two problems. The first is that it keeps the study duration to a manageable length and secondly, based on Arnold's and von Foerster's own descriptions of "deep comprehension" or "begreifen", we believe that if participants have achieved such a comprehension, they will not need longer than 10 seconds of observing a hypercube form.

Ghostcube Task. This task was already implemented in the immersive system but was modified to present participants with a fixed set of eight hypercubes of varying difficulties. The difficulties of ghostcubes are comprehensively described in chapter 7 within subsection 7.3.1. Once a user completes one matching task, the next one is presented (see Figure 5.1C). Participants have a maximum time limit of eight minutes.

5.4 Measures

The measures we utilise and implement for this experiment are detailed below.

Entry Test. We propose a test which participants take upon entry which evaluates two primary abilities: 1) their spatial reasoning and mental rotation abilities, and 2) their logical reasoning abilities. The purpose of the test is to support our categorisation of participants as experts. The entry test is developed using questions from several sources. We took eight questions of varying degrees of difficulty from the New Zealand Mensa online IQ test (*mensa.org.nz*), and six spatial rotation questions from *www.fibonicci.com* to assess participant's capability at mental rotations and geometric operations.

Demographics Questionnaire. Data are collected including age, gender, ethnicity, vision, and prior VR experience.

Semi-structured Interview. Recordings are made of a brief discussion at the beginning and at the end of the study. The purpose is to gauge participants' self-perception of their knowledge of 4D space and of the hypercube. The interviews are semi-structured, with the two primary questions being asked at the beginning: 1) Do you know what 4D space is? And 2) Do you know what a Hypercube is? And the two primary questions asked at the conclusion of the study: 1) Do you feel that your understanding of 4D space or the Hypercube construct has increased? And 2) Do you feel that you attained any measure of grasping 4D space or the Hypercube? Participants were encouraged to express their ideas. Interviews are all recorded and transcribed for later analysis. Transcription also ensures participants' anonymity.

Hypercube Assessment and Ghostcube Task. The assessment and ghostcube task are conducted in the system as described above (section 5.3). The system stores participants' answers to the hypercube assessments and completion times from the ghostcube matching tasks.

Simulator Sickness Questionnaire (SSQ). Finally, due to long exposure times in our VR system, we use a SSQ to collect data regarding user's potential simulator sickness symptoms [96].

5.5 Procedure

Upon arrival participants are greeted, introduced to the study, and given a consent form which they name and sign. They are then presented with the demographics questionnaire followed by the first interview. Participants are introduced to the VR training scene and asked to complete the training task (section 5.3). Participants then enter the first hypercube assessment followed by their first ghostcube matching task. Once participants either finish all eight cubes or run out of time, they enter the experience phase where they are asked to be seated and then have an interactive experience with a single hypercube for a time of 10 minutes (Figure 5.1D). They are told there is no specific goal other than to experience the hypercube. After they have completed the experience phase, they once again enter the hypercube assessment followed by the identical ghostcube matching task (eight minute time limit). Upon completion of the second ghostcube matching phase, participants are asked to complete the SSQ followed by a final interview asking the follow-up questions described above. Participants are compensated for their time and released.

5.6 Results

In total we had 70 participants complete the study (40 males, 30 females) with a mean age of 24.1, within a range of 18 to 60.

5.6.1 Expertise

We have three ways of determining expertise within our study:

1. by background of the participant
2. by the content of the discussion in which we query participants of their knowledge
3. according to our entry test

Our primary determinant for expertise was the background of the participant. We were only able to recruit 22 experts resulting in an unbalanced sample of 22 experts and 48 laypeople. Our secondary determinant for expertise was by the content of the pre-study interview. The transcriptions of the interviews were iterated, and if participants demonstrated theoretical knowledge of the fourth spatial dimension, they were considered an expert. When categorising participants by each of the first two determinants, we found a 72.7% correlation between the two categories.

5.6.2 Entry Test

The entry test was designed as a supporting measure of expertise. There were two components to the test as described above: (1) mental rotation and (1) logic components. The overall entry test scores are shown in Figure 5.2.

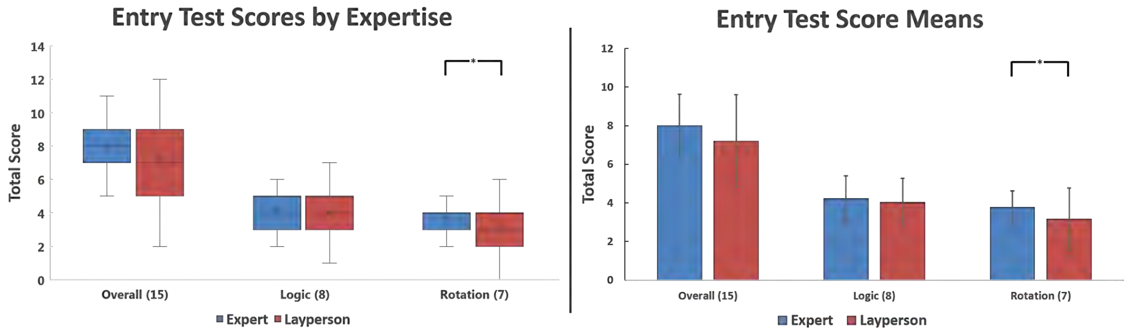


Figure 5.2: Presents the entry test scores by participant categories. Scores are presented for all questions (Overall), the logic questions (Logic), and the mental rotation questions (Rotation). There were significant differences found between expert and laypeople groups for the rotation portion of the entry test. ** denotes statistical significance where* $p < 0.05$*.*

When applying Shapiro-Wilk tests to the entry test data, the overall scores returned a normal distribution, but when split into logic and rotation scores, distributions are not normal. The overall score on the entry test was out of a total of 15 questions. The expert group yield a higher mean score (*M=8.0, S.D=1.62*) than the layperson group (*M=7.21, S.D=2.40*) but an unpaired t-test revealed no statistical significance ($p > 0.05$).

The logic portion of the entry test was out of a total of eight questions. The expert group (*M=4.23, S.D=1.17*) scored similarly to the layperson group (*M=4.04, S.D=1.22*) where an unpaired t-test found no statistical significance ($p > 0.05$). The rotation portion scores were out of a total of seven questions. Experts scored a higher mean score (*M=3.77, S.D=0.85*) than the laypeople (*M=3.17, S.D=1.60*). An unpaired t-test revealed a statistically significant difference for the rotation scores ($p < 0.05$).

5.6.3 Hypercube Assessments

We apply the same analysis of positive and negative prediction power as in the first study (see section 4.5 - Hypercube assessment questionnaire). Shapiro-Wilk tests revealed the PPP results to have a parametric distribution, while the NPP results have a non-parametric distribution. PPP/NPP values achieved for all participants increased between the first (*M=0.58/0.63, S.D=0.11/0.12*) and the second (*M=0.67/0.81, S.D=0.12/0.17*) hypercube

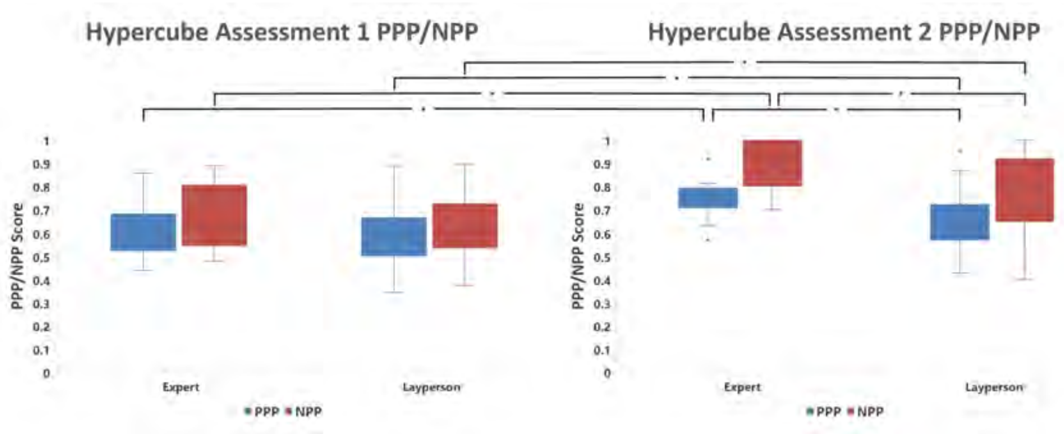


Figure 5.3: Hypercube Assessment results between expert and layperson groups for the first and second assessments broken down into PPP (blue) and NPP (red). Both groups made significant improvements (especially for NPP). The expert group's gain was also significantly higher than the layperson group. ** denotes statistical significance where $p < 0.01$.*

assessments. Figure 5.3 shows a graph of the overall PPP/NPP means for the first and second Hypercube assessment. Significance was found for both PPP (*t-test*, $p < 0.01$) and NPP (first assessment median = 0.61, second assessment median = 0.87, $p < 0.01$ *(Wilcoxon)*). Significance testing between expert and layperson groups revealed no significance for either PPP (layperson mean/s.d = 0.57/0.11, expert mean/s.d = 0.61/0.11, $p > 0.05$ *(t-test)*) or NPP (layperson median = 0.57, expert median = 0.61, $p > 0.05$ *(Man-Whitney U)*). Significance tests for the second assessment revealed the expert group scored higher than the laypersons group on both PPP (layperson mean/s.d = 0.64/0.12, expert mean/s.d = 0.74/0.07, $p < 0.01$ *(t-test)*) and NPP (first assessment median = 0.61, second assessment median = 0.87, $p < 0.01$ *(Wilcoxon)*) scores.

As in the first study, the prediction analysis is primarily indicative of participants' accuracy so we need to analyse the overall scores as well (see Figure 5.4). Participants' score data was found to be normally distributed for both assessments with a Shapiro-Wilk test. Once again, the scores reflect the PPP/NPP results. The expert group achieved similar scores (M=*31.86*, $S.D$=*5.55*) to the layperson group (M=*29.52*, $S.D$=*5.03*) for the first assessment where an unpaired t-test could not find any significance ($p > 0.05$). For the second assessment, the expert group scored an observably higher score (M=*39.73*, $S.D$=*3.49*) than the layperson group (M=*33.75*, $S.D$=*6.33*) and this time an unpaired t-test found statistical significance ($p < 0.01$). Significance was also found using paired t-tests within groups, between assessments, for both groups ($p < 0.01$).

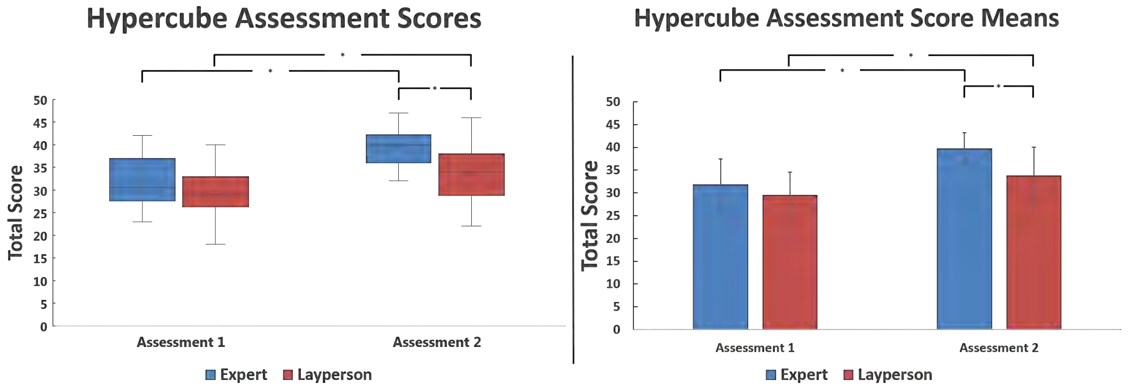


Figure 5.4: Hypercube assessment scores out of 50 between expert and layperson groups for the first and second assessments. As is reflected in the PPP/NPP results, both groups made significant improvements between their first and second assessments. Additionally, the expert group's improvement was significantly higher than the layperson group. * *denotes statistical significance where* $p < 0.01$.

5.6.4 Ghostcube Matching Task

Ghostcube results are analysed in two ways: 1) number of cubes solved and 2) time taken for each cube solution. There were a total of eight cubes to be solved during the ghostcube matching task. All data pertinent to the number of cubes solved is non-parametric so either Wilcoxon or Mann-Whitney U tests are used for significance testing. For all participants, there were no significant differences found ($p > 0.05$) for the number of hypercubes solved between the first (median no. cubes = 3.0) and second (median no. cubes = 3.0) task instances.

Significance tests between expertise groups for the first ghostcube task revealed no significant differences ($p > 0.05$) in the number of cubes solved (expert group: $M=3.05$, $S.D=0.57$; layperson group: $M=2.85$, $S.D=1.17$). Significant differences were found ($p < 0.05$) for the second ghostcube task between the expert group ($M=3.82$, $S.D=1.5$) and the layperson group ($M=2.94$, $S.D=1.02$). When testing within groups from the first to second task instance, significance was found only for the expert group ($p < 0.05$). See the number of solutions presented in Figure 5.5.

More detail is revealed in the completion time of each ghostcube. The solution frequencies for each ghostcube are presented in Figure 5.6, and show a sharp drop in solutions for ghostcube (GC) four. The number of solutions (7, and 12 for the first and second task respectively) does not provide us with enough completion time data to perform any robust statistical analyses. For that reason, we only consider the first three target hypercubes (1,

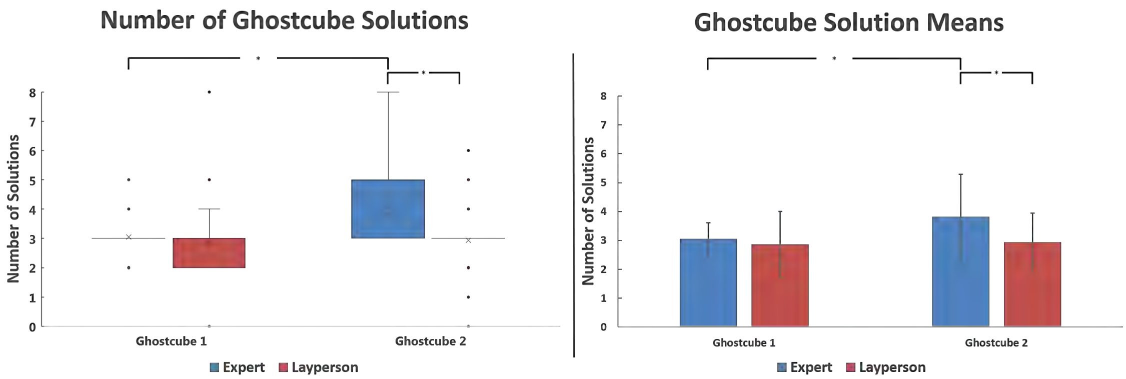


Figure 5.5: Demonstrates the number of cubes solved per assessment between the expert and layperson groups. Significant improvement is found between the first and second ghostcube task for the expert group only, and in the second ghostcube task, the expert group perform significantly better than the layperson group. ** denotes statistical significance where p < 0.05.*

2, 3) in the matching task for our numerical analyses.

Completion time data is shown in Figure 5.7. Median completion times dropped significantly between the first (19.45, 34.65, and 96.45) and second (19.00, 24.95, and 48.30) ghostcube tasks for the second and third target cubes ($p < 0.01$, *Wilcoxon*), but not for the first cube where the mean rose slightly (first GC mean = 37.37, second GC mean = 44.44, $p > 0.05$, *Wilcoxon*).

The expert group's median completion times reduced for all three cubes between the

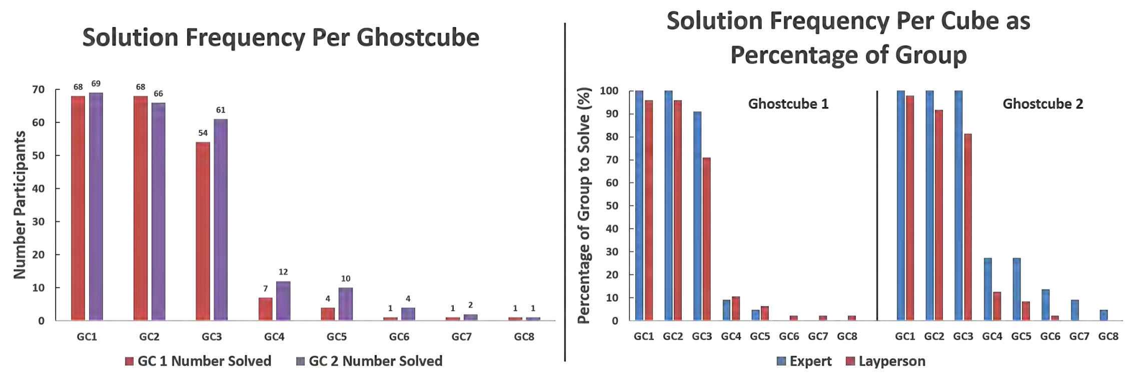


Figure 5.6: The frequencies of ghostcube solutions between the first and second ghostcube task exposure (left) shows slight improvement from the first to the second. The percentage of the expert and layperson groups to solve each ghostcube (right). A larger proportion of experts were able to improve their solution frequencies to a greater degree than laypeople participants (right).

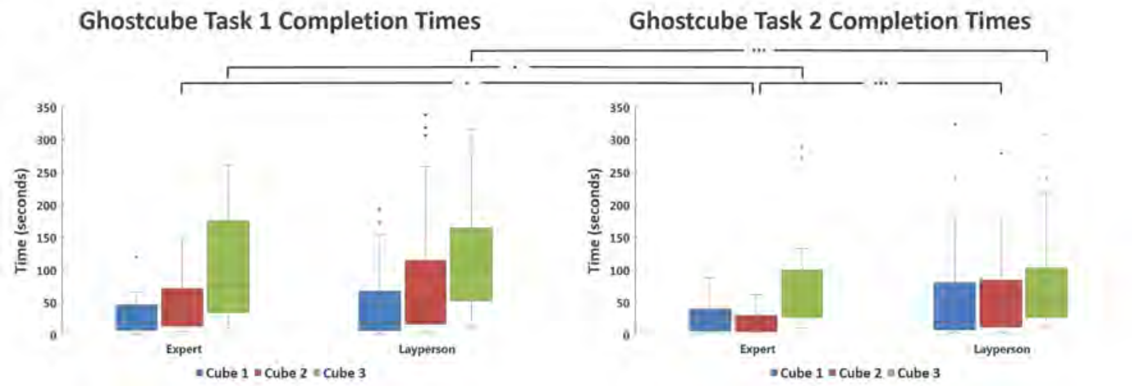


Figure 5.7: Ghostcube task completion times for the first three target ghostcubes. The difficulty of the cubes is reflected in the completion times. Similarly to the hypercube assessment, both groups mostly improved, namely for the 2nd and 3rd cubes, but the expert group have the most significant completion time reductions. **/*** denotes statistical significance where p < 0.05/0.01.*

first (19.95, 24.15, and 76.50) to the second (15.95, 9.05, and 40.95) ghostcube tasks where Wilcoxon tests revealed significant reductions for the 2nd and 3rd cubes ($p < 0.05$) but not for the 1st ($p > 0.05$). The same test revealed a similar trend for the layperson group with means mostly dropping from the first (18.50, 41.95, and 99.20) to the second (19.30, 27.80, and 52.20) task, with significant reductions found only for the 3rd hypercube ($p < 0.01$) but not for the 1st or 2nd ($p > 0.05$).

The expert group's completion time for the second cube in the second ghostcube task (17.88) was significantly less than the layperson group (51.90) when applying a (unpaired) Mann Whitney U significance test ($p < 0.01$) but not between the first or third cubes ($p > 0.05$).

5.6.5 Simulator Sickness Questionnaire

The questionnaire can be split into two symptom components: 1) nausea and 2) oculo-motor [21]. Participants reported a mean nausea total of 1.8 with a S.D of 1.87. Oculo-motor ratings were higher with a mean of 2.96 and a S.D of 2.38. Kennedy et al. provided a categorisation of SSQ scores [95] where any score less than five is considered as "negligible symptoms". A small number of participants gave higher reports, particularly in the oculo-motor dimension, with oculo-motor totals of 10 and 12. In the categorisation from Kennedy et al., scores between 10 and 15 are considered as significant symptoms, however, the categorisation table is not designed to evaluate individual SSQ reports but rather the mean report as an evaluation of the system. The mean total of all SSQ scores was 4.75 with a S.D of 3.83.

Each item on the questionnaire was rated from zero to three where the options were 'None', 'Slight', 'Moderate', and 'Severe' (see Appendix B). The highest ratings were placed on "eye strain" and "fullness of the head", with mean ratings of 0.70 and 0.64, respectively.

5.7 Discussion

We have shown through our experiment that even for those with full theoretical knowledge of subject matter, interactive experience can provide a gain in comprehension. Furthermore, we have shown that the interactive experience facilitates the growth of comprehension in experts more so than in laypeople.

Our first hypothesis (H1) was that expert participants would perform more effectively between the first and second assessments (after having had the experience) despite their theoretical expertise. This hypothesis is supported in our findings for both the hypercube assessment (visual), and the ghostcube task (interactive). The layperson participant group also improved significantly which is an expected outcome, but when testing the second assessment differences between expert and layperson groups, the expert group's performance was significantly better. This is the case for the hypercube assessment, and the number of cubes solved between the first and the second ghostcube assessments. This also contributes to our second hypothesis (H2) that experts would perform more effectively than layperson participants.

The expert group had similar completion times in the first ghostcube task, though for the second, they had a significantly improved time for the second ghostcube. This supports our second hypothesis. The first hypothesis is supported through the improvements demonstrated between the first and second ghostcube tasks. Figure 5.6 demonstrates the improvements made by a larger proportion of the expert group. More were able to pass the fourth, more difficult ghostcube.

The overall mean time taken to solve the first target ghostcube rose from 37.37 to 44.44 seconds. This was also the easiest of the target cubes to solve as is demonstrated by the solution frequency graphs. The rise is likely due to the starting position of (nervous) participants being very close to the solution position. This time actually dropped slightly for the expert group from 27.78 to 26.17 seconds, and for the layperson group, rose from 41.95 to 52.99 seconds which is the reason for the overall rise again supporting our second hypothesis.

Our 3rd hypothesis (H3) is that experts would report on achieving an improved comprehension from having had the experience. The general consensus among the expert participants was similar to that of laypeople. They reported the experience as useful and while they often reported on gaining something, it was intangible or, for them, indescribable. This could be hinting towards Arnold's "deep non-verbal comprehension" but given the lack of complete competency in the tasks, it is unlikely this was totally achieved. Participants have likely gained what Arnold referred to as "partial mastery" [8]. Experts made statements such as "... I think being able to see it and manipulate it yourself really helps ...". There were reports of still using trial and error instead of knowing what to do: "... I had no idea how to do it so I just had to try several axes slowly and see which one did it and which did not ...". Experts also mentioned it could be useful as a tool for learning how projections work. Hypothesis H3 can not be supported through statistical methods, and by iterating the interview transcriptions and highlighting answers to pertinent questions, we can not say that we found evidence of complete certainty from any participant with regards to their comprehension of the subject matter. This is expected given the limited time of exposure.

Nonetheless, this highlights the idea that there is a difference in certainty of participants between the beginning of the study and the end, mainly for experts. In some cases there was an explicit expression of confidence in their comprehension of the subject matter upon entering the study. However, after their exposure, and especially when no major increase in success was made on the second ghostcube task, confidence in their comprehension of 4D space appeared to abate. Their theoretical understanding had not changed, and they were no less confident in their mathematical abilities, but there was acknowledgement that there is more to it.

5.8 Conclusion

We have been able to conclude that the red apple would provide Mary, in her black and white world, with a new or improved form of comprehension. To return to constructivist theory, knowledge is constructed based on our experiences and interactions with the world and this forms our resulting internal representations of concepts and ideas. We posited that mathematicians or physicists for instance would have an internal representation of 4D space rich with theoretical input, just as Mary's internal representation of colour was. We have been able to demonstrate that providing somebody with an immersive interactive experience of a construct which they have a rich or seemingly complete internal representation of, increases their comprehension of that construct. Constructivist learning can benefit from Virtual Reality learning and Virtual Reality Learning can benefit from constructivist learning. Within our scope of design, implementation, and study so far, namely abstract geometric comprehension and “begreifen”, we have shown this reciprocal relationship identifying the complexities of constructivist learning with and in VR.

The challenges associated with assessing comprehension became evident in our first study (chapter 4) by presenting participants with 2D images of complex 4D geometries however the use of VR for learning and assessment as such is working, and so are the measurements. Study two could show that our experts gained a learning experience by manipulating 4D geometry—the immersive, interactive experience of a projected 4D cube improves “begreifen”. In addition, we could show that not only experts but non-experts benefit from such an experience, and that experts benefit more than non-experts. Such a finding has implications for (a) the design of interactive VR learning environments and consequently the HCI and VR research communities and (b) for computer supported learning and learning in general. It raises the question of how much expertise is needed and sufficient for effective interaction with the subject matter but it answers the question whether both expertise and interaction are needed. They are.

In that sense, Arnold and von Foerster triggered the right questions about “seeing and doing”, even if they could not really tackle them appropriately because of the lack of technological advancement. Nowadays, we are in a position to actually put constructivist learning practices into an immersive experience and with this open up a wide field of research and practice on future forms of learning.

There is much room for future work in this space to focus on interesting and important questions around: the influence of individual differences on learning style and efficacy, the underlying questions about what efficacy actually means and how it can be measured, the different and novel ways we might learn with VR in the future, or the range and areas of learning applications VR might be suitable for. How much and what forms of understanding and experience are needed for Mary to deeply comprehend a multi-coloured, multi-dimensional apple?

This study concludes the investigative branch specifically focused on interaction and expertise in immersive VR learning environments. We continue in this thesis with an investigation of the occurrence of significant instances of learning called Aha! moments, or moments of insight. This is a second branch to stem from Arnold’s work, but is more specifically discussed by von Foerster [194]. We also investigate a new method for measuring individuals' performances and emotional engagement with their learning experiences in IVRL environments.

Chapter 6

The Search for Insight: Presence and Aha! moments

Contents

6.1 Introduction **70**

6.2 Emotional Response in VR **72**

6.3 Chasing the Eureka! - Usability Study **75**

6.4 Chasing the Eureka! - Presence and Insight Learning **79**

6.5 Summary **86**

We have been able to provide learning experiences within immersive VR learning environments, and have shown that interactive experience in IVRL environments provides value towards those learning experiences. This is largely in line with Arnold's idea of constructivist learning, though von Foerster proposed that learners would, with enough exposure time, experience a moment of insight and would exclaim "AHA!". Aha! moments are sudden moments of insight characterised as a psychological phenomenon and it has an existing body of research behind it. The proceeding investigation explores whether we can lead learners to Aha! moments, and whether we are able to measure such moments. We also investigate the sense of presence in IVRL environments and its relationship with learning experiences. In this study, we employ the physiological measurement of emotional response due to the implicit relationship emotional engagement shares with both the presence and learning aspects of IVRL environments. This chapter describes a two part study. First, we give an introduction to the concepts and their relationships. We then discuss our approach to the measurement and structured analysis of IVRL experiences. We report on a preliminary usability experiment with the results and finally, we introduce our main experiment. Parts of this chapter are included in the following work:

1. Collins, J., Regenbrecht, H., Langlotz, T., Can, Y. S., Ersoy, C., and Butson, R. (2019). Measuring Emotional Response in Virtual Environments: A Methodology Exemplified in a Learning Context. *In International Symposium for Mixed and Augmented Reality '19 (Accepted).* [39]

6.1 Introduction

One of the most well-known stories of an Aha! moment is that of Archimedes in a bathtub, while pondering over a problem, exclaiming Eureka! In a moment of sudden comprehension, all becomes clear and he solves his problem. This is the moment of insight referred to by von Foerster in his work following on from Arnold [194]. The general idea, linking again back to constructivism, is that only if a learner is able to engage in an interactive experience with subject matter, can they attain von Foerster's "begreifen" or Arnold's "deep comprehension". In the case of Arnold's proposal, it was the presentation of a visually interactive application that would facilitate this process of learning. Von Foerster went one step further and identified the problem-based environment presented to learners, and suggested they will experience this knowledge gain through Aha! moments, indicative of "begreifen".

Learning by Aha! moments is often called insight learning and is described as the sudden discovery of "new relationships within ... prior knowledge as a result of reasoning or problem solving processes" [9]. A further more elaborate definition defines insight as "experience during or subsequent to problem-solving attempts, in which problem-related content comes to mind with sudden ease and provides a feeling of pleasure, the belief that the solution is true, and confidence in this belief" [189]. These moments have been studied by psychologists for nearly a century using mostly behavioural methods [104]. Danek et al. conducted an investigation into the Aha! experience to determine if false solutions can result from such insights [44]. Three key components of insight are identified in their work and include pleasure, suddenness, and certainty. They utilise in-the-moment questionnaires to assess participants' Aha! experiences. Physiological measures have also been applied to Aha! moments such as electroencephalogram (EEG), and functional magnetic resonance imaging (fMRI) in an attempt to isolate the cognitive processes which take place during moments of insight [104, 158]. They discovered emerging patterns in brain states preceding and during the Aha! moments suggesting intervention opportunities to facilitate insight learning [104].

There also seems to be a recurring affective component in the discussion of the Aha! experience. An inherent element of insight learning is the capability to achieve insight through creative thought, rather it relies more on creativity than traditional analytic procedural thinking, and typically results in higher levels of engagement [136]. The VR problem-solving environment we have implemented and used throughout this work provides the platform for engaged, insight-oriented thinking to take place.

As we have discussed previously, one of the defining elements of VR is the sense of presence. Involvement as a core indicator of presence is well established. Schubert et al. in their factor analysis of presence in VR identified three predominant factors: spatial, involvement, and realism [161]. They showed that spatial, and in particular involvement, are powerful catalysts for the emergence of presence. Involvement is described as "a psychological state experienced as a consequence of focusing one's energy and attention on a coherent set of stimuli or meaningfully related activities and events" [205]. This is coherent with the form of engagement discussed above as a requirement for insight learning. It follows then that a users' sense of presence shares an inherent relationship with the kind of learning we are exploring. The phenomenon of insight learning and presence are the primary focuses of the work in the second branch of investigation in this thesis. Given the importance of emotional engagement in learning and VR experiences, we include this in our investigation.

In the current study, we investigate three primary research questions:

RQ1: Can we lead users towards and measure Aha! moments in IVRL environments?

RQ2: What is the effect of presence on users' learning experiences, particularly the occurrence of insight?

RQ3: Are there measurable relationships between physiological measures of emotional response and learning experiences in IVRL environments?

We utilise the same research apparatuses that have been described and used for the studies in previous chapters in an experiment where participants interact with a hypercube uninterrupted for one hour. The purpose of the extended exposure time is to allow participants to engage with the problem-solving elements of the environment, develop strategies, and form a more detailed internal representation. This will, in the ideal case, lead to quality Aha! moments which users can report on. This investigation involves a complex set of relationships. Presence and Aha! moments exist as independent phenomena, however they share and rely upon the emotional states of the user. So we can provide more depth to our investigation, we introduce the measurement and structured analysis of three separate data dimensions. With our analysis lies our measures of presence and Aha! moments in addition to the measure of emotional response via physiological signals. This process still allows for independent measurement and analysis of phenomena according to traditional approaches, but also facilitates analyses of the complex relationships that emerge.

6.2 Emotional Response in VR

We propose a methodology for measuring and analysing separate data dimensions to provide an informed perspective on the complex relationships comprising a users' experiences in IVRL environments. We have emphasised the importance of engagement for IVRL environments, so at the heart of our methodology lies the physiological measurement representative of emotional response. Traditional measures of presence and Aha! moments, our phenomena of interest in the current study, are also applied within the methodology. We isolate three separate data dimensions, each of which provide unique insight into the reality of users' experiences in IVRL environments.

Below we describe the primary measures to be utilised within our current study in terms of the data dimensions. A more detailed description of the procedure and direct application of the measures will be discussed at a later time together with the details of the main experiment.

6.2.1 Sensed Dimension: Emotional Response

A user's sensed reality is an objective measure of subjective states. Individuals behave, act out, and react differently in various situations and environments and exhibit different emotional responses. Physiological measures sense a user's subconscious state of being. Electrodermal activity (EDA), Heart Rate (HR), Brain Activity (electroencephalogram (EEG), functional magnetic resonance imaging (fMRI)), and Heart Activity (electrocardiography (ECG)) are all examples of physiological measures applied in different contexts.

We established earlier that both the sense of presence in VEs and insight learning each depend on engagement. Therefore, it is imperative that a user achieves an engaged affective state in an IVRL environment. Measuring emotional activity by monitoring physiological states is not a novel idea [99, 101, 139]. A common measure of emotional states used within various contexts is that of electrodermal activity (EDA), otherwise known as skin conductance. This measures the sympathetic nervous system (SNS) and is a sensitive index of sympathetic arousal which is integrated with emotional and cognitive states [40]. This measure has prior application in VR environments in the context of anxiety and therapy treatment [49, 55, 199, 200], emotional replication studies [56], behavioural studies [63], and presence [48, 123, 124, 134].

In the sensed reality component of our current study, we apply physiological measurement to assess users' emotional states throughout the duration of their learning experience in our IVRL environment. We use an Empatica E4 wristband device* to primarily measure EDA and HR data. The Empatica E4 wristband measures EDA, HR, peripheral skin temperature, accelerometer data, and blood volume pulse (BVP). Application of this measure and analysis of the resulting data are discussed at a later time.

6.2.2 Reported Dimension: Presence and Aha! moments

A user's reported reality is subjective by nature, and measures a user's own perception of their experiences. Questionnaires and interviews are common examples of measurement apparatuses applied in the psychological dimension.

As discussed in chapter 2, the predominant form of measuring presence to date is by self-report questionnaires. Many efforts have been made to develop questionnaires for assessing the sense of presence of a user in a virtual environment [11, 112, 117, 161, 174, 205]. Despite emerging evidence that the questionnaire is not always a representative measure of the sense of presence [170, 192], and despite the efforts of alternative developments [124], they are still arguably the most prominently used measure of presence. For our current study we utilise the Igroup Presence Questionnaire (IPQ) developed by Schubert et al. [161]. We are particularly interested in users' engagement in our IVRL environment which we find to be potentially analogous with involvement which was identified as a key factor of presence as measured on the IPQ.

The moment of insight or Aha! experiences are a subjective psychological phenomenon. A commonly employed method for measuring insight is by self-report, often in the moment. Recent work by Danek et al. [44] measured insights by asking participants to notify experimenters upon comprehension of task solutions. Users would press a button on a keyboard at which time several questions would be presented to them assessing the quality and magnitude of users' Aha! experiences with respect to key factors previously identified

*www.empatica.com/research/e4/

in the literature. We employ a similar approach in our work, though it is tailored to our IVRL environment. We are not investigating the factors that comprise the Aha! moment, rather we are interested in the occurrence of perceived Aha! moments, when they occur throughout exposure to our IVRL environment, and whether they are indicative of the "begreifen" von Foerster refers to.

There is a further key reason we do not present participants with questions during the exposure in the virtual environment. The importance of engagement has been an emphasis of this work and an issue with mid-exposure questionnaires is the resulting break in that engagement. Pulling a participant out of a VR environment is detrimental to their sense of presence in that environment, and it forces a stop to any (immediate) internal processes regarding the task at hand. This issue has been addressed by presenting questionnaires within the virtual environment to facilitate the retention of presence which has found some success [61], although the sense of presence is arguably still interrupted. Furthermore, mid-exposure questionnaires do not solve the problem of breaking engagement with problem-solving processes. In our case, the environment depends on an organic flow of engagement. Rather than asking questions about specific Aha! moments during the exposure time, we opt to include questions regarding the participants' general Aha! experiences in a post-exposure interview.

6.2.3 Observed Dimension: Environment, Performance

A user's observed reality is an objective dimension capturing the interplay between a user and the environment, and any outcome of that interaction. Video and audio recording devices are commonly used together with devices capable of storing data which is consequential of a user's interactions with the environment (i.e. performance times).

We employ multiple forms of observational measures. Our most important observational measure is users' achievements and successes within the environment. The system participants are exposed to is the same principal apparatuses described in chapter 3, but tailored for this experiment (described in more detail at a later time). When participants achieve a task, the system records data relevant to the achievement such as the time, and which task was achieved. We also make different kinds of video recordings depending on the experimental condition (more detail later) where the users' hand movements are captured. We also make audio recordings of the entire exposure session to capture anything the user says throughout.

6.3 Chasing the Eureka! - Usability Study

Presence is a major component of the research questions we outlined earlier. Although only one question (RQ2) makes direct mention of presence, it is again a defining component of VR and therefore has an implicit relationship with the entire user experience. Given that we are measuring for the effect of presence, we must control for it which we accomplish through immersive factors. Increased immersion has been understood for some time to correlate with the sense of presence in virtual environments [24]. In chapter 2 we discussed the differences between immersion and presence, and immersion was defined as the extent to which a system is inclusive, extensive, surrounding, and vivid [176]. We use the two systems we have implemented and utilised so far throughout this work.

The first is the desktop stereo system on which users interact with six dials as input for manipulating the hypercube according to Arnold's original system description. The second system is the fully immersive application where users interact with two HTC Vive controllers to manipulate the hypercube. We were able to validate both systems in terms of operation as an interface to 4D constructs (see chapter 4). An issue becomes apparent when we plan the study design of our main study around these two systems because they implement different interactive mediums.

One of our observational measures is participants' achievements so we want to minimise confounding factors potentially effecting performance. It is possible that usability differences between the two interactive mediums of our systems emerge such that they effect those performance outcomes in our main experiment. Therefore, we run a preliminary usability study to identify potential differences between the two interfaces. The results can be used to inform the main experiment. Each system provides a different set of immersive factors according to the definition of immersion.

Immersiveness. When using the desktop stereo system, users are wearing 3D stereoscopic glasses for viewing content in 3D. In terms of inclusiveness, the external reality is not shut out of the users' view, and they can look around freely in their real environment. The visual modality is the only sense catered to on the desktop system resulting in one dimension of modality extension. In terms of providing a surrounding environment, the stereo environment is viewed from a fixed perspective by the user so they can not look or move around within the virtual environment. The vividness of the desktop system is measured by the quality of the visual medium which for us is a 1920x1080 resolution monitor running at 120Hz (60Hz per eye).

A HMD is used as the visual medium in the immersive VR system which provides an inclusive quality that shuts out most of the real environment. Both visual and auditory modalities are catered for within the immersive system providing two dimensions of modality extension. The fully immersive virtual environment is fully surrounding, and the user is able to look and move around in that environment in six degrees of freedom. In terms of vividness, the HMD provides a resolution of 2160x1200 updating at a 90Hz refresh rate, and users wear Steelseries Siberia 150 over-ear headphones.

Given the breakdown of each system in terms of their immersive qualities, it is clear the immersive VR system provides an objectively higher degree of immersion.

6.3.1 Experimental Design

We use a single factor within-subject design with two degrees of freedom. The factor is the level of immersion with the two levels being 1) desktop stereoscopic 3D system with the six dial interface shown in Figure 6.1 (left) and 2) our fully immersive VR system shown in Figure 6.1 (right). Both systems are described in detail in chapter 3.

Participants. An A priori power analysis was carried out using G*Power (v3.1.9.2). With a power of 0.85, an effect size of 0.5 (medium effect size)[35], and an error of 0.05, we calculated the recommended sample size of 31. Due to our within-subject design we are required to randomly assign the participant's starting condition, therefore we require an even number of participants. The random assignment was generated on random.org.

6.3.2 Task and Measurements

The study is designed to run between eight and twelve minutes in length. Participants will attempt to complete a task using two different systems. The task is exactly the same for both systems. The task is based on the ghostcube matching task demonstrated in chapter 4 and chapter 5, but this time participants will be tasked to solve two ghostcubes. They are given four minutes to complete both ghostcubes.

Participant demographics are collected by questionnaire and collects data on age, gender, vision status, prior VR experience, and susceptibility to motion sickness. Timing data from the participants' completion of cubes is stored to a text file. Finally, the usability of each interactive medium is evaluated using the 10-item System Usability Scale (SUS) questionnaire [25].

Figure 6.1: Usability study conditions with a user demonstrating the stereo condition (left) and the immersive condition (right)

6.3.3 Procedure

Subjects are greeted and presented with the study information sheet, consent form, and the demographics form. They are introduced to their first pre-randomised system and are given 4 minutes to solve two ghostcubes. When either the time runs out, or they are able to match the two ghostcubes, they complete the SUS questionnaire. They are then introduced to their second system where they complete the same task as previously. They are given the same time for the second task after which they will complete the SUS questionnaire a second time. Participants are thanked, compensated, and released.

6.3.4 Results

In total, we were able to recruit 36 participants (22 male, 14 female) with a mean age of 23.1 ranging between 18 and 48. Participants are recruited from staff and students at the University of Otago. All 36 participants completed the experiment and therefore we ended up with a total of 72 completed SUS usability questionnaires (36 per condition).

Usability. The SUS questionnaire poses 10 questions with answers on a Likert-like scale from one to five. The final result of the SUS questionnaire is a single number ranging from

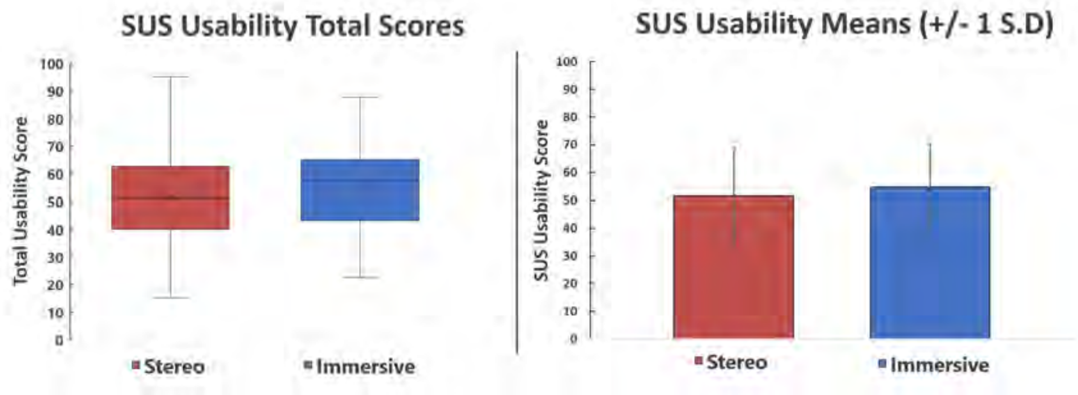


Figure 6.2: Mean SUS Questionnaire Results for Fully Immersive and Stereo Conditions.

0 to 100 where 0 is the lowest score and 100 is the highest. This represents a composite measure of the overall usability. Overall usability scores are presented in Figure 6.2. SUS data revealed a normal distribution with a Shapiro-Wilk test. The fully immersive system yielded a slightly higher mean SUS score (*M=54.65, S.D=15.78*) than the desktop stereo system (*M=51.60, S.D=17.44*), though a paired t-test did not find any significance ($p > 0.05$).

We decided to test for any bias effects based on the participants' starting conditions. Unpaired t-tests were applied to immersive and stereo first SUS results against immersive and stereo second SUS results respectively where no significance was found ($p > 0.05$).

We need to interpret the SUS usability scores to give us an idea of what they actually mean. Bangor et al. were able to present a one item adjective rating scale which correlated highly (0.822) with the overall SUS scores [10]. The adjective scale was marked on a 7-point Likert-like scale, where items directly mapped to the adjectives: 1) Worst Imaginable, 2) Awful, 3) Poor, 4) OK, 5) Good, 6) Excellent, and 7) Best Imaginable. The question asked was "Overall, I would rate the user-friendliness of this product as:". The SUS scores reported for our interfaces correlate almost directly to the "OK" adjective rating. This rating is sufficient for the purpose of our research apparatus and we would expect the difficult nature of the task to influence the SUS reports in this way. We are also mainly interested in the comparison between the two interfaces rather than individual product usability ratings.

Task Completions. On the immersive system, all 36 participants solved the first ghostcube, and only two solved the second ghostcube while on the stereo system, 32 participants solved the first ghostcube, and zero participants solved the second ghostcube.

Shapiro-Wilk tests revealed a non-parametric distribution on completion times. The stereo system had a higher completion time for the first cube (*M=86.85, S.D=58.35*) than the immersive system (*M=53.12, S.D=50.94*) where a paired Wilcoxon found significant differences ($p < 0.05$).

6.3.5 Discussion

The results of the preliminary usability study reveal no significant differences in usability between the two systems as perceived by the participants. In fact, the mean usability scores are very similar but slightly favoured towards the fully immersive system. This tendency could be due to the novelty factor of immersive VR as the report on the demographics questionnaire of prior VR experience was quite low. Despite the faster completion time of the immersive system for the first cube, participants' preference was still unclear, though some participants expressed after the study that they even preferred the stereo interface because of the more fine grained control it provided. They made particular mention of being able to isolate all inputs and move them one by one, and thought it was advantageous.

The results suggest that, in terms of interface usability, we expect no major differences will have any confounding effect on presence. We have been able to establish that our apparatuses are both usable according to the adjective scale described above. We do need to be aware though of the performance differences measured in this preliminary study, and observe if they may carry over to our main experiment. With these results though, we are confident we can move forward with our study measuring presence, insight learning, and emotional response.

6.4 Chasing the Eureka! - Presence and Insight Learning

With the experience we have so far provided in previous experiments, learners have only had a maximum of 10 minutes of uninterrupted interaction with a hypercube. This was enough for us to identify the value of interaction and experience, though now we shift our focus towards presence and insight learning. Learners often take time to engage with problem-solving processes and construct solution strategies, especially when the subject matter is more difficult in nature. We want to facilitate the users' state of engagement and allow them to solve many problems and while doing so, gaining comprehension of the subject matter hopefully leading them to the moments of insight ("begreifen") referred to by von Foerster.

To answer the research questions defined earlier, we conduct a study where we have participants spend an hour in a VR problem-based learning environment in which they are manipulating a hypercube the whole time. Their primary goal is to experience and gain competency over manipulation of the hypercube. Many participants that take part in VR studies have often never experienced VR environments before. The longer exposure time provides the advantage of mitigating for any "wow effects" caused by the novelty of users' experiences. A consequence of this is that users' self-report responses and subjective reflections are more representative of a usual experience, rather than a short first time experience which is often biased by enjoyment. Given our three research questions and our measures of presence, Aha! moments, and emotional response, we formulate the following hypotheses:

H1: *We can lead users to a reported Aha! moment during their time using the system*

H2: *Users' reported Aha! moments will be more frequent with a higher sense of presence in the VR environment*

H3: *Users' reported Aha! moments will lead to more achievements*

H4: *Emotional Response measures will correlate with presence, Aha! moments, and achievements*

The remainder of this chapter describes the design, measures, and procedure of our main experiment.

6.4.1 Experimental Design

A single factor within-subject design with two conditions is used where the independent variable is immersiveness. The two interactive hypercube systems described in chapter 3 and evaluated in the previous usability study, provide our two conditions: 1) Desktop 3D stereoscopic system where the user wears stereoscopic glasses and uses analogue dials for interaction, and 2) a fully immersive HMD-based VR system with controllers for interaction. Participants' first condition is pre-randomised.

Participants. An A priori power analysis was carried out using G*Power (v3.1.9.2). With a power of 0.85, an expected large effect size of 0.6, and an error of 0.05, we calculated the recommended sample size of 22. The only inclusion criterion for participant recruitment is age (between 18 and 65). We recruited primarily from the student and staff populations within the university.

6.4.2 Apparatus and Measurements

Both systems, desktop stereo and fully immersive, are configured to run the same way for the study. The one difference between the two systems is that the immersive system is playing a relaxing soundtrack in the background which users listen to through the Steelseries headphones described earlier. The desktop stereo system is running on an Alienware M17X with a 3D monitor and the six dial interface, and the immersive system is running on a desktop PC, and an HTC Vive HMD and controllers are used as the display and interaction mediums. These systems are described in chapter 3.

Upon commencement of the experiment session, one hypercube is presented to users to introduce them to the environment and the controls they will use throughout the session. This phase lasts for three minutes after which time the "passive" task begins which is based on the ghostcube task described and validated in previous experiments. Upon commencement, the system records a time-stamp marking the beginning of the exposure time.

Extended Ghostcube Task. The ghostcube task has been implemented for use in previous experiments. In the task, users are presented with two hypercubes, one which is static, does not change, and is preset to some rotation, and another which the user is able to manipulate. The task for the user is to manipulate their hypercube to match the second static hypercube (called a ghostcube). In this experiment, there is a set list of 30 ghostcubes for participants to solve. We have implemented new functionality for the purpose of the experiment.

An interface is presented to users with a forwards and a backwards arrow on it. This is designed to allow users to iterate through the list of ghostcubes to the one they wish to attempt. Figure 6.3 (left) demonstrates the interface. On the desktop stereo system, participants use the forwards and backwards arrows on the laptop keyboard to switch through ghostcubes. In the immersive system, participants hold a button on the VR controller to render a laser pointer which they use to point at the arrows. When they release the button on an arrow, the pointer disappears and the ghostcube changes. The amount of time a participant spends on each ghostcube is stored and written to disk upon completion of the session. The ghostcubes that were solved and the corresponding instances in time they were solved are also stored and written to disk upon session completion. This comprises our task completion measure which lies in the observed dimension of data. This has been indicative of subject matter comprehension.

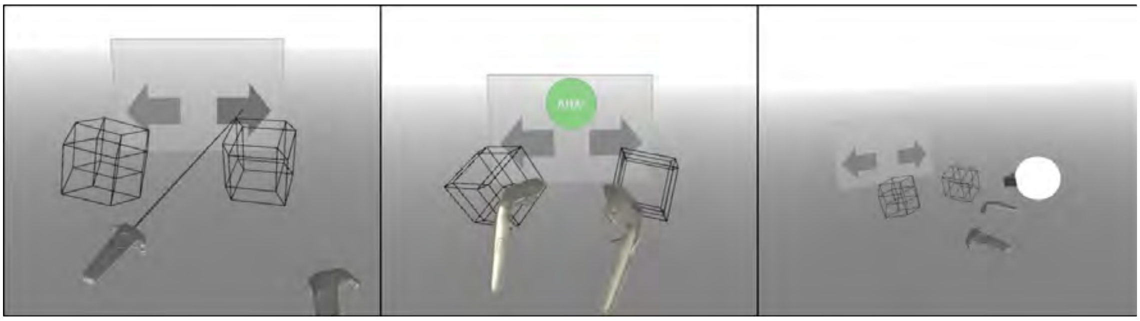


Figure 6.3: The interface presented to users throughout the study on which they can use the forwards and backwards arrows to skip through the list of target ghostcubes (left), the feedback mechanism for when participants report an Aha! moment (middle), and the third-person view of the immersive virtual environment recorded by the screen capture (right).

Aha! moments. We make users aware of the Aha! moment phenomena, and ask that should they experience such a moment they press a button. On the desktop system, we labelled the 'spacebar' key as the Aha! button and asked users to press it in the event of an Aha! experience. In the immersive system, both VR controllers could be squeezed (resulting in a button press of the palm buttons of the Vive controllers) upon experiencing an Aha! moment. When the 'Aha' button is pressed (in either system), a green circle appears on the interface in front of them as feedback to recognise their action (see Figure 6.3 (middle)). When these moments are experienced and the respective buttons are pressed, the system stores a timestamp of the event. This is the second measurement applied within our study, and sits within the psychological (reported) dimension of data.

Environment Capture. We implemented a method for capturing the virtual environment throughout the duration of the study. We want to capture as much as we can without compromising the anonymity of the participants. Therefore, for the immersive system, we implemented a third-person virtual camera placed in the virtual environment in such a way that it sees the relevant portion of the environment. We created an object to represent the users head in the environment so we could see where they were directing their attention at any given time. The object was made up of a white sphere representing the user's head, and a black cuboid representing the head-mounted display. It was placed in the environment as a child object of the head-mounted display so that it it moves and rotates together with the HMD, and it is clear from the third-person camera where the user is looking. The controllers are already rendered so we can see the participants hand movements, and the hypercubes are in view also. Figure 6.3 (right) demonstrates the view from the virtual third person camera. When we run the system, we start the NVIDIA screen capture software to record the scene for the session.

This is more difficult for the stereo system. Not all relevant actions and environmental content can be captured by a screen capture. Instead, we place a GOPRO camera† next to the laptop such that it records the screen and the six-dial interface the participant is interacting with. We get less data with this recording method, but we are still able to record the interactions and what the participant is seeing on the monitor while maintaining anonymity.

In addition to video capture of each condition, we record the audio for the exposure duration of each session to capture anything the participants might say of any relevance to their experience or performance. This summarises the objective measurement of environmental data, both video and audio, for each condition which fits into the observed dimension of data.

Demographics. Data are collected including age, gender, ethnicity, vision, and prior VR experience through a demographics questionnaire.

Presence. We assess the participants' sense of presence upon completion of each session. The 14-item Igroup Presence Questionnaire (IPQ) [145, 161] is applied immediately upon completing the one hour exposure (or less time if they completed all ghostcubes) for each session.

Simulator Sickness. We use the same simulator sickness questionnaire (SSQ) as in previous studies to collect data regarding users' potential symptoms [96].

Emotional Response. We described the Empatica E4 wristband device we employ for this study in an earlier section. We are particularly interested in the EDA, HR, temperature, and accelerometer data. EDA and HR signals are used as our primary measure of emotional response. We describe in the following section (Procedure) how the device is applied. More detail will be given in the proceeding chapter on how these measures are analysed and interpreted.

† www.gopro.com

Figure 6.4: Empatica E4 wristband device for collecting physiological data. The exposed sensors can be seen in each view of the device (circled in red).

Interview. We conduct a semi-structured interview upon completion of each session. After the participants' first session, we use the following questions as guidance for the interview:

1. Do you think you understand how the Hypercube works?
2. Do you understand at all how it relates to 4-dimensional space?
3. Did you feel any sense of frustration at any time? To what degree?
4. Did you feel any sense of relief or pleasure at any time? To what degree?
5. How do you feel about the Aha! moments you reported?
    - Were they sudden?
    - How certain were you at that moment that you knew how to solve the ghostcubes?
6. What do you study/do?
7. Where are you from? Where did you grow up?
8. What do you do for fun (hobbies)?
9. Do you enjoy challenging activities?

Questions (1) and (2) are taken from our earlier studies, and are used to probe the participants for any tangible knowledge gains with respect to the hypercube construct, or 4D space itself. We also want to get an idea if participants demonstrate anything resembling the deep non-verbal comprehension, or "begreifen" of von Foerster. Questions (3) and (4) were to inform the sensed measure of emotional response. Physiological measures are sensing activity of the sympathetic nervous system which can be due to positive or negative emotional stimuli. Therefore, we ask about general frustration levels, relaxation, relief and pleasure. Question (5), (5a), and (5b), are our equivalent of Aha! moment questions which we took from the previous work of Danek et al. [44]. The remaining questions (6), (7), (8), and (9) are to assess the context of individuals. We ask about domain of expertise, hobbies, and enjoyment of puzzles, which provides pertinent data with respect to users' potential performances. We also ask where participants are from, and where they grew up in case any tendencies emerge in our participants' backgrounds.

Following their second session, which completes the study, we only ask questions pertinent to the other measures in our study, i.e. questions one to five. The interviews are all recorded.

6.4.3 Procedure

Each participant completes two conditions (immersive and non-immersive) held on separate (not necessarily consecutive) days with each having 1.5 hours dedicated time. Participants' first condition is pre-randomised. Upon entry to the study participants are presented with an information sheet, and complete a consent form along with the initial demographics questionnaire. The experimenter describes the process of the study and the technology involved. The E4 bracelet is started and given to the participant to put on whichever wrist they are most comfortable with. The wristband needed to be attached relatively firmly on the wrist such that the sensors are not able to frequently lift from the skin. In most cases participants need assistance to attach the wristband. Participants are also provided with a definition of an Aha! moment, or moment of insight, and are told what to do should they experience such a moment.

They are asked to spend an hour within the system in an attempt to solve all 30 of the presented ghostcubes. If they complete all 30 ghostcubes, the system will stop automatically. For the fully immersive condition, the participant takes a seat and fits the HMD and headphones until comfortable and are given the controllers. For the desktop stereo condition, the participant is seated in front of the laptop and interface and given 3D stereoscopic glasses to wear. In each case the controls are described for the participants. We chose to have the user seated in the immersive VR condition due to the long exposure time of one hour. We noticed in our previous investigation (chapter 5) that participants

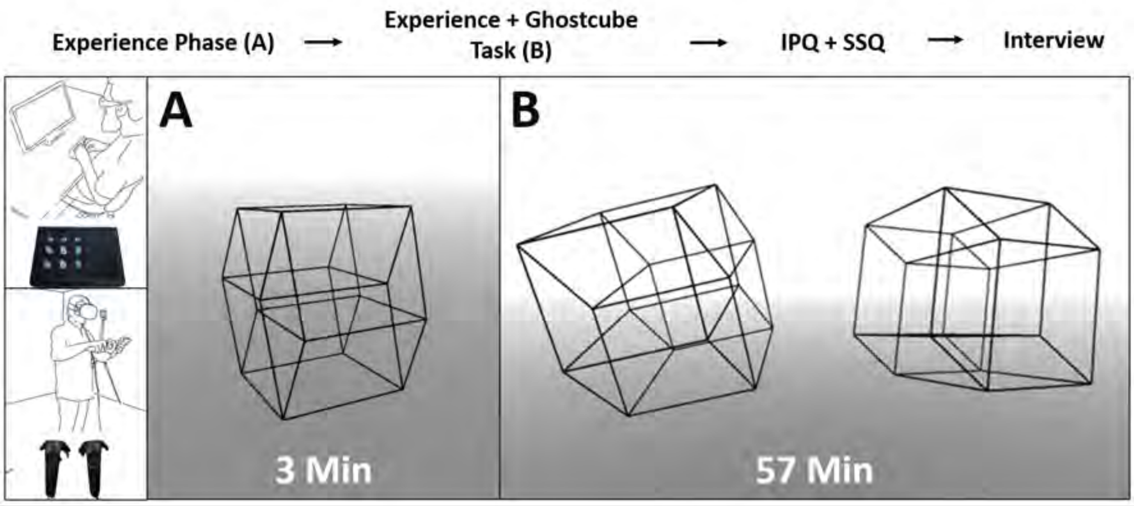


Figure 6.5: The procedure of the main study. The study is of a within-subject design so all participants perform both the stereo and the immersive condition. For both, the procedure is the same: 1) a 3 minute introduction to the environment and the controls (A), 2) 57 more minutes of experience with the ghostcube task running passively (B), 3) the IPQ and SSQ self-report questionnaires are completed, and finally 4) a semi-structured interview is conducted.

began to fatigue after standing for approximately 15 minutes. We wanted to mitigate any fatigue participants might develop in the current study. Participants in the desktop stereo condition must be seated regardless due to the fixed point-of-view.

The system is started and participants have three minutes to get used to the environment and controls (see Figure 6.5 (A)) before the matching phase begins for the remainder of the hour (see Figure 6.5 (B)). Once the hour is up or if the participant completes all of the ghostcubes, we remove the equipment and ask participants to complete the IPQ and SSQ questionnaires. They are also asked nine questions in the semi-structured interview. That concludes the first condition. The same process is repeated when participants return for their second condition with the exemption of consent and demographics questionnaires, and only the first five questions are asked for the semi-structured interview.

6.5 Summary

We have given an outline of the main concepts in this investigation: presence, Aha! moments, and emotional response as measured through physiological signals. We examined engagement in immersive VR learning environments as a shared element between the sense of presence and moments of insight, and evaluated physiological tendencies with respect to learning experiences. We have reported on a preliminary usability study to verify the usability differences are not considerable enough to introduce confounding effects. Finally, the main study of this investigation has been introduced and we continue in the following chapter to report on the results. We will describe the structured analysis in more detail and conclude the investigation with an extensive discussion on the implications and future work.

Chapter 7

Presence, Insight, and Emotional Response: Results

Contents

7.1 Introduction 90

7.2 Psychological Dimension 91

7.3 Observational Dimension 95

7.4 Physiological Dimension 99

7.5 Interpretations and Discussion 104

7.6 Conclusion 111

In the previous chapter we introduced the primary concepts we focus on in the current study, and the complex relationships that exist. A methodology was presented for the systematic measurement and structured analysis of emotional response in VR environments, and how it relates to other phenomena under investigation. After conducting a brief usability experiment and introducing the measures and procedure of the current experiment, we now present the results. The results are presented in the context of our three data dimensions: 1) physiological, 2) psychological, and 3) observational. We discuss the individual treatment and analysis of each, and follow up with a more complex analysis of some of the relationships between the data dimensions. We conclude the chapter with a detailed discussion on the results, as well as the potential application and the relevance of the methodology. Parts of this chapter are included in the following work:

1. Collins, J., Regenbrecht, H., Langlotz, T., Can, Y. S., Ersoy, C., and Butson, R. (2019). Measuring Emotional Response in Virtual Environments: A Methodology

Exemplified in a Learning Context. *In International Symposium for Mixed and Augmented Reality '19 (Accepted).* [39]

The machine learning algorithms were implemented by Yekta Said Can, the fourth author on the paper cited above, who is supervised by Cem Ersoy, the fifth author. Yekta was responsible for training the model based on our data, and performing the classifications. The remainder of the work in this chapter was conducted by the author of this thesis with guidance from his supervisors Holger Regenbrecht and Tobias Langlotz. Russell Butson also advised on the educational elements of the work multiple times throughout the process and facilitated the physiological analysis.

7.1 Introduction

In the previous chapter we introduced our investigation into presence, insight learning, and emotional response in IVRL environments. We first conducted a usability study of our two systems to evaluate for any outstanding usability issues which could confound our other measures in any serious way. We found no major usability differences between the two systems which comprise our two conditions. Following our report on the usability study and results, we introduced the primary experiment for this investigation. It is through this experiment that we answer the previously specified research questions around whether Aha! moments are attainable in immersive VR environments, what impact presence has in IVRL environments, and whether the measure of emotional response is correlated with learning experiences in VR environments.

To assist with our investigation, we outlined a methodology to facilitate the structured analyses of the various phenomena of interest in our environment. We identified three primary dimensions of data: 1) physiological, 2) psychological, and 3) observational. We provide a detailed analysis of each measure within these dimensions beginning with the psychological and observational dimensions, then moving to physiological analyses. Our hypotheses were as follows:

H1: *We can lead users to a reported Aha! moment during their time using the system*

H2: *Users' reported Aha! moments will be more frequent with a higher sense of presence in the VR environment*

H3: *Users' reported Aha! moments will lead to more achievements*

H4: *Emotional Response measures will correlate with presence, Aha! moments, and achievements*

In total, we were able to recruit 24 participants (17 male, 7 female) with a mean age of 23.9 within a range of 18 and 45. All participants completed both conditions. In the remainder of this chapter we report on the results from the experiment, and conclude with a detailed discussion on the interpretations and implications of these results.

7.2 Psychological Dimension

We have three reported measures to analyse in the psychological dimension: 1) presence, 2) Aha! moments, and 3) interviews.

7.2.1 Presence

We measured presence via the Igroup Presence Questionnaire (IQP) [161]. Before performing any analysis, we needed to wrangle the data to reverse negatively weighted questions. Questions three, nine, and eleven are reversed on their scales (see Appendix A for the IPQ questions). Overall, the presence data was found to be normally distributed through a Shapiro-Wilk test. The immersive system yielded higher presence ratings overall (*M=4.74, S.D=0.60*) than the stereo system (*M=2.88, S.D=0.93*) where statistical significance was found through a t-test ($p < 0.01$). Presence ratings are shown in Figure 7.1.

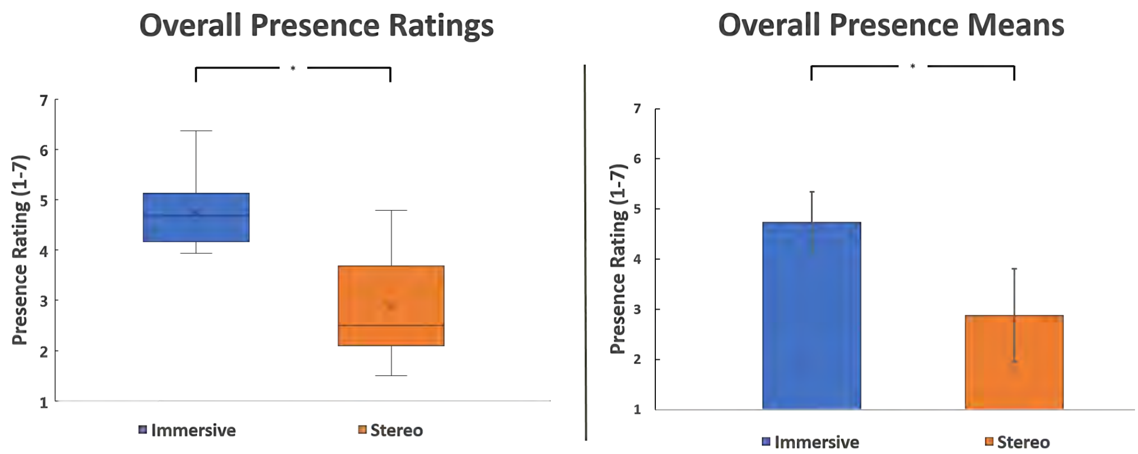


Figure 7.1: Overall presence ratings between the immersive and stereo systems. Significant differences were found using a t-test ($p < 0.01$). ** denotes statistical significance where p < 0.01.*

We also analysed presence ratings for effect of starting condition by comparing each system's ratings as a first condition, against their rating as a second condition. No significant differences were found ($p > 0.05$). We also tested the immersive condition's first ratings against stereo's first ratings and found significant differences ($p < 0.05$). The same was found of the second ratings. Therefore we could find no effect from starting conditions.

Three primary factors can be analysed by grouping questions from the questionnaire which have certain factor loadings: five questions loading on spatial (SP), four questions loading on involvement (INV), and four questions loading on realism (REAL), with one additional general presence question (PRES). For the immersive presence ratings, internal consistencies for *involvement* ($\alpha = 0.67$), and *realism* ($\alpha = 0.65$) were satisfactory, though the *spatial* ($\alpha = 0.32$) factor was reporting a relatively low Cronbach's alpha. Internal consistencies of the stereo presence ratings for *involvement* ($\alpha = 0.68$), *realism* ($\alpha = 0.7$) and *spatial* ($\alpha = 0.73$) were all reporting satisfactory alphas. Presence ratings by factor are demonstrated in Figure 7.2.

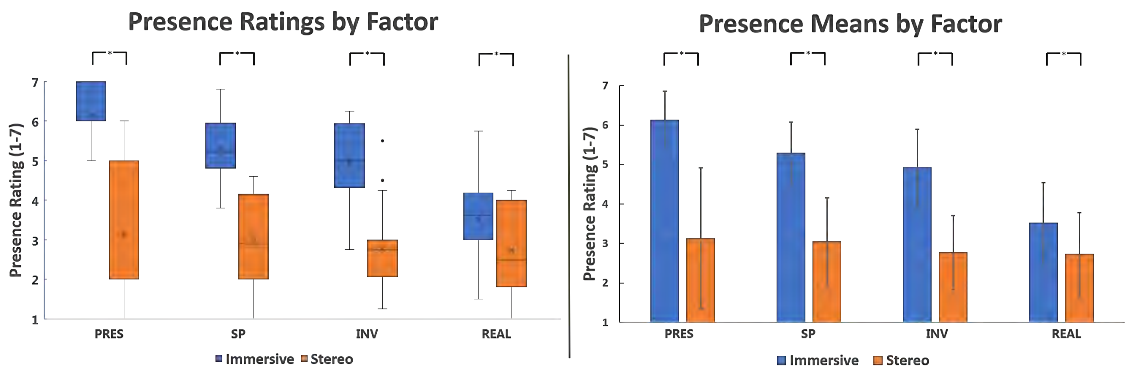


Figure 7.2: Presenting the presence ratings grouped by factor loading. The realism factor rated lower than the spatial and involvement factors. We speculate the difference to be due to the minimalism of the virtual environment.

Shapiro-Wilk distribution tests for the factor analysis found the the immersive SP, INV, and REAL factor data to be parametric, but the PRES factor was not. All of the stereo factor data was non-parametric. Significance tests found differences between the immersive and stereo factors (reported immersive/stereo, SP median = 5.2/2.9, INV median = 5.0/2.75, REAL median = 3.63/2.5, PRES median = 6.0/2.0, $p < 0.01$ *(Wilcoxon)*).

7.2.2 Aha! Moments

We have measured multiple event related data sets throughout this study. The first to be analysed here is moments of insight as reported by participants. Events in our experiment are represented by time-stamps which are recorded by the system. These make up one of our primary forms of data for the overall analysis, especially with respect to analyses of the relationships between phenomena.

Participants in the immersive condition reported a higher number of Aha! moments (M=*1.79*, *S.D*=*3.24*) than when in the stereo condition (M=*1.04*, *S.D*=*1.84*) although these were not found to be significantly different when applying a paired Wilcoxon signed rank ($p > 0.05$). In terms of the raw number of reported Aha! moments, participants in the immersive system reported a total of 43 Aha! moments. When in the stereo system, 25 Aha! moments were reported. There were two observable outliers in the immersive condition with one participant reporting 13 Aha! moments, and one participant reporting 11. Only one outlier is worth noting in the stereo system with seven reported Aha! moments, but three other outliers are registering at three, four, and five Aha! moments.

Out of the 24 participants, a total of 15 reported an Aha! moment in the study leaving nine that did not. Thirteen participants reported Aha! moments in the immersive system where nine participants reported Aha! moments in the stereo system. Of the 13 participants reporting Aha! moments in the immersive condition, seven also reported Aha! moments in the stereo condition. Only two participants reported Aha! moments in the stereo condition alone. We visualise this dataset on a horizontal axis with event markers where the start of the axis is the time recorded by the system upon commencement of the exposure, and the end of the axis is the end of the exposure which for the majority of participants is one hour from commencement of the exposure. Figure 7.3 demonstrates a visualisation of a participant with 13 reported Aha! moments. We will report further on the Aha! moment data at a later time with respect to other dimensions.

7.2.3 Interviews

We use interview data to support our discussion of the results. Interviews can be a rich source of information and they give us a good indication of participants' attitudes towards the environment and the tasks. We also used the opportunity to ask about their perceptions of positive and negative feelings such as frustration and pleasure experienced throughout the session. Due to the emotional nature of such feelings, they can have effects on the sympathetic nervous system consequentially affecting the physiological measures.

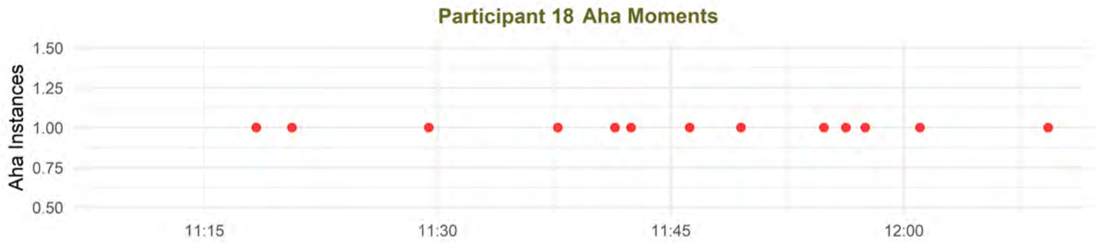


Figure 7.3: An example of how Aha! moments are visualised as events. A participant has 13 reported insights (bottom).

Therefore, it is valuable information to get from participants.

Interviews further provide us with detail into participants' perceptions of their Aha! moments. As we discussed in the previous chapter, we do not present questions to participants regarding each specific Aha! moment due to the consequential break in engagement. Instead, we ask more general questions about their insight moments (or lack thereof) in our post-experience interviews. Therefore, we don't report on interview findings here, but they will be used anecdotally in the discussion of our results.

7.2.4 Simulator Sickness Questionnaire (SSQ)

We measure simulator sickness to evaluate for any symptoms that may have effects on other measurements. This is especially relevant due to the longer exposure times of each experimental session.

As in chapter 5, the nausea and oculo-motor components of the SSQ are analysed [21]. Participants' reports on the oculo-motor symptoms were generally higher than the nausea symptoms. The immersive condition yielded higher oculo-motor scores (*M=5.0, S.D=4.22*) than the stereo condition (*M=4.08, S.D=2.84*). The same trend was found for nausea symptoms where immersive ratings were higher (*M=2.83, S.D=3.4*) than the stereo condition (*M=1.67, S.D=3.75*). Given the scores of the each symptom, the mean overall rating for the immersive condition was higher (*M=7.83, S.D=7.49*) than the stereo condition (*M=5.75, S.D=6.34*). Referring again to the SSQ score categorisation from Kennedy [95], means lying between five and ten are considered as "minimal symptoms".

We conducted distribution and significance tests of the SSQ data to investigate how different the two systems are in terms of simulator sickness symptoms. All SSQ data was found to be of a non-parametric distribution using a Shapiro-Wilk test. No significant differences were found between immersive and stereo SSQ reports for either of oculo-motor or nausea components (reported immersive/stereo, oculo-motor median = 3.5/3.0, nausea median = 1.0/1.0, $p > 0.05$ *(Wilcoxon)*).

We expected the emergence of simulator sickness symptoms due to the long exposure times, particularly for the fully immersive condition. This was reflected in the SSQ results, however while the symptom ratings were observable, they were not a cause for concern in terms of overall confounding effects.

Each item on the questionnaire was rated from zero to three where the options were 'None', 'Slight', 'Moderate', and 'Severe'. Item four on the questionnaire, "Eye Strain", registered as the highest rated symptom for the immersive condition with a mean rating of 1.17. Other items registering higher mean ratings were "Difficulty Focusing" (item 5) with a mean of 0.83, and "Fullness of the Head" (item 10) with a mean of 0.75. Stereo reports were very similar with eye strain reporting the highest (mean=1.04), except instead of "Fullness of the Head" reporting higher, it was "Difficulty Concentrating" (item 9) with a mean of 0.75. The SSQ items can be viewed in Appendix B.

7.3 Observational Dimension

We collected multiple forms of observational data including video and audio recordings of the environment, and system records pertinent to different actions of the user throughout their experience. Similarly to the interview data, video and audio data is extensive and requires an in-depth qualitative analysis involving multiple third party coders evaluating the footage. This type of analysis was not within the realm of our expertise, was out of the scope of this work, and was not immediately pertinent to our fundamental research questions. We discuss the potential of this data in a later section.

All actions and achievements made by the participant throughout the exposure were recorded. Aha! moments were also recorded, but are considered part of the reported dimension of data so we have already covered it above. The remaining event types were recorded by the system:

- *Ghostcube Solutions* - when the participant solved a ghostcube, the system recorded the cube that was solved, and the timestamp of the achievement.

- *Ghostcube Changes* - when the participant chose to change the ghostcube target the system stored the cube which they changed from, and the timestamp of the change.

7.3.1 Ghostcube Solutions

Participant solutions from the ghostcube task are expected to be a strong indicator for learning achievements. The immersive condition had a significantly higher number of solutions (*M=13.54, S.D=7.57*) than the stereo condition (*M=7.63, S.D=5.06*) when using a paired t-test for evaluating means differences ($p < 0.01$). Overall, the participants on the immersive system solved a total of 345 ghostcubes as compared to the stereo system in which 183 ghostcubes were solved. We are interested in two aspects of the ghostcube solutions: 1) the timestamps of the events, and 2) which cubes were solved, and how difficult each was.

The timestamps of each solution were stored at the time of the event as UTC Date/Time strings. As we discussed with the Aha! moments, timestamps are representations of our different momentary events and we rely on these throughout our analysis. We are able to visualize our solution events similarly to the Aha! moments (above), however we provide more detail with the solutions in the form of difficulty. Ghostcube difficulty is determined by different means: 1) by rotational complexity, 2) expert ratings, and 3) completion rates.

Ghostcube Difficulty. Rotational complexity refers to the combinations of 4D rotational planes (xw, yw, zw, xy, xz, yz) that are rotated, and the extent of the rotation. The rotational complexity in 4D space in terms of rotational planes is not intuitive, i.e. we expect that rotation in only one plane must be easier than rotations in all six. In fact, certain rotational planes compliment each other. For instance, if the 'xw' plane only is rotated 75 degrees, it will be a much harder ghostcube to solve than if the 'xw' and 'xy' planes are both rotated 75 degrees. Based on these differences, three categories of difficulty ratings were assigned to the ghostcubes resulting in nine easy difficulty, 11 medium difficulty, and 10 hard difficulty. The difficulty ratings for each cube and their rotations are presented in Figure 7.4

Rotational complexity was also rated by an expert user of the system whose ratings were similar to the difficulty table presented here. There were a number of ghostcubes the expert rated as easy difficulty that were determined as medium by the rotational analysis, and the same from medium to hard. As further support of the categorisation of the ghostcubes, we analysed the solution frequency of each ghostcube (see Figure 7.5 (left)). The frequencies are largely correlated with the difficulties in the table, although there is some crossover between easy and medium difficulty cubes. A more robust indication of the difficulty would be the amount of time participants spent on each cube before being able

	Easy							Medium							Hard					
Cube #	XW	YW	ZW	XY	XZ	YZ	Cube #	XW	YW	ZW	XY	XZ	YZ	Cube #	XW	YW	ZW	XY	XZ	YZ
1	0	120	0	0	120	0	2	140	0	210	20	0	0	3	0	0	0	60	60	0
4	120	60	0	120	50	0	5	60	60	0	60	0	60	7	100	300	210	0	0	50
9	85	60	120	240	45	60	6	0	0	250	300	45	60	8	0	0	0	60	235	80
10	10	60	0	10	60	0	11	200	315	0	0	0	0	12	275	0	0	0	0	110
19	0	0	0	0	0	0	14	180	60	0	0	0	340	13	85	90	230	0	0	0
20	130	0	275	60	0	275	15	0	30	0	0	320	190	18	10	275	10	275	0	0
21	0	85	0	0	85	0	16	30	30	170	0	35	0	23	60	0	120	120	0	60
22	0	0	300	0	0	200	17	0	225	0	125	0	325	25	200	0	0	230	0	100
26	360	0	0	20	0	0	24	160	230	0	0	0	0	27	235	60	0	190	0	0
							29	230	10	90	60	165	45	28	0	0	0	140	0	90
							30	20	220	80	10	0	80							

Figure 7.4: Ghostcube difficulty assignments based on their rotations.

to solve it, however it was observed that many participants spent a lot of time learning and forming their solution strategies on the earlier ghostcubes which would skew the data for such an analysis.

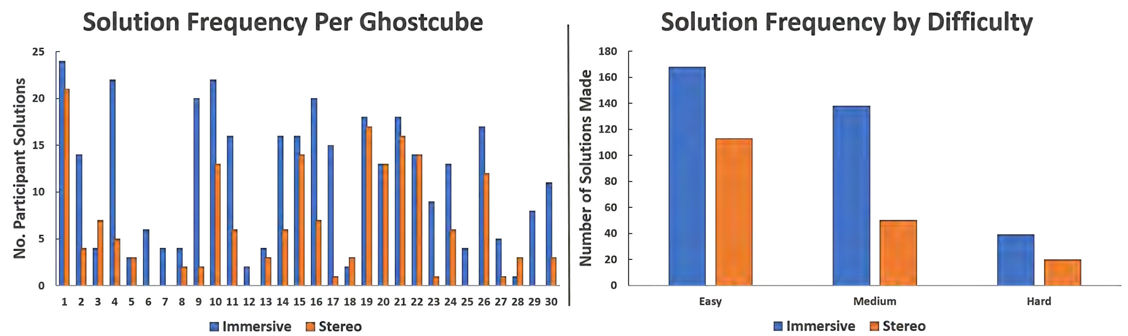


Figure 7.5: The frequency that each ghostcube was solved through the study. The solution frequencies largely correlate with the difficulties in the table although there is some crossover between medium and easy, and medium and hard difficulties (left). The number of solutions reached for each category corresponds to the difficulty rating assigned (right).

When analysing the solution frequencies based on difficulty, the number of solutions produced corresponds to the assigned difficulties. There are approximately the same number of ghostcubes in each difficulty category with there being slightly more medium (11) than easy (9), and 10 hard ghostcubes. Figure 7.5 (right) demonstrates how many solutions were reached for each category of difficulty on each system.

We further analysed the achievements with respect to Aha! moments. We separated the solution frequencies of participants that had Aha! moments from participants that did not. In the immersive condition, participants reporting Aha! moments had a higher mean number of cubes solved (*M=17.08, S.D=7.38*) than those that did not (*M=9.36, S.D=5.31*). A t-test found the differences to be significant ($p < 0.01$). Participants in the stereo condition however were solving very similar numbers regardless of Aha! moments with the insight reporters only very slightly higher (*M=7.67, S.D=4.67*) than participants not reporting insight (*M=7.6, S.D=5.28*) where no significant differences were found ($p > 0.05$). This analysis indicates that Aha! moments experienced within the immersive VR environment are more indicative of effective learning outcomes for the participants.

7.3.2 Ghostcube Changes

The system recorded a timestamp each time the participant changed the ghostcube. In addition to the timestamp is a record of which ghostcube the participant changed off. Participants made frequent use of the switching function. Participants on the stereo system changed cubes more times (*M=81.13, S.D=87.35*) than on the immersive system (*M=59.79, S.D=74.03*). No significant differences were found between the two groups through a paired Wilcoxon significance test ($p > 0.05$). In the immersive system, participants switched a total number of 1435 times, and in the stereo system, 1947 times. Particularly for the immersive system, the ghostcube change frequencies approximate the inverse of the solution frequencies. In comparison, the stereo system changes were uniformly distributed.

Participants' solution and change events are recorded as timestamps and can be visualised in the same manner as the Aha! moments. Figure 7.6 demonstrates an example of our plotted event data.

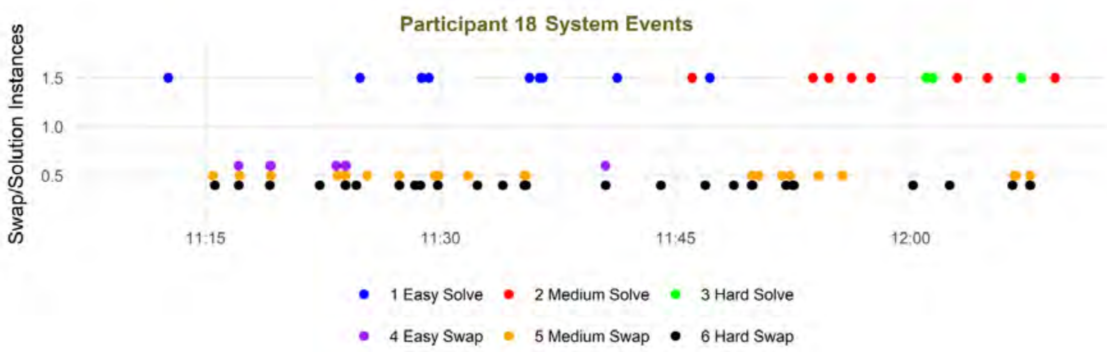


Figure 7.6: Demonstrates how we plot the ghostcube solutions and changes. The solutions are the upper sequence of events and the changes are the lower sequence.

7.4 Physiological Dimension

Participants are required to wear the E4 bracelet for the duration of the study. The bracelet was firmly fitted to the wrist/forearm of the participant. Upon conclusion of participants' sessions, the device is plugged into a PC containing the E4 software and the data is automatically uploaded.

There are multiple stages of analysis for the physiological data. We use both electrodermal activity (EDA) and heart rate (HR) data for assessing the emotional states of participants, and the accelerometer and peripheral skin temperature data are used for treatment (noise reduction) of EDA and HR data. Before conducting the numerical analyses, we first visualise all of our data for the purposes of sanity checking the data, and to visually assess whether there are any obvious trends. This involves plotting the EDA, temperature, accelerometer, and all event related data together. Figure 7.7 presents an example of a resulting plot. All of the participants' data were captured in full except one participant where approximately nine minutes of data were lost from the beginning of the session. The participants' EDA readings from each condition can be viewed in Appendix D for the readers viewing.

The immersive system has produced significantly higher solution frequencies, and a significantly greater sense of presence. The reported Aha! moments also appear to be of a higher quality. Given our results together with the observably higher levels of activity in the EDA data, we opt to conduct our physiological analysis based only on the data from participants' immersive condition. The research questions are pertinent to emotional response and learning outcomes in immersive VR learning environments so for this part of the analysis, we do not see the requirement to include the physiological data from the stereo condition. Additionally, this type of analysis is being applied in a novel manner, and there are much less solutions and Aha! moments generated in the stereo condition for the analysis to work with. The proceeding section describes the analysis of participants' emotional responses in our IVRL environment.

7.4.1 Data Preprocessing and Operations

Physiological data is often filled with noise from movement and particularly in the case of EDA data, skin temperature. Below we describe the preprocessing methods used for cleaning and separating the data where necessary. We also describe the early stages of analysing the physiological data (sensed data) against our event data (observed/psychological data). The first stage involves extracting features from the EDA and HR data which are later

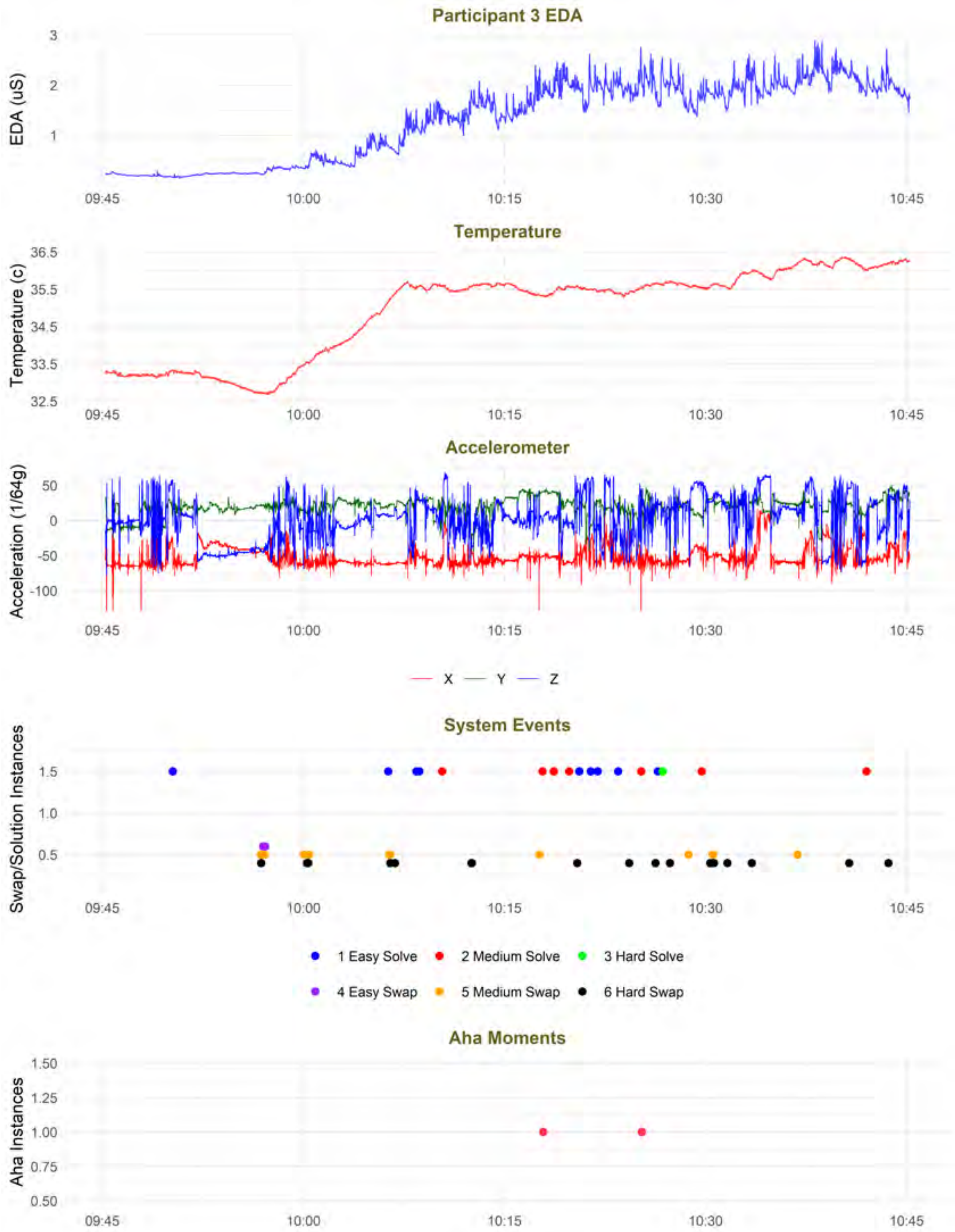


Figure 7.7: Demonstrates how we plot the data from our various dimensions. It includes the EDA, temperature, accelerometer, and event related data which includes solutions, changes, and Aha moments.

processed with a machine learning algorithm to identify relevant emotional responses.

Electrodermal Activity Signal Preprocessing and Feature Extraction Tools. The clarity of the Electrodermal Activity (EDA) signal is affected by the varying intensity of physical activity and alterations in skin temperature. To mitigate these effects, we applied the EDA-Explorer tool developed by Taylor et al. [187] to filter the raw signal data. In their work, they employ two experts to manually label EDA data collected from 32 participants. Data points are labelled as either clean or artefacts based on a specified set of criteria. A total of 1301 labelled data points are given to a Support Vector Machine (SVM) to train a model (Radial basis function (RBF), β=0.1, C=1000, 60/20/20 split). This tool has a classification accuracy of 95.67% for artefact detection with those labels [187].

After removing artefacts from the signal, the two common components of EDA (tonic and phasic) are extracted. The tonic component refers to long-term skin conductance levels and slow changes, whereas the phasic component refers to short-term (event-related) changes in the signal [15]. We employed a convex optimisation approach to decompose the EDA signal into phasic and tonic components by applying the cvxEDAtool [68]. We then extracted seven features from the EDA signal: mean, standard deviation, peak, strong peak, 20th percentile, 80th percentile and quartile deviation (75th percentile, 25 percentile). These features are noted in the literature as the most discriminative for the EDA signal [1].

Heart Activity Signal Preprocessing and Feature Extraction Tools. The heart activity is also exposed to signal contamination due to the movement of subjects. To address this, a preprocessing tool has been developed in MATLAB which employs the 20 percent rule on data and a local average. In this rule, every data point is compared with the local average and detected as an artefact if the difference is higher than 20%. The data points that do not satisfy this rule are deleted. The 20% rule for artefact detection is commonly used in the literature [33]. We have implemented parameters that can be used to either remove the artefact points, or adjust artefact points by applying shape preserving cubic spline interpolation similar to that presented by Hussain et al. [83]. If the artefacts are removed and not interpolated, the cleaning tool can impose new constraints on the remaining clean data. It requires N consecutive data points, or it can set a minimum consecutive time rule similarly (non-interrupted with deleted artefacts) to evaluate the segment worth processing. The percentage parameter of the artefact detection rule and window length of local mean calculation for data point comparison can also be adjusted from the preprocessing tool.

MATLAB's built-in functions along with Marcus Vollmer's HRV (heart rate variability) toolbox* [193] was applied to extract heart activity features. The following time domain features were extracted: HR mean, standard deviation of the inter-beat interval, mean value of the inter-beat intervals, root mean square of successive difference of the inter-beat intervals, the percentage of the number of successive inter-beat intervals varying more than 50ms from the previous interval, the total number of inter-beat intervals divided by the height of the histogram of all inter-beat intervals measured on a scale with bins of 1/128s (HRV triangular index), and triangular interpolation of inter-beat interval histogram. We applied a Fast Fourier Transform (FFT) approach to isolate the separate frequencies within the data. We were able to determine low frequency power (LF), high frequency power (HF), very low frequency power, prevalent low frequency, prevalent high frequency, and the ratio of LF to HF (LF/HF). Lomb-Scargle periodogram was applied to analyse the periodicity of the LF, HF and LF/HF time series data. The features that were extracted are the most widely used, and the most discriminative according to the literature [1].

7.4.2 Detecting Emotional Response

With the EDA and Heart Activity data preprocessed, we conduct the classification with respect to observed data (ghostcube solutions) and reported data (Aha! moments).

Electrodermal Activity and Solution Events. We first analyse the relationship between EDA data, and solution events. We estimate cognitive load trends based on the solution difficulties established earlier. During easier solutions, participants would be experiencing less cognitive difficulty, and harder solutions require more cognitive effort. The difficulty levels of the ghostcubes (easy, medium, and hard) solved by participants in our experiment were used as labels for machine learning algorithms (1, 2, and 3 respectively). In order to classify three cognitive load levels, we have used the Weka toolkit [53]. These classes were imbalanced due to the nature of the data where "hard" labels represent the minority class. We employed a re-sample method from Weka toolkit to balance the data (i.e added samples of the minority class) to prevent classifiers from biasing towards the majority class. To ensure we are providing an exhaustive analysis, we test our EDA features (extracted earlier) using multiple classifiers. We have applied five different classifiers

*marcusvollmer.github.io/HRV/

on our cognitive load data:

A. PCA and SVM with linear basis function

B. MultiLayer Perceptron (7-5-3) (MLP)

C. K-nearest neighbours (k=1)

D. J48 Decision Tree

E. Random Forest (RF, 100 trees)

These classifiers are selected due to their common application for physiological signal processing in the literature. All classifiers in the Weka toolkit were run with the algorithms' default values. For each classifier, results are validated with 10-fold cross validation. Results have been provided for 3-class classification of low, medium and high cognitive load levels. We tried the selected machine learning algorithms on our EDA data (see Figure 7.8 (left - EDA-CL)). The resulting average classification accuracy across the five different classifiers was 48.36% when discriminating using three difficulty levels of cognitive load. The most successful classifier for EDA data was the Random Forest approach which achieved 50.83% accuracy with 7.62% variance.

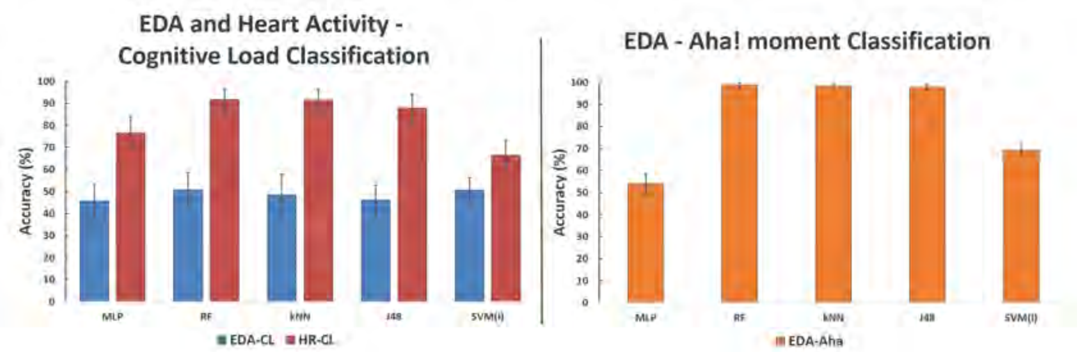


Figure 7.8: Resulting classifications of the Electrodermal and Heart Activity features against cognitive load (solution events), and Aha! moments (reported events). Bar plots show five classifiers, indicating that HR-CL and EDA-Aha moment classifications yield highest prediction accuracy for subjects' difficulty state and approximate moments of insight (Aha!): 1) Cognitive Load classification accuracy using Heart Activity features (HR-CL, mean=82.79%), 2) Cognitive Load classification accuracy using Electrodermal Activity features (EDA-CL, mean=48.36%), and 3) Aha! moment classification accuracy using Electrodermal Activity features (EDA-Aha, mean=83.65%).

Heart Activity and Solution Events. We applied the same process to the heart activity data collected from the Empatica E4 device using the same cognitive labels specified above (based on solution difficulty). We were able to discriminate the three cognitive load levels with a resulting average classification accuracy of 82.79%. For the heart activity data, the most successful classifier is the Random Forest approach which achieved 91.75% with variance 4.87% (see Figure 7.8 (HR-CL)).

Electrodermal Activity and Aha! Moments. Aha! experiences (moments of insight) are times the participants felt that they made a conceptual breakthrough in determining a solution. These events occur rarely. In the duration of our experiment, less than two of these moments occur in sixty minutes on average. This means that the detection of the events is not trivial. Since these moments represent a form of arousal, we have used the EDA signals which can detect instant arousal of individuals. We have employed the same EDA tools as specified above in subsection 7.4.1 for feature extraction of the EDA data.

We were able to classify against multiple classes for the cognitive load classification due to our difficulty categorisation. Aha! moments can not be categorised in such a way, so a binary classification is required. We divided each 60 minute session into 60 segments, each one minute long. A one or a zero was assigned to a segment depending on whether a participant reported an Aha! moment within that minute or not. These values are then given to the classifier.

By applying the Weka toolkit with the same machine learning algorithms, we achieved an average accuracy of 83.65%. The most successful classifier was again the Random Forest approach achieving an accuracy of 98.81% with a variance of 0.9 (see Figure 7.8 (right - EDA-Aha)).

7.5 Interpretations and Discussion

The results of this experiment have provided valuable insights pertinent to the research questions of this investigation and this thesis. We posed several hypotheses from our research questions which we have addressed through a complex and detailed analysis of the various measurements applied in our experiment.

We have analysed each of our measures independently, and we have investigated emotional response in the context of participants' successes and moments of insight. Below we discuss the outcomes in further detail, identifying potential emerging relationships.

Presence. The immersive system produced a significantly higher sense of presence in participants. There was a particularly high report for questions loading on spatial and involvement factors, with a slightly lower rating given to the realism factor. This is coherent with the system as we intentionally designed a minimalist environment to reduce potential distractions.

Aha! moments. While participants reported more Aha! moments in the immersive system, the differences were not significant from reports in the stereo condition. Aha! moments are supposedly an indication of a significant gain in comprehension or learning in general. When we look at the performance outcomes in the context of the Aha! moments, it appears as though the insights did not yield anything for the participants. Despite there being little difference between the reported insights, there was a significant difference in terms of achievement.

However, upon further analysis we found that participants that reported Aha! moments in the immersive system performed significantly better than those that did not. The same is not true of the stereo system which indicates the immersive system facilitated higher quality Aha! moments.

Achievements. There was a significantly higher level of achievement observed in the immersive condition. In our preliminary study, we evaluated the two conditions for any obvious usability issues that could influence the achievement outcomes of this experiment and found no reported differences. The immersive condition in that study did however yield faster solution times on the first ghostcube that was presented. That observation, together with the achievement results from this experiment, could be an indication of the benefits immersive VR environments provide for learning over desktop systems.

Emotional Response Outcomes. The average classification accuracy for the cognitive load with the EDA data was 48.36% with the lowest classifier yielding an accuracy of 45.8%, and the highest classifier yielding 50.83%. The primary reason the EDA classification has returned low accuracy is due to the short consecutive segments of cognitive load data. Ghostcube solution timestamps were often quite close together (in the order of 10 seconds). The more delayed nature of EDA makes it more difficult to isolate significant emotional responses against such frequent cognitive events. For instance, a user spends three minutes attempting to solve a ghostcube of a hard difficulty which results in a steadily rising EDA, then upon completion they are presented a ghostcube with an easy difficulty which they complete in 10 seconds. The EDA takes longer to stabilise making correct prediction difficult for the classifier.

The HR data does not have the issue of delay as it is one of the first responses to the sympathetic nervous system. This explains the consistently higher classification accuracy of cognitive load against the HR data averaging 82.78%. By looking at the HR data alone, we are able to predict the difficulty of environmental subject matter at the approximate rate of the achieved accuracy. For example, we could tell when a subject is struggling with a particular problem, or when they are finding particular problems easy.

Aha! moments were generally infrequent with an average of less than two for each session. Therefore the EDA analysis does not suffer the same drawbacks as with the frequent cognitive load data. The Aha! moment classification yielded an average accuracy of 83.66%. This means we can identify emotional responses in the EDA data which indicate the subject has had, is having, or is about to have an Aha! moment.

A potential limitation with the analyses conducted with the type of data we have lies in the physical movement of participants which can impact the measurements taken from the wristbands. We mitigate for such effects by ensuring the wristbands are firmly attached, and we employ the artefact detection technique of Taylor et al. [187] to clean the raw data of any noise generated from movement. Despite these efforts, if data are to be compared to or used in different contexts, the potential effects of movement differences should be considered.

7.5.1 The Aha! Hypotheses

Our first hypothesis (H1) was that we would be able to lead users to Aha! moments within our immersive VR learning environment. We have certainly achieved this albeit not so reliably that we can state with confidence all factors which contributed to the moment—this is still an investigation of multiple domains such as the education and psychological sciences. When we talk about Aha! moments, sudden insights, or eureka moments, there is an assumed significant gain in knowledge. We can say that Aha! moments experienced within our immersive VR environment appear to be valid learning moments, though it seems they do not hold the significance often characterising such moments. When participants were asked about their Aha! moments in the post-experience interview, they often reported as having grasped a certain aspect of the manipulations. For instance, one participant reported with respect to one of their reported Aha! moments: "... it just made sense when I would see it moving ... it would kind of like cross over in a way that just kind of made sense to me and I was able to use that." The participants that reported at least one Aha! moment in the immersive system (13) solved significantly more ghostcubes than participants that did not report insight (11), so learning is a result of the insights.

We could control for and measure presence reliably in our experiment. The results support the expected significantly higher presence ratings of the immersive system than the desktop stereo system. Our second hypothesis (H2) was that participants' Aha! moments would be increased with a higher sense of presence. There were no significant differences between the number of Aha! moments experienced in the immersive and stereo conditions, so we cannot reliably support this hypothesis. We already covered the significant differences in achievements between those reporting insights and those that did not in the immersive system. The same significance was not found in the stereo system. This is possibly indicative of the quality of the Aha! moments experienced under conditions facilitating presence. In some cases, participants explicitly expressed their preference for the immersive system with statements such as "I preferred it to that [stereo] one. I felt like it was in the real world, and it was easier to control the cube, as opposed to using knobs, actually using my hands." There is huge potential identified here to utilise immersive virtual environments for focused analyses of insight moments. We discussed the potential issues with interrupting VR experiences to ask questions mid-exposure, though perhaps with different subject matter and task delivery, this could be possible.

The third hypothesis was with respect to the relationship between users' Aha! moments and their achievements which we have already covered. We find support for this hypothesis in our immersive condition where the Aha! moments appear to have had an effect on learning outcomes but again, we cannot support this hypothesis with total confidence due to the lacking effect of Aha! moments on achievements in the stereo condition.

7.5.2 Emotional Response in Virtual Environments

Our final hypothesis (H4) has two components. One component states a relationship between emotional response measures (EDA/HR) and learning outcomes. Firstly, we conducted an in-depth analysis of the physiological signals with respect to both Aha! moments and solution events. We found the electrodermal activity data, applied through our methodology, to be a poor indicator of cognitive load due to the high frequency of solutions. However the heart rate data was able to provide a high confidence in predicting cognitive load. The EDA data was however able to provide high confidence in prediction of Aha! moments. We have already established that Aha! moments in our environment have been indicative of learning experiences, so for both of our emotional responses analyses, achievements and Aha! moments, we have found supporting evidence of a relationship between our emotional response measures and learning outcomes. We can therefore support the first component of our final hypothesis (H4).

The second component states a relationship between emotional response measures and presence. Before conducting the analysis of the physiological signals, we discussed the reasoning behind limiting our classification analysis to the physiological readings of the participants' immersive condition only. Furthermore, we did not conduct a numerical analysis of the physiological data based on averaged or instance-based measurements such as those found in the work of Felnhofer et al. [56]. A technical requirement of these types of analysis is the collection of baseline physiological readings. We did not collect this data as it was not a requirement of our methodology, and the purpose of the process we have developed is to provide a more continuous and targeted analysis of the user experience. However, it would be prudent for an evaluator to include baseline measures in future work if the experimental design allows for it, therefore enabling different types of numerical analyses if desired.

When we consider the findings of the previous hypotheses: H1) leading users to Aha! moments, H2) Aha! moments increase with higher presence, and H3) Aha! moments corresponding to achievement, it appears there is a common trend that the immersive condition supports the learning experiences we are investigating. Given these results, we only find weak support for our hypothesis with respect to the immersive condition and therefore we cannot support our fourth hypothesis (H4) overall. Further investigation is required to refine the hypothesis and provide more detailed answers to the related research questions.

We asked participants about their feelings of frustration and pleasure throughout their experiences due to the potential impact of such feelings on their emotional response data. In general, participants expressed frustration at moments in time within the immersive system with reports such as "I got frustrated on this [immersive] one because I couldn't get it to do what I wanted it to." On the stereo system, frustration levels were consistent due to participants' difficulty with the fixed point-of-view, reporting "I could get it to look the same, but couldn't get the [match]." Participants needed to be able to adjust the rotations of their hypercube for perception. The same issue was present in the immersive system, but participants were able to move around and change their perspective. Even with this advantage, those participants still had trouble with the issue of perspective, reporting for example, "I had a bit of an issue, I was trying to match the pattern exactly. I found that didn't work, [but a] slight rotation to the left, and it works!"

Participants also reported high satisfaction levels upon solving ghostcubes on both systems. After being asked if they had a sense of satisfaction, reports were similar to the following response: "at the start yeah, I would cycle through them and an easier one would pop up and I'd feel really good after I'd got one of those." Other reports specifically mentioned a greater sense of satisfaction after solving more difficult ghostcubes, "Yeah definitely, cause especially more with that [harder] one because when I did get [solve] one it wasn't just luck, I actually did something to get there." Overall, participants reported a tendency to experience a mixture of frustration and feelings of pleasure or relief within both systems.

7.5.3 Implications and Future Work

These findings have multiple implications for the research community. Firstly, the emotional response analysis methodology we have outlined and implemented provides the foundation for future application of these measures in immersive VR environments. Being able to predict the cognitive load of learners in real-time has obvious application in educational VR. For example, if we can determine when a learner is beginning to struggle with certain problems, or when they are finding content too easy, we can design virtual environments to tailor the content, provide indicators, or give hints, at moments in time that will benefit learners. We could also trigger dynamic changes in user interfaces by manipulating user control, feedback mechanisms, or all parts of the interactive feedback loop we discussed earlier in the thesis.

The application of our methodology goes beyond the landscape of educational VR. The growing number and quality of compact wearable sensors means they will likely be used with increasing frequency in all areas of VR research. They are of particular use in fields investigating emotional stimuli, although we have found this measure a useful indicator of subconscious emotional response in more general, naturally occurring VR environments. Many applications of physiological measures evaluate participants' emotional states under conditions intended to induce stress [49, 55, 123]. We present a method with potential application outside the realm of stressful virtual environments.

Task workload is commonly measured in virtual environments. The most common tool for measuring task workload is the NASA-TLX (NASA - Task Load Index) [72] questionnaire which is typically presented to participants after completing one or more given tasks

[12]. The method of analysis presented in this work provides researchers with a new way to investigate the task and cognitive load of participants throughout an experience. This provides more detail to the analysis in addition to the application of questionnaires, especially in the case a task has multiple stages, a continuous measure of emotional response indicative of cognitive load could provide insight into which stages of task execution are more or less stressful.

We identify a range of outstanding analyses and potential studies to be conducted as a result of our work. In terms of the current study, we did not utilise the environmental (observed) video or audio data, and we did not formally analyse the reported interview data. Further depth could be provided by analysing the video capture of participants' actions in the virtual environment, and marking when a learner acts in a way that is conducive to success. For example, we can dichotomise any given solution to a specific number of actions. This could be called a solution taxonomy. In the case of a ghostcube solution, the formation of any given solution can be described in six unique movements. We observed that often participants would pick up on a set of those six movements, but they did not often establish all six, or the relationships between them required for complete mastery. We would include awareness of perspective as a seventh component in such a taxonomy because movement in the environment was advantageous for participants in the immersive condition. Such an analysis coupled with our existing continuous measure of emotional response in virtual environments could provide much insight into problem-solving processes. Furthermore, we could analyse any common behaviours exhibited by participants leading up to moments of insight.

Other future work includes further analysis of presence and emotional response as measured through our method. A potential limitation of this work is routed in the measurement of presence via self-report questionnaire only. We emphasised earlier that various prior work have found self-report questionnaires to produce potentially incoherent results when measuring between mediums [192] or to be sub-optimal measures altogether [170]. It is our hope that we can contribute to the current space of continuous measures of presence through our systematic approach to measurement and analysis of physiological signals. Recently, Liebold et al. applied continuous measures of presence using an analogue slider throughout a study for example, although this still requires a users' conscious awareness of what it is they are responding to [113]. In other work, physiological measures of subconscious experience often take an overall composite measure of the tonic component of EDA from throughout an exposure and correlate this with self-reported presence [56, 63]. A continuous physiological measure of sub-conscious experience could contribute to measures of presence in immersive VR environments.

To further analyse the value of the insights provided in our study, a correlative analysis of Aha! moments to solution difficulty could be conducted. We could also conduct a more focused study of insight moments in VR environments and attempt to assess the quality dimension of the Aha! moment.

7.6 Conclusion

We have been able to show through this investigation that users of IVRL environments are able to attain insights indicative of learning gains. We could also measure the Aha! moments in the context of users' tangible achievements, and found that with high confidence, we can predict moments of insight based on emotional responses sensed through physiological measurement. We can also predict users' cognitive load based on a continuous stream of physiological signals. We found evidence that the users' sense of presence in immersive VR affected their learning outcomes, and potentially the quality of the reported Aha! moments. These results further demonstrate the validity of, and the unique opportunities provided by, immersive VR learning environments. We had an opportunity to apply our work in an immersive VR learning application in a real-world scenario. We aim to verify the efficacy of our systematic application of measurement and structured analysis. We report on our findings in the following chapter.

Chapter 8

Immersive VR Education: A Real World Exploratory Study

Contents

8.1 Introduction **114**
8.2 Prototypical Development **116**
8.3 Commercial Development - Study Apparatus **121**
8.4 Exploratory Study **126**
8.5 Discussion **129**

This chapter summarises a real-world project that is a collaboration between academia, industry, and an entity dedicated to community work. This work is exemplary of the reason we conduct our research—how is our work applicable outside of the research community? We have developed a prototypical immersive VR application designed for delivering low-level literacy and numeracy content to illiterate adults. This development has been taken to the commercial sector and is currently under product development. The initial target population for the application are those currently held in a correctional facility, but who have the motivation and determination to educate themselves. In this chapter we discuss the current life-cycle of this project including the development, initial tests, and an exploratory study we conducted. We conclude with a discussion of logistical issues, potential research opportunities, and current outcomes.

The work described in this chapter involved multiple parties including the Methodist Mission Southern (MMS), Animation Research Limited (ARL), and University of Otago researchers. This project is a collaborative effort. The prototypical application was designed between the MMS and the university team, and implemented with contributions from ARL in the form of the 360 degree photography and photo stitching. The commercially developed VR application used for the exploratory study is a collaboration between the MMS and ARL who granted their permission to use figures of their development. The exploratory study design, data collection, and preliminary analyses are attributed to the university research team.

8.1 Introduction

The purpose of the research conducted throughout this thesis is not intended to remain in the theoretical realm of contribution. It is the hope that it can be applied in a tangible way, producing a positive effect on the learning potential of individuals. Throughout the course of this work, an opportunity arose to demonstrate the potential effectiveness of VR in a real world educational scenario.

The Methodist Mission Southern* (MMS) are a non-governmental organisation based in Dunedin, New Zealand, who are involved in community work throughout the region. One of their focuses is the education of prisoners at a regional correctional facility where they help to deliver basic literacy and numeracy classes to high-needs foundation learners. A common issue with delivering educational content to high-needs learners is their lack of confidence, motivation, and engagement with content. To mitigate this issue, the MMS use a contextualised learning approach in which they wrap the literacy and numeracy content in vocationally-relevant contexts to improve the engagement and motivation levels of their learners. To take contextualised learning one step further, they conceived of the idea to deliver this content using VR technologies. This would further provide an immersive environment capable of captivating learners within their contexts of interest.

The first step in pursuing such a project was to select a context that would be relevant to a large enough subset of the learners attending classes at the correctional facilities. Upon questioning the prisoners of their interests there were a number of commonalities, although most were unethical in nature such as gambling, tattooing, or drugs. However one very clear interest among many was automobiles. At this stage of the project we, at the HCI laboratory at the University of Otago, were approached by the MMS where we discussed the potential of the project and its relevance to our research. We committed to the development of an early prototype to demonstrate the idea in practice allowing it to be more easily presented to interested parties, investors, and the New Zealand Department

*www.dmm.org.nz

Figure 8.1: Presenting the immersive Virtual Mechanic application. A virtual car is placed inside a real workshop which users can look around in (left). Specific car parts can be isolated for inspection where parts can be exploded (brake system). We show the literacy content specific to the car parts with multiple possible functions to help learners (right).

of Corrections whose support was required.

We began the design process with the MMS who provided the pedagogical guidance and contributed to the design of the content delivery mechanisms. After several design and development iterations, we had a functional prototype we call the "Virtual Mechanic" project which demonstrated the basic concepts of immersive contextualised learning in VR where a mechanics workshop was the primary context (see Figure 8.1). The prototype was presented to a small number of prisoners from the facility, teaching staff, and other third parties such as investors, and was very well received. This prompted the MMS to take the project to the next step, and begin collaboration with a commercial partner who would develop the project into a product. The MMS began a collaboration with Animation Research Limited† (ARL) based in Dunedin, New Zealand.

The project development team at ARL has started to build the application from the ground up in a scalable manner with the intention this product will be adopted on a large scale. The project is being developed with an initial goal of 40 hours of educational content which is planned for early 2019. Before adding a considerable amount of content, ARL has developed the principal components of their system integrating a small number of literacy and numeracy achievement tasks. As a preliminary evaluation, ARL and the MMS opted to present the current state of their system to groups of learners of different levels at the facility. The goal of the evaluation for ARL and the MMS was to observe the prisoners' interactions with the interfaces and content of the system, and attain any user feedback. We were prompted to conduct an evaluation at this stage utilising the same group of users.

We conducted an evaluation where prison literacy and numeracy learners were exposed

†https://arl.co.nz/

to different content delivery mediums, one of which was the immersive VR Virtual Mechanic application. The study is exploratory in nature, and investigates how our methodology and the early stages of our structured analysis process could work in a real-world scenario, and to identify potential issues in measuring different dimensions of data. In the remainder of this chapter, we report on the project development life-cycle to date, and the exploratory study conducted at the Otago Correctional Facility. We conclude the chapter with recommendations for analyses, lessons learned, and implications of future real-world immersive VR learning developments.

In the following two sections, we describe two separate application developments. The first was developed primarily by the research team from the University of Otago as a prototypical proof of concept for the purposes of demonstrating the concepts and ideas to interested parties. The second application is in the early stages of development as a collaboration between Animation Research Limited and the Methodist Mission Southern. The second system is used in an exploratory study described in a later section.

8.2 Prototypical Development

There were two base requirements for the system: 1) an immersive VR application where the user would wear an Oculus Rift headset for visual immersion, and 2) the scenario context: a mechanics workshop. The Virtual Mechanic prototype went through several design iterations with respect to the environment, interface, and tasks. The prototype was fully developed within the Unity3D engine[‡]. This section summarises the design choices and the development steps for the prototype application.

8.2.1 Environment

There were two primary options for the design of the environment: 1) an entirely virtual scenario with a surrounding virtual workshop and virtual contents, or 2) a realistic environment generated by mapping photographs to surfaces in the environment to achieve the illusion of a real workshop, and integrating virtual content within. We opted for option (2). The realism of 360 degree panoramic images would provide a convincing experience for prisoners.

We were able to contact staff at a local mechanics workshop and arrange a time to take photos. ARL provided the stereoscopic 360 degree camera rig and joined us to take the

[‡]https://unity3d.com/

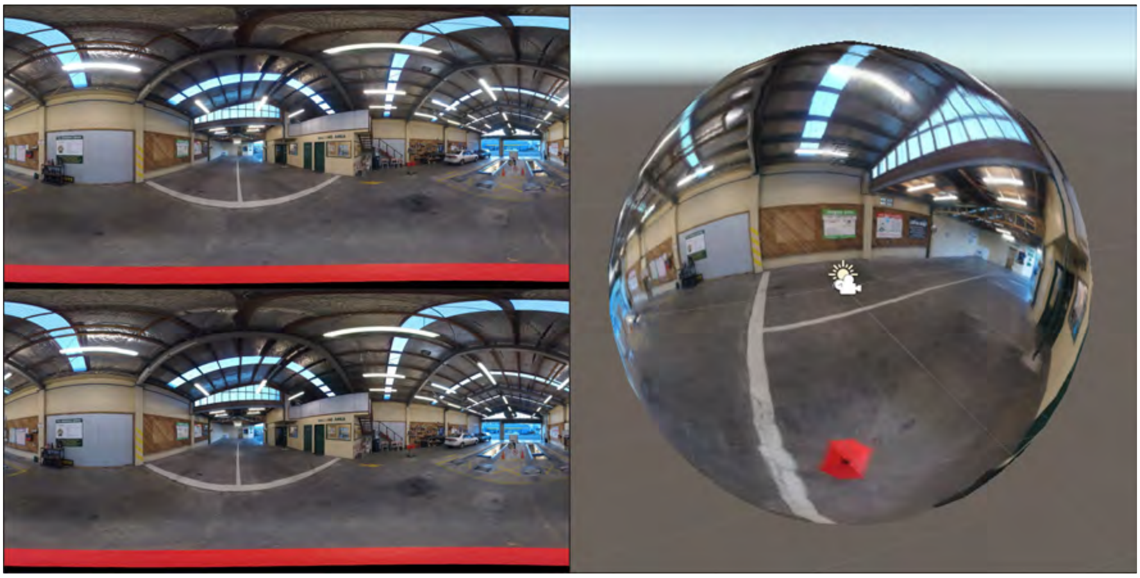


Figure 8.2: The resulting stereoscopic 360 panorama photograph of the mechanics workshop. The photos and stitching were provided by ARL (left). An example of the image textured to a sphere in the Unity environment (right).

photographs. While at the workshop, we took images from four different positions in the environment so we could potentially provide different perspectives in the virtual environment. Figure 8.2 (left) demonstrates a stereo pair of one of the stitched images where the left and right eye were mapped to the top and bottom halves of the image respectively. Figure 8.2 (right) shows one of the images mapped to a sphere. We implemented two different methods: 1) a monoscopic implementation where just one image of the stereo pair would be mapped to one sphere that the user would see with both eyes, or 2) a naive stereoscopic implementation where the left and right eye halves of the image were mapped to separate spheres, and the two virtual cameras representing the users' eyes were set to cull the sphere representing the opposite eye. Spheres were displaced the same distance apart as the stereo panoramic camera rig (66mm).

8.2.2 Virtual Content Integration

We decided to insert a basic car model in our environment as a place-holder for what would eventually be more complex and detailed car models (see Figure 8.3). Because the surrounding environment was mapped to a sphere, there was no ground plane which meant the model car would cast no shadow. This resulted in a break of the visual coherence of the scene [36]. To remedy this, we inserted a ground plane and applied a modified custom

Figure 8.3: The figures above demonstrate the virtual car inserted into the environment surrounded by the panoramic workshop. Each figure is taken from a different position in the workshop. A basic model car is used as a place holder for demonstration purposes. We perform shadow rendering onto an invisible plane to achieve visual coherence of the scene and provide depth cues.

shader which would render all pixels of the ground plane invisible except pixels that received shadows. The resulting shadows are demonstrated in Figure 8.3.

8.2.3 Navigation and Interaction

This prototype was under development before Oculus touch controllers were readily available, otherwise they would have been the clear choice as a device for interaction. At the time, two devices were considered for navigation and interaction. Firstly, we discussed the use of a 3D mouse such as a space mouse by 3Dconnexion[§]. Using this device for navigation did not make sense after our selected environment within which we can only place users at specified positions. Therefore, we went with an XBOX controller given that they are a prolific device, and we expected our target users to be familiar with them.

There were four positions a user could be placed in to view the virtual car in the scene. We implemented navigation about these four positions on the two shoulder buttons of the XBOX controller (commonly known as L2 and R2). Figure 8.4 (left) shows the display above the main car model prompting the user with possible actions.

For interactions, we implemented a centre gaze indicator. The main interaction metaphor required users to place the gaze marker over the object they wish to interact with, and if an interaction is available, a prompt will appear telling users to press the ‘A’ button on the XBOX controller (for example, see Figure 8.4 (right)).

[§]www.3dconnexion.com

Figure 8.4: The virtual car presents a display guiding the user to which controls can be used to navigate (left). The small black gaze dot is always centred within the users' view frustum. When the user gazes at an object that is interactive, the gaze dot turns into a picture of the 'A' button of the XBOX controller, prompting the user to press the button.

8.2.4 Additional Virtual Content

The environment developed so far had provided the overall context for the educational content to be integrated with. We decided that the educational content would need to be based on potential real world tasks. We decided that literacy content would be our main focus for the prototype. We would present a further scenario where a user would focus on a specific car part as an exemplar method of delivering literacy content. The car part we selected was the brake assembly (see Figure 8.5 (left)).

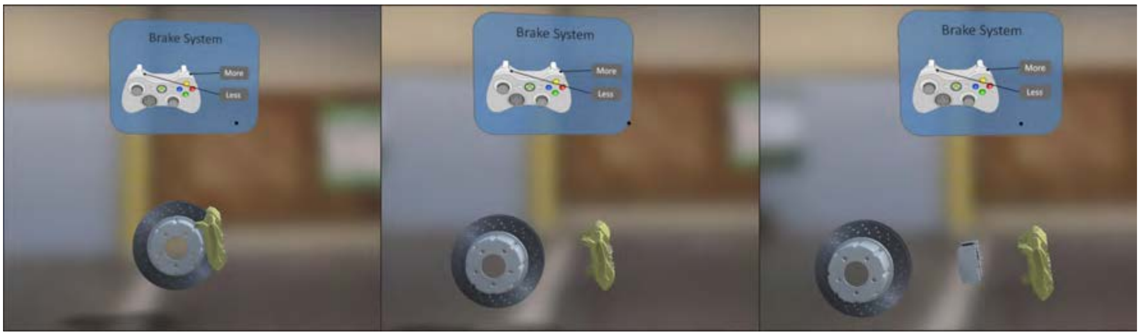


Figure 8.5: The figures above demonstrate the virtual brake assembly (left) with an accompanying display presenting the assembly name and a picture of the interaction device (XBOX controller) identifying possible controls. The brake assembly can be separated into its main components (middle and right).

The wheel of the virtual car model was programmed to pulse indicating that an action was possible. When the user gazed at the wheel of the virtual car with their gaze dot, they would be prompted to press the '"A' button (see Figure 8.4 (right)). When they did, the scene would smoothly fade away, and then back into view, the virtual car would be gone, and the brake assembly would remain in a close up state as in Figure 8.5 (left). We also applied a gaussian blur to the background environment to direct the focus of the user towards the brake assembly.

We programmed the brake assembly to allow the user to perform a simple two-level explode function which separated the brake assembly into three primary sub-components (see Figure 8.5 (middle and right)). We experimented with different controls for this interaction, but we settled on using the upper shoulder buttons of the XBOX controller (often referred to as L1 and R1) to disassemble and reassemble the sub-components.

8.2.5 Educational Content

We ran through several scenarios in a storyboarding session to design the educational content that would be integrated with the current environment. We conceived of one passive, and one interactive activity. For the passive activity, users were provided a prompt when gazing at specific brake assembly parts (Figure 8.6 (left)). We presented the names of specific brake assembly parts upon the users prompt (e.g. Figure 8.6 (middle)). By gazing at the speaker symbol on the display and actioning the prompt, users trigger a voice over which demonstrates the pronunciation of the part name, and would break the word down into its syllables (Figure 8.6 (right)). This was implemented for all three sub-components of the brake assembly.

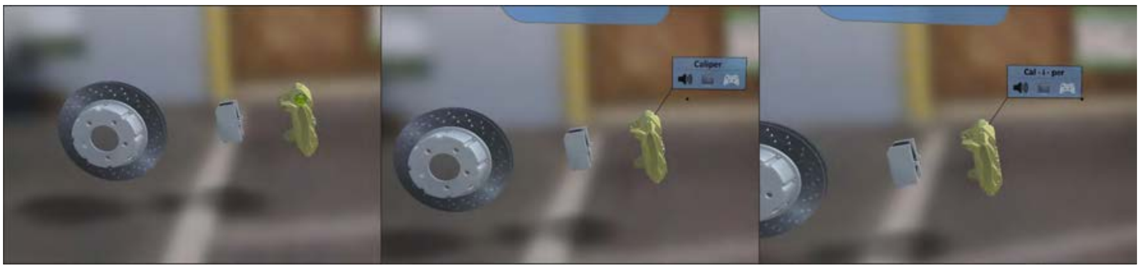


Figure 8.6: The passive task in which users click 'A' when prompted on a sub-component (left), can see what the sub-component part name is called (middle), and can click on the speaker symbol to have a voice over say the name of the part first normally, and then by syllables (right).

The active task was implemented based on a rhyming task. This task was only implemented based on the "brake pad" sub component of the brake assembly which users could activate by hovering their gaze dot over the controller symbol on the brake pad display and pressing the ‘A’ button (Figure 8.7 (left)). Upon activating the rhyming task, a voice over provides instructions for the user stating they must select the first letter of a word ending in ‘ad’ from the set provided which makes a word that rhymes with ‘Pad’ (Figure 8.7 (middle)). Users hover over letters with their gaze control as usual and press ‘A’ when they wish to select a word. If the word is correct, it is said aloud by the voice over, and is added to the list of correct words on the left of the display (Figure 8.7 (right)). If an incorrect word is selected, it appears but flashes red and is not added to the list of correct words.

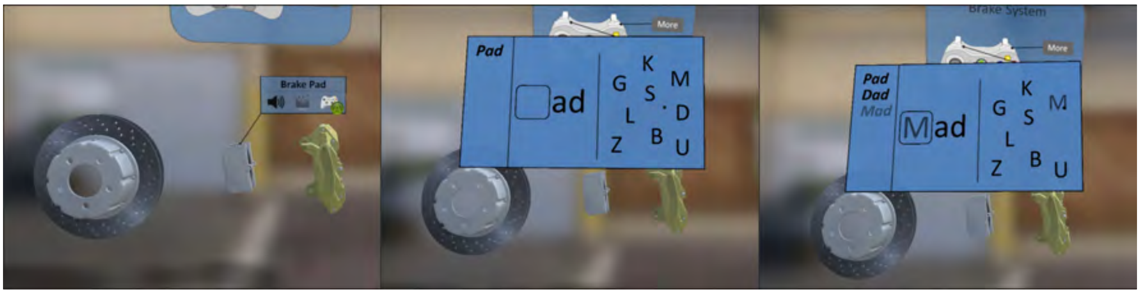


Figure 8.7: The active task in which users click ‘A’ when prompted while hovering over the controller symbol of the brake pad sub-component display (left). A rhyming task is activated in which users must select the first letter from a set of letters to make words that rhyme with ‘pad’ (middle). When words are correct, the voice over says the word, and adds it to the list of correct words (right). If an incorrect word is selected, it appears but flashes red and is not added to the list of correct words.

With the environment implemented and the exemplary educational tasks integrated with the context, the prototype was ready for presentation to various parties. This was a successful undertaking and facilitated the project’s progress by gaining the support of entities such as the Department of Corrections and further investment from other third parties. A collaboration was formed between the MMS and ARL to further develop a commercial product based on this idea. This led to an early stage development which we could use in an exploratory study. We would have the opportunity to test aspects of our previous investigations in a real world scenario implementing an immersive educational VR application.

8.3 Commercial Development - Study Apparatus

In the months leading up to the planned date of the exploratory study, the leading member of the Methodist Mission Southern and one of the researchers from the university attended meetings at ARL to discuss various parts of the development. The commercial product is developed by Animation Research Limited. The application is developed in the Unity3D engine where the target head-mounted display is an Oculus Rift, and Oculus Touch controllers as primary interactive tools. In the proceeding section we describe the environment they have built, and the early stage educational tasks they have implemented for the purpose of user testing and the exploratory study reported on in this chapter.

8.3.1 Virtual Environment and Content

A fully virtual environment was selected for the product development due to the increased flexibility and control it provides. It also facilitates more dynamic navigation within the environment. A virtual mechanics workshop was modelled with typical workshop equipment placed throughout the environment (see Figure 8.8 (left)). A large roller door was built into the model to allow users to navigate outside the workshop into a car-park. The environment is surrounded by a cityscape with real-world buildings built in (see Figure 8.8 (right)). Users are not able to navigate beyond the car-park boundaries. All tasks are designed to take place within the workshop and the car park areas.

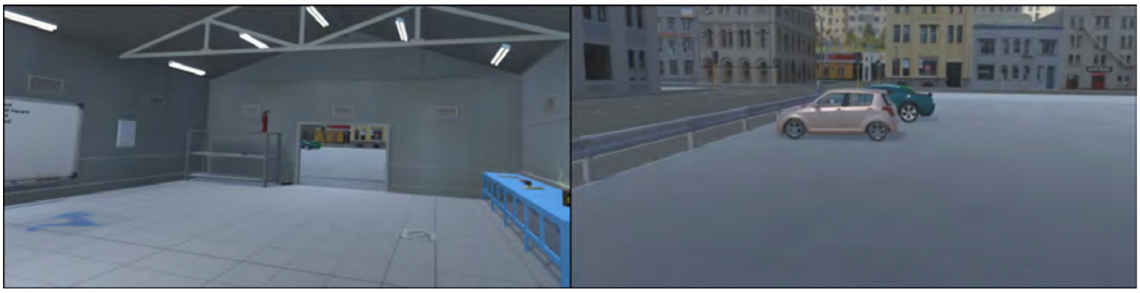


Figure 8.8: The virtual environment produced for the commercial Virtual Mechanic application. The workshop environment contains various workshop equipment (left). Users can navigate outdoors to a car park where they are surrounded by a city texture with real world buildings (right).

Navigation. A common navigation technique implemented in VR environments is the teleport metaphor. There are variations of its implementation, but generally, a user holds down a button on the controller (Oculus Touch) and points the controller where they want to move, then upon letting go of the button they teleport to that location. For visual ease and acceptance, a screen fade is often implemented on the teleport action. The screen will briefly fade out and then back in at the new location. This is the navigation metaphor implemented in the Virtual Mechanic application.

An issue previously identified by the development team was that users would get disoriented once the screen faded back in at their new location. To solve this, the teleport interface was modified. The user would still hold down a button to select their new teleport location, however a 3D arrow and a model of the Oculus Rift HMD is added at the new teleport location which the user can direct using the joystick on the Oculus Touch controller (see Figure 8.9 (left)). When the user lets go of the teleport button, they will reappear facing the direction the arrow was pointing (see Figure 8.9 (right)).

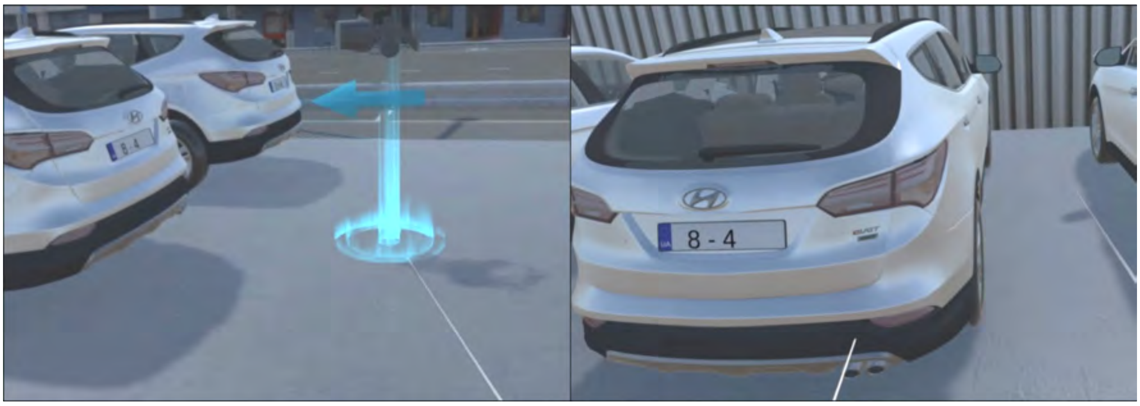


Figure 8.9: Demonstrates the teleportation metaphor used to navigate throughout the environment. When the user holds a button down, the circle on the ground represents the new location to teleport to, and the arrow and HMD model point in the direction the user would be facing upon teleporting (left). The user is facing the expected direction upon teleportation (right).

8.3.2 Educational Tasks

Three active tasks have been implemented within the Virtual Mechanic application. One is a literacy task, and two are a combination of literacy and numeracy tasks.

Task 1: Identify Workshop Hazards. Each task presented to users is written on a virtual whiteboard in the workshop environment Figure 8.10 (A). For each task, they are required to navigate to the whiteboard and read the task board to find out their current task. The first task users are presented with is a hazard identification task. There are two motivations behind this task: 1) to teach users about health and safety awareness within workshop environments, and 2) to have users read and select hazard names from a list that correspond to the hazard presented to them. Hazards to identify are presented to users on the whiteboard, and they must navigate the workshop until they find and correctly identify them all. Shock, slip, and trip hazards are examples pictured in Figure 8.10 B, C, and D respectively.

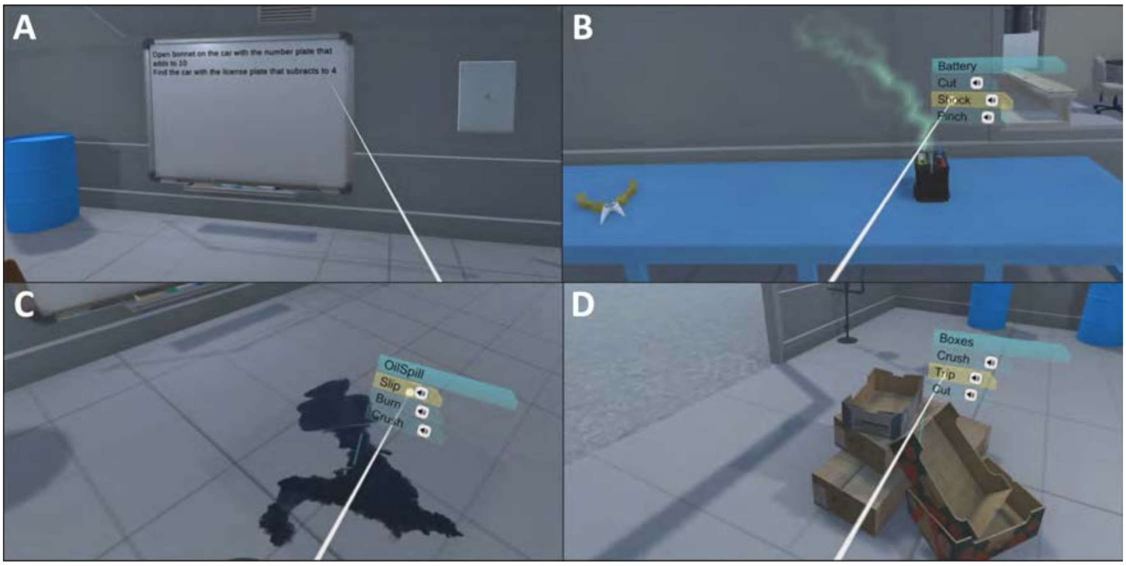


Figure 8.10: Tasks are presented to users on a virtual whiteboard in the workshop environment (A). Different hazards are presented which users must navigate through the environment to identify, i.e. shock (B), slip (C), and trip (D) hazards

Each hazard option in the list presented to the user has a speaker symbol attached to it so they can hear the word spoken aloud by voice over. When the user selects the correct hazard from the list, the voice-over pronounces the hazard aloud, and visual feedback is provided by the animation of a safety cone being placed on the hazard. The virtual whiteboard updates accordingly.

Task 2: Fill the Oil Drums. In the second task users are presented with oil drums that are partially empty. Signs on the wall above each drum show how full it is, and users must select how many litres of fluid are required to fill the drum. The example pictured in Figure 8.11 (left) shows an oil drum that contains six out of eight litres, and the user must select how many litres from the list to fill the barrel. When they select the correct amount, the selection turns green, and the barrel fills (Figure 8.11 (right)). There are three barrels for users to fill. The voice-over is once again implemented on the buttons so the user can hear the options spoken aloud if they have trouble reading it.

Task 3: Number Plate Arithmetic. The final task presented to users required more navigation and exploration of the outdoor environment. Search tasks with numerical arith-

Figure 8.11: Users are presented with barrels which are partially full. They must select the correct amount of fluid required to fill the barrel from a list of values (left). Upon correct selection, the barrel fills, and the correct option is coloured green (right).

metic built into them were written on the virtual whiteboard for users to read. For example, "Open the boot on the number plate that adds to 6" was a search task. Users would have to navigate outside to the car-park and search for a car with the correct arithmetic on the number plate (Figure 8.12 (left)). Once they found the correct number plate, they had to action the instruction correctly (Figure 8.12 (right)). This requires the users to remember the tasks that are written on the whiteboard.

Figure 8.12: Search tasks are given to users based on arithmetic problems. For instance, "Open the boot on the number plate that adds to 6" is a search task. They must evaluate the number plates of the cars in the car-park until they find the correct one (left). Once they identify the correct number plate, they must action the corresponding instruction (i.e. open the boot) from the whiteboard (right).

Prior User Testing. Earlier versions of the Virtual Mechanic application have been through two prior user tests at the correctional facility where the exploratory study takes place. One group of five learners had previously been exposed to the system, and many of the changes based on their feedback and observed interactions have been integrated into the system described above. The resulting system is used in the exploratory study described in the following section.

8.4 Exploratory Study

We were presented with the opportunity to explore the efficacy of the method that emerged from our prior research, namely the structured analysis of physiological data against other dimensions of data in IVRL environments. We are also able to take this opportunity to investigate how IVRL applications might be evaluated in the future, and what kind of requirements will emerge based on conducting evaluations in real-world learning environments.

The nature of the environment our evaluation takes place in comes with inherent instabilities, more so than arguably most other learning environments. When plans are made with the correctional facility to conduct certain lessons, there can often be changes due to events external to the classroom. They can not be certain which prisoners will be showing up to classes. Our study was planned to be conducted with three separate groups of learners, each group having six learners. In total, we were able to evaluate three groups with six, five, and four learners in each, giving us a total of 15.

Each of the three class sessions were run the same way where the learners would, at any given time, be in one of three conditions: 1) immersive Virtual Mechanic application, 2) tablet-based activities, or 3) taking a break, talking, or giving feedback. Due to the environment, tight regulations, and consequential protocols, we are only able at this stage of the project to take a certain amount of equipment into the correctional facility. In total, we had two Virtual Mechanic systems and two tablets for each session. Therefore, if there were more than four learners, they were either passively observing, or interacting with the staff from the MMS, ARL, or ourselves.

The general aim for each session was to have two learners in the Virtual Mechanic application, two learners on the tablets, and two learners observing. The longer the exposure the better, but due to time and content constraints, we settled on approximately 15 minute segments after which time the learners would rotate their activities. Once again, due to the environment, and somewhat chaotic nature of each session, it was difficult to maintain each session in this manner, though they did approximate this flow of execution.

8.4.1 Study Design

One of the more specific goals of the work in this thesis is to investigate the value of emotional engagement with IVRL environments, and what that engagement means in the context of learning processes. We take this opportunity to test our previous findings in a real-world learning environment. In the previous two chapters, we presented a structured analysis which dichotomises a users' experimental conditions into three measurable dimensions of data: 1) physiological, 2) psychological, and 3) observational. In this exploratory study, we measure dimensions (1) and (3). We once again utilise the physiological measure of emotional response through electrodermal activity (EDA) and heart rate (HR) data, and we observe users achievements and environmental behaviour. The measures we used and the procedure we followed are described in the proceeding sections.

Participants. As previously described, our participants are prisoners at the Otago Correctional Facility who voluntarily commit to classes held at the facility.

8.4.2 Apparatus and Measures

We utilise three different methods of measurement within this exploratory study: 1) we sense EDA and HR data using an Empatica E4 wristband, 2) we record participants' achievements using the immersive Virtual Mechanic system, and 3) two investigators observe the sessions and record events as they are observed.

Physiological Data. As in our previous investigation, we employ the Empatica E4 wristband[¶] for measuring EDA, HR, accelerometer, and skin temperature. It also contains an internal clock storing unix-based UTC timestamps.

Achievements. Participants are given at least one exposure session to the immersive Virtual Mechanic VR system. The implementation of the tasks were described earlier. When a participant completes a part of a task, or a whole task, the system is programmed to store the timestamp of the achievement, and which achievement was completed.

Observations. This measurement is complex but is a requirement due to the nature of the study. We need to know which participants are currently active in which tasks. Two investigators have notebooks and pens with which they record environmental events throughout the session. For instance, the times in which participants would begin the VR tasks, tablet tasks, or if they were just sitting around observing, are recorded by the observers. It is imperative that it is clear and known which participants are performing which tasks at any given time, especially because they are wearing physiological devices. It would be ideal to have a video camera set up for the duration of this study so observations

[¶] www.empatica.com/research/e4/

could be recorded retrospectively, but ethical considerations and correctional regulations prevented us from doing so.

8.4.3 Procedure

At the beginning of each session we introduced ourselves to the participants and described our investigation. We expressed our interests in exploring how VR technology can improve learning engagement and potentially learning outcomes. In total, there were five people running the session including two people from ARL, one person from MMS, and two researchers from the University of Otago. The two ARL representatives present were guiding the users of the Virtual Mechanic application if they needed help and taking notes on user interactions and feedback. The MMS representative was guiding the overall operation and was the primary contact for the visit. And the two researchers were conducting the exploratory study.

We described the details of the Empatica wristbands explaining what they measure, and why we use them. They were given the option of wearing the wristbands during the experiment if they were comfortable with it. If they were comfortable, we firmly attached the bracelets onto participants' wrists so they could not move around much. This way all participants would wear a bracelet for the entire duration of the class session.

Upon commencement of the session, participants split off into an activity. Two would use the immersive Virtual Mechanic application, two would use the tablets, and the remaining participants observed, or engaged in conversation with others. Participants were expected to spend approximately 15 minutes in the VR environment, though this varied depending on how long it took participants to complete the content, and how long they wanted to spend exploring the VR environment. After approximately 15 minutes, participants were rotated so the VR participants would move to the tablets, the tablet participants would observe, and those observing would move to the VR activity. This part of the procedure was flexible, and did not always execute this way. It was heavily dependent on the participants want, and how many participants were in each session. These rotations happened however until all participants had attempted all activities. In some cases, some participants attempted certain activities more than once. Upon completion of each session, we removed the bracelets and turned them off which stores the data on the bracelet as a separate session. Participants were thanked, and were escorted away.

8.5 Discussion

VR shows promise as an engaging method of educational content delivery, although some issues came to light throughout our investigation highlighting the difficulty of conducting robust and thorough educational VR research. All participants to attend the class attempted the Virtual Mechanic application with our desired time of exposure meeting our expectations with a mean overall time of 14.13 minutes. There were some differences however between the three groups using the application.

Of the three groups, one group had previous exposure to an earlier version of the Virtual Mechanic application. Furthermore, the groups had different abilities and were sitting on different levels of education. The content currently integrated into the VR application consisted of literacy and numeracy tasks at a level similar to that of the third group's education level. While this exposure was not designed to evaluate for learning outcomes, the participants' level of education was reflected in the time they spent completing the tasks. The first group of six had no previous experience, but were at a higher level of education already, and they had a mean exposure time of 13.16 minutes. The second group of five had previous experience with the system and were of a high level of education. They spent a mean time of 11.2 minutes in the system. The third group consisted of only four participants, and they had a lower level of literacy and numeracy education, and no previous experience with the system. Their mean exposure time was 19.25 minutes. Additional exposures were allowed although the majority of participants only had one exposure with only two participants, both from the first group, having an additional two exposure times each.

The tasks within the Virtual Mechanic application were designed to gauge how target users generally interact with the environment. Therefore, our analyses is not to be conducted in the context of learning gains, but rather in terms of general engagement with the environment. From here we will discuss the potential analyses that can be conducted from this exploratory study. We also give an overview of issues with the study conduct and the implications of these on future immersive educational studies. The data captured from the study will be made available to the project developers at ARL together with this thesis for any analyses they may wish to conduct to facilitate their developments.

Physiological Analyses. In the previous investigation we used our physiological measures of EDA and HR to successfully predict moments of insight and cognitive load. Given the exploratory nature of this evaluation, we were not measuring for moments of insight, and the conditions are arguably not conducive to insight learning. Cognitive load analysis however could be possible.

In our own work, we were able to fully control the content which users were exposed to, i.e. four-dimensional spatial constructs, and we had total control of the tasks we presented participants. In the case of the Virtual Mechanic application, at the time of this evaluation, it is only in the early stages of development and consists of a small number of simple tasks. Both literacy and numeracy tasks were all of approximately the same difficulty. With respect to our evaluation of cognitive load, the current state of the system does not facilitate the same analyses from our previous investigation.

The cognitive load analyses we conducted previously was based on three categories of task difficulty, i.e. easy, medium, and hard. Therefore, we could use a multi-class classifier to determine degrees of cognitive load through heart activity analyses. In the current approach, with only several tasks of similar difficulty, the analysis would have to be based on a binary classifier which gives less information about the current state of the participant in the environment, but would still tell us, perhaps, whether the participant has surpassed a particular level of cognitive load. Furthermore, the small number of tasks currently implemented would not provide enough achievement data for such a classification. Consequently, a comprehensive numerical analysis of the physiological data is not possible.

Observations. Several observations were made throughout the study. We required observation to ensure we knew which participants were active in which activities, but in particular it was important we knew when they were active in the immersive VR application. In addition to our main observations, we noticed several issues with the study conduct that affect robust analyses.

Of the outside personnel present, two were heavily involved with the system development. They were tending to the two VR applications to guide the users should any issues or questions arise. A lot of the time however they would provide more instruction than was required, rather than letting the participants attempt to complete the tasks themselves. One consequence of this is the participants' break in focus. Often they might ask a question about the interface and the helpers would answer the question but then continue to tell them what to do next. This prevents the participant from becoming engaged with the task at hand.

Furthermore, throughout most of the sessions, the ambient noise in the classroom was quite loud and distracting. The participants would often be prompted by other classmates in the room resulting in a break in focus. This is exemplary of a typical classroom environment, and presents a further issue with such studies.

Implications on Future Educational VR Studies. At the end of the previous chapter we discussed the implications of our structured analysis, and its potential uses in the context of educational VR developments. We have been able to show here that it is possible, in a real-world scenario, to implement the method for the purposes of a structured analysis. Unfortunately, the study we conducted here was too unstructured and, with various observed confounding factors present, did not provide robust enough data for thorough analyses.

The first key lesson learned is with respect to the tasks that system users are expected to complete. These should be well structured against a known hierarchy of expected difficulty. In the context of immersive VR systems, tasks should also be self-directed, and self-explanatory as much as possible. If a user constantly requires help from an external source, the effectiveness of the virtual environment will be hindered.

We were able to collect physiological data in the correct manner however there were many sources of external stimuli participants were exposed to. Measurements of EDA and HR are highly responsive to the sympathetic nervous system (SNS) and the more variables present in an environment means more possible effects on the SNS. This in turn means less reliance on the effects of stimuli from the virtual environment. Therefore, it's important to try and maintain a controlled environment for the duration of a given exposure to allow users to achieve a focused state. This issue is highly relevant for the users' sense of presence as it is well established that external stimuli of a distracting nature causes a decrease in the users' sense of presence [113, 172]. We established in our previous work that reported presence approximated emotional engagement and correlated highly with achievement levels.

An issue related to users' engagement factor and sense of presence is exposure time. Once again, the stage of development limited the exposure time we could provide participants with. It was sufficient for the purposes of exploring potential class session structures and, for our purposes, whether or not our methodology holds real-world validity. Longer exposure times are however desirable for maximising engagement with material. Users should be allowed to gain momentum in their learning experiences so they can develop strategies and construct their knowledge accordingly. If they are interrupted, their momentum can be broken as in any environment. A further reason for longer exposure times is to allow for more robust physiological analyses. Particularly in the context of system evaluation, it can be difficult to establish patterns in users' physiological states with short exposure times.

A qualitative dimension of measurement will be highly beneficial in future evaluations of educational VR systems. One of the natural outcomes of the learning process is that learners will ask questions. Any feedback that is given at a point in time should be noted and ideally integrated into any analyses. This requires a more complex qualitative analysis, though it could provide important insights, particularly on individual learning efforts.

Insights for the Classroom. The various insights we attained from conducting this study are relevant not only for the highly regulated prison environment, but also for the more general educational classroom environment. However, applying these insights in evaluations and IVRL environments varies between these two scenarios.

A distraction free environment is desirable to help attain reliable physiological measures and also to facilitate a user's presence and engagement. In the context of the prison, it is a highly controlled but unpredictable environment with multiple distractions. Apart from noise from other learners in the space, the users' trust in their environment arose as an internal stimuli which could constantly distract one from their task. These issues can be addressed with smaller classroom sizes and potentially more isolated immersive learning spaces. Collaboration is a beneficial strategy in learning environments, however this should be implemented through telepresence systems, or in other elements of the classroom. In a more general classroom context with younger children, it is more likely going to be noise from other students and potentially performance anxieties that will detriment genuine engagement and presence attributes in IVRL environments. Smaller classroom sizes are not as common in regular schools, though an effort should be made to isolate the immersive learning space to facilitate students' presence and engagement, and if sensed measures are applied, to help with targeted physiological readings.

Controlled and isolated immersive environments are highly applicable for the prison learning environment. As VR systems and applications improve and become more portable (such as wireless technologies), more mobile solutions will be employed. The interactive element of IVRL systems will also have users holding controllers which means there is potential for abrupt and unexpected actions such as swinging controllers. These aspects are also applicable for student classrooms as children are also prone to such impulses, however in the volatile prison environment, avoiding any unnecessary incidence is a high priority.

If one is conducting an evaluation of an IVRL application, and they plan to incorporate physiological measures, then longer exposure times are desirable. There are multiple benefits to increased exposure time including an improved sense of engagement and presence. A drawback of long exposure to current immersive technologies is the potential for simulator sickness symptoms to emerge. As these are address, longer exposure times will be possible. In terms of prison environments, system designers should use a targeted persona to create content and develop the system. Frustration and boredom can begin to detriment the user experience and consequently, the learning outcomes. This is true of a classroom environment, however student users are more generally tolerant of a typical class length experience and can have their limit pushed at a greater rate.

Qualitative data is important in any evaluation. If a system is being evaluated on learners at a prison, it is likely evaluators are dealing with a smaller sample size than normal. This makes qualitative data collection and analysis more accessible. It could also be beneficial to tailor an experience to prison learners individual needs. Tailoring individual experiences is desirable in any case, however given the type of learner often found in the prison learning environment, short term benefits may be achieved by employing such an approach.

Conclusion. From the outcomes of this exploratory study we have drawn up preliminary guidelines for robust evaluations of educational virtual environments. The issues identified, and consequential solutions, are particularly relevant for structured analyses as conducted in the second investigation of this thesis (see chapter 6 and chapter 7). As a community that focuses on user experiences in VR systems, it is desirable to see their success not only in the education domain, but throughout multiple domains. It has been shown here that VR applications can be delivered even in logistically difficult classroom environments surrounded with ethical and regulatory constraints. Our methodologies have demonstrated potential to help with the success of these applications, and it is our hope they will be adopted in the further development as immersive VR proliferates.

Chapter 9

Discussion & Conclusion

Contents

9.1 Contributions **135**

9.2 Implications and Generalisability **139**

9.3 Limitations **141**

9.4 Future Work **142**

9.5 Conclusion **144**

Learning in Virtual Reality is the overarching theme of the research conducted throughout this thesis. Two primary research questions guided the thesis: 1) Does interactive involvement in IVRL environments provide unique learning experiences? And 2) How can we lead learners towards, measure, and analyse interactive learning experiences in IVRL environments? We addressed each of these questions through two branches of investigation, both of which stem from the original work of Arnold [8] and von Foerster [194]. We breathe life into the discussion of experience and knowledge, and at the same time address fundamental questions about learning in VR. In this chapter, we proceed to summarise the primary contributions, external validity and implications, limitations, and future work of this thesis.

9.1 Contributions

Three contributions were outlined in chapter 1: 1) the relationship between interaction and expertise in IVRL environments, 2) presence, insight learning, and the measurement and analysis of emotional engagement in IVRL environments, and 3) a real-world exploratory study. Contributions (1) and (2) comprise the backbone of the thesis and are the result of our two primary investigations. Our third contribution verifies the application of our second investigation, and allows us to provide guidelines for the future evaluation of IVRL environments.

9.1.1 Interaction and Expertise in Immersive VR

Our first investigation evaluated experiential value in IVRL environments for the comprehension of the fourth spatial dimension. This investigation was inspired by the early work of P. Arnold [7, 8] and the later work of von Foerster [194]. Based on their proposals, we implemented a principal research apparatus that was realised through two separate interactive and visual mediums. We built a system to correspond with their original proposal, and a modern immersive equivalent. The research apparatus is designed to allow users to interact with 4D constructs, namely a 4D cube (hypercube), with the intention that they will gain a deep form of comprehension ("begreifen") of the construct, or even of 4D space.

The first study of this investigation was conducted to validate the subject matter, the research apparatuses, and the measurements of subject matter comprehension. We were able to determine the operational validity of the research systems, and partial validity of the measures of comprehension. Based on the outcomes of the validation study, we improved our measurements, and tailored the system implementation for the second component of this investigation.

We designed our second study based on the knowledge argument where we allowed both subject matter experts and laypeople to interact with 4D space. The hypothesis was that by providing theoretical experts with interactive experience of 4D space, they will still achieve a knowledge gain despite the prior theoretical mastery of the subject. We also exposed laypeople users to the experience to assess the value of interaction for them in comparison to experts. The outcome revealed experts and laypeople both achieved significant gains in comprehension showing that interaction in this context has resulted in a unique form of learning that experts had apparently not previously experienced. Furthermore, we could show that while both experts and laypeople made significant gains in their comprehension, experts were able to improve their comprehension significantly more.

Together, these two studies contribute to the first overarching research question (RQ1) described at the beginning of this thesis: Does interactive involvement in IVRL environments provide unique learning experiences? This was motivated in part by the work from P. Arnold and von Foerster. We were able to investigate and draw conclusions contributing to the constructivist learning space, and the unique learning environments provided by VR technologies.

9.1.2 Presence, Insight, and Emotional Response in Immersive VR

The second investigation focuses on the sense of presence, insight moments (Aha! moments), and the physiological measure of emotional response in immersive virtual environments. This investigation is separated into two main analysis components: 1) we study the emergence of Aha! moments and the effect of presence on learning experiences in virtual environments, and 2) we present a methodology for the structured analysis of physiological signals representative of the emotional response of users within IVRL environments.

The same research apparatuses were utilised throughout this investigation. We used both systems, the stereo and immersive systems, validated in chapter 4 to investigate the effect of presence on learning experiences in VR environments. We conducted a preliminary study evaluating for usability differences between the two systems which resulted in no significant reported differences. Following the preliminary study, we ran an experiment where participants spent one full hour in each system (on separate days) experiencing 4D space.

In the first analysis component, we studied the effects of presence and moments of insight experienced by participants. Overall we discovered that the immersive system produces a higher sense of presence, and that this correlated with higher achievement levels. While participants reported similar numbers of Aha! moments between the two systems, those that reported insight moments in the immersive system produced significantly more solutions than those with zero reports in the same system. The same analysis within the stereo condition did not yield the same results and those reporting moments of insight solved similar numbers of solutions as those that did not.

The second analysis component of the investigation presents a methodology for the analysis of physiological (sensed) signals indicative of users' emotional states in our IVRL environment. The methodology includes an analysis of the sensed data against psychological (reported) and observational (observed) data. During the experiment, participants wore Empatica E4 wristbands to measure electrodermal activity (EDA), heart rate (HR), peripheral skin temperature, and accelerometer data. We utilised machine learning algorithms for the preprocessing and analysis of the EDA and HR data. We used participants' achievements to classify cognitive load labels based on easy, medium, and hard solution difficulties. We were able to establish with high confidence the cognitive load of users in our IVRL environment through their heart activity data. We could also determine with high confidence, based on the EDA data, approximate moments of insight (Aha! moments) experienced by participants. While there were no formal analyses of the physiological data against presence, there were observable differences between participants' electrodermal activity plots with respect to presence ratings of each condition indicating that higher levels of emotional engagement were achieved in the immersive system.

The outcomes of this investigation and analysis contributed to our second overarching research question (RQ2): How can we (1) lead user towards, (2) measure, and (3) analyse, interactive learning experiences in IVRL environments? By employing insight learning we could lead users to learning experiences based on their interactions with abstract 4D spatial constructs of varying difficulties. We utilised different measures as a lens to understanding users' experiences. Our goal was to provide a more continuous form of analysis with users' emotional response measures at the center, and considering them in conjunction with data of a different nature (i.e. psychological or observational data). We were able to build a methodology which can, and hopefully will, be applied in various domains which utilise immersive VR.

9.1.3 Educational Immersive VR in the Real World

The third investigation presented in this thesis is of an exploratory nature, and provided us the opportunity to verify the findings from our previous work. The investigation was grounded in a project we call the Virtual Mechanic which was developed over time from an initial prototypical proof of concept to a commercially developed product. Although the product is in its early stages of commercial development, we were able to capitalise on the user testing phase by conducting an exploratory study. We aimed to verify our prior work by implementing the method of measuring separate data dimensions as we previously identified, and then conducting the structured analyses in search of significant emotional responses to the immersive VR application.

We were able to verify the process of applying the measurements in the real world scenario, however several factors present throughout the study sessions introduced an instability to the data collection resulting in potential confounding effects. From the study, we were able to summarise a brief set of recommended guidelines should one attempt to implement the structured analysis we demonstrated in our second investigation. The primary requirements include: 1) a quiet controlled environment facilitating learners' focus and engagement, 2) longer exposure times to facilitate momentum, 3) well documented variable task difficulties, and 4) the addition of qualitative data analyses. It is imperative that requirements (1), (2), and (3) are met for such a study, and requirement (4) will potentially improve the quality of the structured analysis, particularly on an individual basis.

9.2 Implications and Generalisability

Several implications are made apparent through our investigation. Firstly, we are able to support the current educational VR literature by showing that VR, and most notably immersive VR, facilitates learning. However, there are still outlying questions regarding effectiveness which need to be addressed through more longitudinal studies and across multiple subject matter topics. Beyond our more general finding of the usefulness of VR for learning, each of our investigations provided results with various implications.

Interaction Design. We have been able to more specifically demonstrate the unique potential of interactive experience in IVRL environments. This emphasises the importance of interaction design in educational VR systems and raises questions around which interaction metaphors are best suited for educational applications of different kinds. Our results also raise questions with respect to interaction design for users of different levels of expertise. While interaction in VEs is a highly studied field, our results suggest that theoretical expertise is of importance and that it should be a fundamental consideration in the design of IVRL interfaces.

In chapter 2 we discussed the work of Roussou et al. in which they similarly investigated the value of interaction in VR learning environments. They found that with children, interaction contributed to problem-solving skills, but not to conceptual understanding [151, 152]. Our results support the findings of Roussou et al. however we have found both quantitative and anecdotal evidence in our results indicating the improvement of users' comprehension and conceptual understanding of complex subject matter. Furthermore, our studies were conducted with adults as opposed to children which provides a further avenue for investigation and highlights questions around the value of interactivity for different groups of users learning different subject matter.

Our first investigation has provided more theoretical implications, and raised many questions with respect to evaluating experience and interactivity in VR environments. Furthermore, we implemented and applied measures of comprehension which worked for the purposes of this research, but this has also raised questions around subject matter assessment methods in educational applications. There is space within this research field for investigating how assessments can be integrated within educational VR systems, and what that means for traditional pre- and post-exposure assessment methods.

Presence, Learning, and Engagement. The second contribution of this thesis provides the more significant implications of our work. Through the investigation conducted in chapter 6 and chapter 7, we applied the systematic measurement and structured analyses of presence, learning, and emotional response.

The sense of presence, as a defining element of VR, must be achieved by a user. A user's psychological involvement, as a key factor of presence in VEs, should therefore be facilitated. This means that we need to ensure a user's emotional engagement with the VR environment. Learning, and in our case insight learning, also requires a user's engagement, particularly in problem-solving based learning environments such as the one we present throughout this work. It is therefore imperative that we are able to measure and analyse users' emotional states throughout their experiences in IVRL environments. If a user demonstrates diminishing engagement levels, it is likely that the virtual environment will become less effective. This is due to the consequential diminished sense of presence, and the learning outcomes are also likely to decrease in frequency and quality.

Our results are indicative of this relationship. The analytical process we utilise in our second investigation provides a method for determining participants' emotional states throughout their learning experiences in IVRL environments. That is, the analysis of sensed physiological signals against observed and reported dimensions of data. We discussed the direct implications of our method in subsection 7.5.3. A practical implication of these results is the production of IVRL environments with dynamic content based on the users' emotional states with the intention of maximising the users' engagement levels. This would result in a maximised involvement and sense of presence as well as continuous engagement with the subject matter for optimising learning outcomes.

The implications of our analysis go beyond educational applications. Given, once again, that presence is a defining element of virtual reality, it is arguably important to measure the impact of any virtual environment on the users' emotional engagement. Previous work has utilised physiological measures including EDA to measure presence in VR environments. Much of this work analyses the EDA data producing a composite value of users' emotional states which is then correlated with presence conditions. This is a useful measure of the overall emotional state of a user, however, many details pertinent to a user's emotional responses are unintentionally omitted from analyses. In addition to overall measures of emotional states, our method allows for the identification and analysis of emotional events throughout an experience presenting opportunities for investigations of a multitude of phenomena in various application contexts. Training and instruction, entertainment, mental health treatment, and the psychological and behavioural sciences are just some of the areas in which our work is applicable.

9.3 Limitations

We have identified several drawbacks of the approaches taken throughout the research in this thesis. One recurring limitation throughout the experiments in this thesis was the lack of exposure time participants had within our interactive IVRL environment. Educational research benefits from longitudinal studies due to the often longitudinal nature of learning. While we have been able to identify learning processes and assess learning comprehension, we could potentially find more informative results if users had more exposures. In Arnold's original study proposal [8], he proposed (seemingly arbitrarily) that a time of 25 hours could be required to achieve his deep non-verbal comprehension. It is possible that exposure times in that order may be required for that kind of fundamentally complete internal representation of 4D space to be achieved (begreifen). The impact of a lack of exposure time emerged as an apparent issue to us in our real-world case study, although there is room for improvement in terms of users' acceptance of VR over long exposure times. We also had participants experience simulator sickness symptoms in our second investigation over a one hour long study.

A further recurring limitation throughout this work is the lack of formal qualitative analyses of data acquired throughout the experiments. In the first investigation, we collected extensive interview data from semi-structured interviews. While the primary and a secondary investigator iterated the interview transcriptions and used them to support the quantitative results, they were not analysed in a formal qualitative analysis. Similarly, in our second investigation we collected participants' responses to various questions pertinent to the phenomena under investigation, and used this data as anecdotal supporting evidence. We also recorded video and audio footage of the environments that participants acted in. This data was not used at all. We did not conduct the formal qualitative analyses of these data because it is not our domain of expertise and we were rather focused on the quantitative results. We acknowledge the value of qualitative analyses and how they could contribute in the area of investigation addressed in this thesis. This data could be particularly useful in the realm of education given the impact of individual characteristics on learning processes.

We used two different systems delivering the interactive experience of 4D space. One was the desktop stereo system on which the user manipulated a hypercube using six dials on a board, and the other was the immersive system where a user manipulated a hypercube by rotating two HTC Vive controllers. We conducted a study in chapter 4 to test the operational validity of the two systems, and in chapter 6, we conducted a study evaluating each interface for usability where no significant differences were found. Despite our efforts to test the equivalency of the interfaces, it is still likely the two interfaces had differing effects on performance outcomes. We can say that the differences do not likely confound our results, but a more in-depth exploration of interface differences using our subject matter and apparatus setups could provide interesting insights for the field of interaction design in the context of IVRL interfaces.

The final limitation worth noting is with respect to the continuous physiological measure of EDA and HR data. These measures are entirely objective and, for the most part, cannot be controlled by the individuals being measured. While it is an objective measurement, it is still based on the subjective experience of the individual. The measures of EDA and HR employed in this work are responsive to the sympathetic nervous system, so it is possible that arbitrary phasic events in the EDA data are in fact representative of individuals' own thoughts (i.e. stressful stimuli in one's personal life). This is something that is virtually impossible to control for. In general, it is safe to assume that these 'false' events are not frequent. This provides further incentive to include qualitative analyses in studies such as the ones presented in this thesis. Data reported in interviews, or particularly observed through video capture, could further inform the physiological measurement of emotional responses, and could perhaps assist in verifying sub-conscious emotional events.

9.4 Future Work

Our exploration of interaction in IVRL environments has emphasised the need for a framework to guide the structured analyses of interaction metaphors which will in turn assist in the design process of IVRL applications. We discussed earlier that different interactive mediums and approaches are likely required not just for different subject matter, but also for different learners based on their theoretical expertise. As VR becomes more prolific and educational VR gains momentum, a solution for this problem will be increasingly required.

It is often desirable for researchers to apply questionnaires measuring different phenomena throughout an experiment. The result of this in the context of VR studies is that a participant needs to be removed from the VR environment to complete the questionnaire leading to a consequential break of engagement with the environment and the users' sense of presence. Researchers have studied the possibility of integrating questionnaires with the environment while being able to maintain levels of presence. This work illuminates some interesting possibilities around utilising measures of emotional engagement in the context of mid-exposure breaks of presence and engagement, virtual environment-based questionnaires, and what effect such breaks might have on other dimensions of users' engagement in VR applications.

We have discussed at multiple points the contribution that qualitative analyses can make to studies such as those conducted in this thesis. Interview data is commonly applied in qualitative analyses in order to extract key themes based on peoples' perceptions on their experiences. Objectively observed data, i.e. by video capture, presents another form of data that could benefit from analysis using qualitative methods. In the context of learning, it could provide much insight into learners' processes in different environmental and topical contexts and if combined with data also giving individual context, we can learn a lot about learning and education practices. This suggestion extends beyond education and learning in VR environments and can be applied across general immersive VR experiences.

Simulator sickness symptoms began to emerge in our fourth study due to the one hour long exposure time. In addition to the environmental exposure time, participants were exerting continuous cognitive efforts to solve problems, and to comprehend the abstract phenomenon presented to them. This resulted in notable simulator sickness symptoms wherein two out of the 24 immersive cases, participants had to exit the system for a period of time. This would undoubtedly have an impact on physiological readings in the system due to the stress imposed on the sympathetic nervous system. When we examine the SSQ results from the second study (chapter 5) and the fourth study (chapter 7), we can see the same tendency in reported symptoms. Oculo-motor symptoms had consistently higher reports, and the same two individual symptoms ("eye strain" and "fullness of the head") had the highest ratings in both studies. This informs the area of investigation focusing on solving simulator sickness issues. It is important that these issues are addressed if VR technology is to be adopted on a large scale in any application space.

More generally, the methods demonstrated in this work show promise for facilitating the improvement of users' experiences in VR applications. For instance, the entertainment and gaming industry relies on users' emotional engagement. It is their aim to create experiences that will not be forgotten. The set of conventions that underlie game development as we know it do not all necessarily apply to VR game development. Our approach to measuring users' emotional experiences will be a useful tool as the industry works towards a new framework for immersive development. It can be integrated as a part of the development process to validate users' emotional engagement with their content. Training applications often work on hard skills that are learned through drill and repeat practices. In some cases, training can mean providing a trainee with situational experiences. VR has the ability to produce such experiences for trainees that are difficult and/or expensive to simulate in the real world. An example of such an experience is an emergency room situation where a person in critical condition is rushed into the hospital. These experiences will be most valuable if the trainees are engaged with the environment and genuine emotional responses are generated. It is the hope that the approach we have presented in this work will contribute to these, and many other immersive VR application spaces.

9.5 Conclusion

Immersive Virtual Reality has many unique benefits that are advantageous for learning practices. Complex interactive experiences are one such benefit afforded by VR environments. Throughout this work, we have revitalised the historic theoretical inquiry of P. Arnold and H. von Foerster that inspired us to pursue this investigation. It was their discussion of experiential impact on learning that sparked our investigation. We have found supporting evidence of their suppositions that experiential practice with subject matter is a necessary component of the learning process due to the unique form of learning comprehension it provides. We have also provided insight on the relationship between interaction and expertise in immersive VR learning environments. We hope these insights will contribute to the discussion on how educational content should be implemented in VR, and the general design principals behind immersive VR learning systems, particularly interaction metaphors.

The engagement of users in immersive environments is a key factor in the success of any VR application. Not only is engagement required for effective content delivery, but it is also a key part of the presence construct which itself is a defining component of VR. We have been able to demonstrate a method in the form of a structured analysis which uses physiological measures representative of emotional response to successfully predict elements of learning processes in VR environments. Namely, we could predict the cognitive effort of participants, and also emerging moments of insight as reported by learners. This measure of emotional response can be an effective tool for measuring engagement and the psychological experiences of users in immersive VR environments. This measure contributes to traditional subjective and objective measurements of users' experiences in VR, therefore expanding our perceptions as researchers, and allowing us to further improve the experiences we are passionate about delivering.

Bibliography

[1] Alberdi, A., Aztiria, A., and Basarab, A. (2016). Towards an automatic early stress recognition system for office environments based on multimodal measurements: A review. *Journal of Biomedical Informatics*, 59:49–75.

[2] Alghamdi, M., Regenbrecht, H., Hoermann, S., and Swain, N. (2017). Mild stress stimuli built into a non-immersive virtual environment can elicit actual stress responses. *Behaviour & Information Technology*, 36(9):913–934.

[3] Allison, D. and Hodges, L. F. (2000). Virtual Reality for Education? In *Proceedings of the ACM Symposium on Virtual Reality Software and Technology*, VRST '00, pages 160–165, New York, NY, USA. ACM.

[4] Alshaer, A., Regenbrecht, H., and O'Hare, D. (2017). Immersion factors affecting perception and behaviour in a virtual reality power wheelchair simulator. *Applied Ergonomics*, 58:1–12.

[5] Anderson, P., Rothbaum, B. O., and Hodges, L. F. (2003). Virtual reality exposure in the treatment of social anxiety. *Cognitive and Behavioral Practice*, 10(3):240–247.

[6] Andreatta, P. B., Maslowski, E., Petty, S., Shim, W., Marsh, M., Hall, T., Stern, S., and Frankel, J. (2010). Virtual Reality Triage Training Provides a Viable Solution for Disaster-preparedness. *Academic Emergency Medicine*, 17(8):870–876.

[7] Arnold, P. (1971). Experiencing the Fourth Spatial Dimension. volume 70, pages 201–215. The Biological Computer Laboratory, Department of Electrical Engineering, University of Illinois, Urbana.

[8] Arnold, P. (1972). A Proposal for a Study of the Mechanisms of Perception of, and Formation of Internal Representations of, the Spatial Fourth Dimension. volume 71, pages 223–236. The Biological Computer Laboratory, Department of Electrical Engineering, University of Illinois, Urbana.

[9] Ash, I. K., Jee, B. D., and Wiley, J. (2012). Investigating Insight as Sudden Learning. *The Journal of Problem Solving*, 4(2).

[10] Bangor, A., Kortum, P., and Miller, J. (2009). Determining What Individual SUS Scores Mean: Adding an Adjective Rating Scale. *J. Usability Studies*, 4(3):114–123.

[11] Barfield, W., Baird, K. M., and Bjorneseth, O. J. (1998). Presence in virtual environments as a function of type of input device and display update rate. *Displays*, 19(2):91–98.

[12] Barré, J., Michelet, D., Truchot, J., Jolivet, E., Recanzone, T., Stiti, S., Tesnière, A., and Pourcher, G. (2019). Virtual Reality Single-Port Sleeve Gastrectomy Training Decreases Physical and Mental Workload in Novice Surgeons: An Exploratory Study. *Obesity Surgery*.

[13] Başer, M. and Durmuş, S. (2010). The Effectiveness of Computer Supported Versus Real Laboratory Inquiry Learning Environments on the Understanding of Direct Current Electricity among Pre-Service Elementary School Teachers. *Eurasia Journal of Mathematics, Science and Technology Education*, 6(1):47–61.

[14] Beck, J. G., Palyo, S. A., Winer, E. H., Schwagler, B. E., and Ang, E. J. (2007). Virtual Reality Exposure Therapy for PTSD Symptoms After a Road Accident: An Uncontrolled Case Series. *Behavior Therapy*, 38(1):39–48.

[15] Benedek, M. and Kaernbach, C. (2010). A continuous measure of phasic electrodermal activity. *Journal of Neuroscience Methods*, 190(1):80–91.

[16] Bhagat, K. K., Liou, W.-K., and Chang, C.-Y. (2016). A cost-effective interactive 3d virtual reality system applied to military live firing training. *Virtual Reality*, 20(2):127–140.

[17] Billinghurst, M. (2002). Augmented reality in education. *New horizons for learning*, 12(5).

[18] Billinghurst, M. and Duenser, A. (2012). Augmented Reality in the Classroom. *Computer*, 45(7):56–63.

[19] Biocca, F. and Levy, M. R. (1995). *Communication in the Age of Virtual Reality*. Routledge. Google-Books-ID: MzaMSbzcz6UC.

[20] Bliss, J. P., Tidwell, P. D., and Guest, M. A. (1997). The Effectiveness of Virtual Reality for Administering Spatial Navigation Training to Firefighters. *Presence: Teleoperators and Virtual Environments*, 6(1):73–86.

[21] Bouchard, S., Robillard, G., and Renaud, P. (2007). Revising the factor structure of the simulator sickness questionnaire. *Annual review of cybertherapy and telemedicine*, 5(Summer):128–137.

[22] Boud, A. C., Haniff, D. J., Baber, C., and Steiner, S. J. (1999). Virtual reality and augmented reality as a training tool for assembly tasks. In *1999 IEEE International Conference on Information Visualization (Cat. No. PR00210)*, pages 32–36.

[23] Bower, M., Howe, C., McCredie, N., Robinson, A., and Grover, D. (2014). Augmented Reality in education - cases, places and potentials. *Educational Media International*, 51(1):1–15.

[24] Bowman, D. and McMahan, R. (2007). Virtual Reality: How Much Immersion Is Enough? *Computer*, 40(7):36–43.

[25] Brooke, J. et al. (1996). Sus-a quick and dirty usability scale. *Usability evaluation in industry*, 189(194):4–7.

[26] Bystrom, K.-E., Barfield, W., and Hendrix, C. (1999). A Conceptual Model of the Sense of Presence in Virtual Environments. *Presence: Teleoperators and Virtual Environments*, 8(2):241–244.

[27] Carlin, A. S., Hoffman, H. G., and Weghorst, S. (1997). Virtual reality and tactile augmentation in the treatment of spider phobia: a case report. *Behaviour Research and Therapy*, 35(2):153–158.

[28] Cayley, A. (1894). *The collected mathematical papers of Arthur Cayley*, volume 7. The University Press.

[29] Chee, Y. S. and Hooi, C. M. (2002). C-VISions: socialized learning through collaborative, virtual, interactive simulations. pages 687–696. International Society of the Learning Sciences.

[30] Chen, P., Liu, X., Cheng, W., and Huang, R. (2017). A review of using Augmented Reality in Education from 2011 to 2016. In Popescu, E., Kinshuk, Khribi, M. K., Huang, R., Jemni, M., Chen, N.-S., and Sampson, D. G., editors, *Innovations in Smart Learning*, Lecture Notes in Educational Technology, pages 13–18. Springer Singapore.

[31] Cheng, A., Yang, L., and Andersen, E. (2017). Teaching Language and Culture with a Virtual Reality Game. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, CHI '17, pages 541–549, New York, NY, USA. ACM.

[32] Chih Hung Chen, Jie Chi Yang, Shen, S., and Ming Chang Jeng (2007). A Desktop Virtual Reality Earth Motion System in Astronomy Education. *Journal of Educational Technology & Society*, 10(3):289–304.

[33] Cinaz, B., Arnrich, B., La Marca, R., and Tröster, G. (2013). Monitoring of mental workload levels during an everyday life office-work scenario. *Personal and Ubiquitous Computing*, 17(2):229–239.

[34] Civelek, T., Ucar, E., Ustunel, H., and Aydin, M. K. (2014). Effects of a Haptic Augmented Simulation on K-12 Students Achievement and their Attitudes towards Physics. *Eurasia Journal of Mathematics, Science and Technology Education*, 10(6):565–574.

[35] Cohen, J. (1992). A power primer. *Psychological bulletin*, 112(1):155.

[36] Collins, J., Regenbrecht, H., and Langlotz, T. (2017). Visual coherence in mixed reality: A systematic enquiry. *Presence: Teleoperators and Virtual Environments*, 26(1):16–41.

[37] Collins, J., Regenbrecht, H., and Langlotz, T. (2018). Back to the future: Constructivist learning in virtual reality. In *Adjunct Proceedings of the IEEE International Symposium for Mixed and Augmented Reality 2018 (Unpublished)*.

[38] Collins, J., Regenbrecht, H., and Langlotz, T. (2019a). The uniqueness of experience for learning in immersive virtual reality. In *In International Journal of Human-Computer Studies (Submitted)*.

[39] Collins, J., Regenbrecht, H., Langlotz, T., Can, Y. S., Ersoy, C., and Butson, R. (2019b). Measuring cognitive load and insight: A methodology exemplified in a virtual reality learning context. In *ISMAR '19 (Accepted)*.

[40] Critchley, H. D. (2002). Review: Electrodermal Responses: What Happens in the Brain. *The Neuroscientist*, 8(2):132–142.

[41] Cruz, L. C. D. and Oliveira, J. C. D. (2016). A CAVE/Desktop Collaborative Virtual Environment for Offshore Oil Platform Training. In *2016 XVIII Symposium on Virtual and Augmented Reality (SVR)*, pages 178–182.

[42] Cruz-Neira, C., Sandin, D. J., and DeFanti, T. A. (1993). Surround-screen Projection-based Virtual Reality: The Design and Implementation of the CAVE. In *Proceedings of the 20th Annual Conference on Computer Graphics and Interactive Techniques*, SIGGRAPH '93, pages 135–142, New York, NY, USA. ACM. event-place: Anaheim, CA.

[43] Dalgarno, B. and Lee, M. J. W. (2010). What are the learning affordances of 3-D virtual environments? *British Journal of Educational Technology*, 41(1):10–32.

[44] Danek, A. H. and Wiley, J. (2017). What about False Insights? Deconstructing the Aha! Experience along Its Multiple Dimensions for Correct and Incorrect Solutions Separately. *Frontiers in Psychology*, 7:2077.

[45] De Corte, E. and Verschaffel, L. (1987). The effects of semantic and non-semantic factors on young children's solutions of elementary addition and subtraction word problems. In *Proceedings of the Eleventh Conference of the International Group for the Psychology of Mathematics Education, Montreal, July 1987. Volume II.*, pages 375–381. University of Montreal; Montreal.

[46] Dede, C., Salzman, M. C., and Loftin, R. B. (1996). ScienceSpace: virtual realities for learning complex and abstract scientific concepts. In *Proceedings of the IEEE 1996 Virtual Reality Annual International Symposium*, pages 246–252.

[47] Dewey, J. (1938). An as experience. *Education.*

[48] Diemer, J., Alpers, G. W., Peperkorn, H. M., Shiban, Y., and Mühlberger, A. (2015). The impact of perception and presence on emotional reactions: a review of research in virtual reality. *Frontiers in Psychology*, 6:26.

[49] Diemer, J., Mühlberger, A., Pauli, P., and Zwanzger, P. (2014). Virtual reality exposure in anxiety disorders: Impact on psychophysiological reactivity. *The World Journal of Biological Psychiatry*, 15(6):427–442.

[50] Difede, J. and Hoffman, H. G. (2002). Virtual Reality Exposure Therapy for World Trade Center Post-traumatic Stress Disorder: A Case Report. *CyberPsychology & Behavior*, 5(6):529–535.

[51] Donalek, C., Djorgovski, S. G., Cioc, A., Wang, A., Zhang, J., Lawler, E., Yeh, S., Mahabal, A., Graham, M., Drake, A., Davidoff, S., Norris, J. S., and Longo, G. (2014). Immersive and collaborative data visualization using virtual reality platforms. In *2014 IEEE International Conference on Big Data (Big Data)*, pages 609–614.

[52] Dunne, J. R. and McDonald, C. L. (2010). Pulse: A Model for Research and Development of Virtual-Reality Learning in Military Medical Education and Training. *Military Medicine*, 175:25–27.

[53] Eibe, F., Hall, M., and Witten, I. (2016). The weka workbench. online appendix for" data mining: Practical machine learning tools and techniques. *Morgan Kaufmann.*

[54] Eleftheria, C. A., Charikleia, P., Iason, C. G., Athanasios, T., and Dimitrios, T. (2013). An innovative augmented reality educational platform using Gamification to enhance lifelong learning and cultural education. In *Information, Intelligence, Systems and Applications (IISA), 2013 Fourth International Conference on*, pages 1–5. IEEE.

[55] Felnhofer, A., Kothgassner, O. D., Hetterle, T., Beutl, L., Hlavacs, H., and Kryspin-Exner, I. (2014). Afraid to Be There? Evaluating the Relation Between Presence, Self-Reported Anxiety, and Heart Rate in a Virtual Public Speaking Task. *Cyberpsychology, Behavior, and Social Networking*, 17(5):310–316.

[56] Felnhofer, A., Kothgassner, O. D., Schmidt, M., Heinzle, A.-K., Beutl, L., Hlavacs, H., and Kryspin-Exner, I. (2015). Is virtual reality emotionally arousing? Investigating five emotion inducing virtual park scenarios. *International Journal of Human-Computer Studies*, 82:48–56.

[57] Finkelstein, S., Nickel, A., Barnes, T., and Suma, E. A. (2010). Astrojumper: Motivating Children with Autism to Exercise Using a VR Game. In *CHI '10 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '10, pages 4189–4194, New York, NY, USA. ACM.

[58] Fowler, C. (2015). Virtual reality and learning: Where is the pedagogy? *British Journal of Educational Technology*, 46(2):412–422.

[59] Freeman, J., Avons, S. E., Meddis, R., Pearson, D. E., and Ijsselsteijn, W. (2000). Using Behavioral Realism to Estimate Presence: A Study of the Utility of Postural Responses to Motion Stimuli. *Presence: Teleoperators & Virtual Environments*, 9(2):149–164.

[60] Freina, L. and Ott, M. (2015). A Literature Review on Immersive Virtual Reality in Education: State Of The Art and Perspectives. *eLearning & Software for Education*, (1).

[61] Frommel, J., Weber, M., Rogers, K., Brich, J., Besserer, D., Bradatsch, L., Ortinau, I., Schabenberger, R., Riemer, V., and Schrader, C. (2015). Integrated Questionnaires: Maintaining Presence in Game Environments for Self-Reported Data Acquisition. In *Proceedings of the 2015 Annual Symposium on Computer-Human Interaction in Play - CHI PLAY '15*, pages 359–368, London, United Kingdom. ACM Press.

[62] Gallagher, A. G. and Cates, C. U. (2004). Virtual reality training for the operating room and cardiac catheterisation laboratory. *The Lancet*, 364(9444):1538–1540.

[63] Garau, M., Slater, M., Pertaub, D.-P., and Razzaque, S. (2005). The Responses of People to Virtual Humans in an Immersive Virtual Environment. *Presence: Teleoperators and Virtual Environments*, 14(1):104–116.

[64] Garcia-Palacios, A., Hoffman, H., Carlin, A., Furness, T. A., and Botella, C. (2002). Virtual reality in the treatment of spider phobia: a controlled study. *Behaviour Research and Therapy*, 40(9):983–993.

[65] Gardner, H. (1987). *The mind's new science: A history of the cognitive revolution*. Basic books.

[66] Goedicke, D., Li, J., Evers, V., and Ju, W. (2018). VR-OOM: Virtual Reality On-rOad Driving siMulation. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, CHI '18, pages 165:1–165:11, New York, NY, USA. ACM.

[67] Gomes de Sá, A. and Zachmann, G. (1999). Virtual reality as a tool for verification of assembly and maintenance processes. *Computers & Graphics*, 23(3):389–403.

[68] Greco, A., Valenza, G., Lanata, A., Scilingo, E. P., and Citi, L. (2016). cvxEDA: A Convex Optimization Approach to Electrodermal Activity Processing. *IEEE Transactions on Biomedical Engineering*, 63(4):797–804.

[69] Grubert, J., Langlotz, T., Zollmann, S., and Regenbrecht, H. (2017). Towards Pervasive Augmented Reality: Context-Awareness in Augmented Reality. *IEEE Transactions on Visualization and Computer Graphics*, 23(6):1706–1724.

[70] Hanna, D., David, I., and Francisco, B. (2010). *The Nature of Learning: Using Research To Inspire Practice*. OECD publishing.

[71] Harman, J., Brown, R., and Johnson, D. (2017). Improved Memory Elicitation in Virtual Reality: New Experimental Results and Insights. In *Conference on Human-Computer Interaction*. IFIP.

[72] Hart, S. G. and Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. In Hancock, P. A. and Meshkati, N., editors, *Advances in Psychology*, volume 52 of *Human Mental Workload*, pages 139–183. North-Holland.

[73] Hauptman, H. (2010). Enhancement of spatial thinking with Virtual Spaces 1.0. *Computers & Education*, 54(1):123–135.

[74] Hedberg, J. and Alexander, S. (1994). Virtual Reality in Education: Defining Researchable Issues. *Educational Media International*, 31(4):214–220.

[75] Held, R. and Durlach, N. (1991). Telepresence, time delay and adaptation. *Pictorial communication in virtual and real environments*, pages 232–246.

[76] Hendrix, C. and Barfield, W. (1996). The Sense of Presence within Auditory Virtual Environments. *Presence: Teleoperators and Virtual Environments*, 5(3):290–301.

[77] Hew, K. F. and Cheung, W. S. (2010). Use of three-dimensional (3-D) immersive virtual worlds in K-12 and higher education settings: A review of the research. *British Journal of Educational Technology*, 41(1):33–55.

[78] Hodges, L. F., Rothbaum, B. O., Watson, B., Kessler, G. D., and Opdyke, D. (1996). A Virtual Airplane for Fear of Flying Therapy. *Virtual Reality Annual Symposium*.

[79] Hoermann, S., Santos, L. F. d., Morkisch, N., Jettkowski, K., Sillis, M., Devan, H., Kanagasabai, P. S., Schmidt, H., Kruger, J., Dohle, C., Regenbrecht, H., Hale, L., and Cutfield, N. J. (2017). Computerised mirror therapy with Augmented Reflection Technology for early stroke rehabilitation: clinical feasibility and integration as an adjunct therapy. *Disability and Rehabilitation*, 39(15):1503–1514.

[80] Hoffman, H. G., Patterson, D. R., and Carrougher, G. J. (2000). Use of virtual reality for adjunctive treatment of adult burn pain during physical therapy: a controlled study. *The Clinical Journal of Pain*, 16(3):244–250.

[81] Hoffman, H. G., Richards, T. L., Coda, B., Bills, A. R., Blough, D., Richards, A. L., and Sharar, S. R. (2004). Modulation of thermal pain-related brain activity with virtual reality: Evidence from fMRI. *NeuroReport: For Rapid Communication of Neuroscience Research*, 15(8):1245–1248.

[82] Huang, H.-M., Rauch, U., and Liaw, S.-S. (2010). Investigating learners attitudes toward virtual reality learning environments: Based on a constructivist approach. *Computers & Education*, 55(3):1171–1182.

[83] Hussain, M. Z., Sarfraz, M., and Shaikh, T. S. (2011). Shape preserving rational cubic spline for positive and convex data. *Egyptian Informatics Journal*, 12(3):231–236.

[84] Ijsselsteijn, W., de Ridder, H., Freeman, J., Avons, S. E., and Bouwhuis, D. (2001). Effects of Stereoscopic Presentation, Image Motion, and Screen Size on Subjective and Objective Corroborative Measures of Presence. *Presence: Teleoperators & Virtual Environments*, 10(3):298–311.

[85] IJsselsteijn, W.A. and Ridder, de H. (1998). Measuring temporal variations in presence. In *Papers from the workshop on presence in shared virtual environments*, Ipswich, UK. UCE (TM), Human Technology Interaction, TM-IPO-01.

[86] Izatt, E., Scholberg, K., and Kopper, R. (2014). Neutrino-KAVE: An immersive visualization and fitting tool for neutrino physics education. In *2014 IEEE Virtual Reality (VR)*, pages 83–84.

[87] Jackson, F. (1982). Epiphenomenal Qualia. *The Philosophical Quarterly (1950-)*, 32(127):127–136.

[88] Jex, H. R. (1988). Measuring Mental Workload: Problems, Progress, and Promises. In Meshkati, P. A. H. a. N., editor, *Advances in Psychology*, volume 52 of *Human Mental Workload*, pages 5–39. North-Holland.

[89] Johnson, A., Moher, T., Ohlsson, S., and Gillingham, M. (1999). The Round Earth Project-Collaborative VR for Conceptual Learning. *IEEE Comput. Graph. Appl.*, 19(6):60–69.

[90] Johnson, A., Roussos, M., Leigh, J., Vasilakis, C., Barnes, C., and Moher, T. (1998). The NICE project: learning together in a virtual world. In *Proceedings. IEEE 1998 Virtual Reality Annual International Symposium (Cat. No.98CB36180)*, pages 176–183.

[91] Johnston, E., Olivas, G., Steele, P., Smith, C., and Bailey, L. (2018). Exploring Pedagogical Foundations of Existing Virtual Reality Educational Applications: A Content Analysis Study. *Journal of Educational Technology Systems*, 46(4):414–439.

[92] Joolingen, W. R. V., Jong, T. D., and Dimitrakopoulou, A. (2007). Issues in computer supported inquiry learning in science. *Journal of Computer Assisted Learning*, 23(2):111–119.

[93] Kaufmann, H. and Schmalstieg, D. (2003). Mathematics and geometry education with collaborative augmented reality. *Computers & Graphics*, 27(3):339–345.

[94] Kaufmann, H., Schmalstieg, D., and Wagner, M. (2000). Construct3d: A Virtual Reality Application for Mathematics and Geometry Education. *Education and Information Technologies*, 5(4):263–276.

[95] Kennedy, R. S., Drexler, J. M., Compton, D. E., Stanney, K. M., Lanham, D. S., and Harm, D. L. (2003). Configural scoring of simulator sickness, cybersickness and space adaptation syndrome: Similarities and differences. *Virtual and adaptive environments: Applications, implications, and human performance issues*, page 247.

[96] Kennedy, R. S., Lane, N. E., Berbaum, K. S., and Lilienthal, M. G. (1993). Simulator sickness questionnaire: An enhanced method for quantifying simulator sickness. *The international journal of aviation psychology*, 3(3):203–220.

[97] Kerawalla, L., Luckin, R., Seljeflot, S., and Woolard, A. (2006). Making it real: exploring the potential of augmented reality for teaching primary school science. *Virtual Reality*, 10(3):163–174.

[98] Kesim, M. and Ozarslan, Y. (2012). Augmented Reality in Education: Current Technologies and the Potential for Education. *Procedia - Social and Behavioral Sciences*, 47:297–302.

[99] Khalfa, S., Isabelle, P., Jean-Pierre, B., and Manon, R. (2002). Event-related skin conductance responses to musical emotions in humans. *Neuroscience Letters*, 328(2):145–149.

[100] Kim, T. and Biocca, F. (1997). Telepresence via Television: Two Dimensions of Telepresence May Have Different Connections to Memory and Persuasion. *Journal of Computer-Mediated Communication*, 3(2).

[101] Koelstra, S., Muhl, C., Soleymani, M., Lee, J., Yazdani, A., Ebrahimi, T., Pun, T., Nijholt, A., and Patras, I. (2012). DEAP: A Database for Emotion Analysis ;Using Physiological Signals. *IEEE Transactions on Affective Computing*, 3(1):18–31.

[102] Kolb, D. A. (1984). *Experiential Learning: Experience as the Source of Learning and Development*. Prentice Hall Englewood Cliffs, NJ.

[103] Kotranza, A., Lind, D. S., Pugh, C. M., and Lok, B. (2009). Real-time in-situ visual feedback of task performance in mixed environments for learning joint psychomotor-cognitive tasks. In *2009 8th IEEE International Symposium on Mixed and Augmented Reality(ISMAR)*, pages 125–134.

[104] Kounios, J. and Beeman, M. (2009). The Aha! Moment: The Cognitive Neuroscience of Insight. *Current Directions in Psychological Science*, 18(4):210–216.

[105] Lawson, G., Salanitri, D., and Waterfield, B. (2016). Future directions for the development of virtual reality within an automotive manufacturer. *Applied Ergonomics*, 53:323–330.

[106] Lee, E. A.-L. and Wong, K. W. (2008). A Review of Using Virtual Reality for Learning. In Pan, Z., Cheok, A. D., Müller, W., and El Rhalibi, A., editors, *Transactions on Edutainment I*, Lecture Notes in Computer Science, pages 231–241. Springer Berlin Heidelberg, Berlin, Heidelberg.

[107] Lee, K. (2012). Augmented Reality in Education and Training. *TechTrends*, 56(2):13–21.

[108] Lee, K. M. (2004). Presence, Explicated. *Communication Theory*, 14(1):27–50.

[109] Lehmann, K. S., Ritz, J. P., Maass, H., Cakmak, H. K., Kuehnapfel, U. G., Germer, C. T., Bretthauer, G., and Buhr, H. J. (2005). A Prospective Randomized Study to Test the Transfer of Basic Psychomotor Skills From Virtual Reality to Physical Reality in a Comparable Training Setting. *Annals of Surgery*, 241(3):442–449.

[110] Lehner, V. D. and DeFanti, T. A. (1997). Distributed virtual reality: supporting remote collaboration in vehicle design. *IEEE Computer Graphics and Applications*, 17(2):13–17.

[111] Lele, A. (2013). Virtual reality and its military utility. *Journal of Ambient Intelligence and Humanized Computing*, 4(1):17–26.

[112] Lessiter, J., Freeman, J., Keogh, E., and Davidoff, J. (2001). A cross-media presence questionnaire: The ITC-Sense of Presence Inventory. *Presence*, 10(3):282–297.

[113] Liebold, B., Brill, M., Pietschmann, D., Schwab, F., and Ohler, P. (2017). Continuous Measurement of Breaks in Presence: Psychophysiology and Orienting Responses. *Media Psychology*, 20(3):477–501.

[114] Limniou, M., Roberts, D., and Papadopoulos, N. (2008). Full immersive virtual environment CAVETM in chemistry education. *Computers & Education*, 51(2):584–593.

[115] Liu, T.-Y. (2009). A context-aware ubiquitous learning environment for language listening and speaking. *Journal of Computer Assisted Learning*, 25(6):515–527.

[116] Lombard, M. and Ditton, T. (1997). At the heart of it all: The concept of presence. *Journal of computer-mediated communication*, 3(2).

[117] Lombard, M., Ditton, T. B., and Weinstein, L. (2009). Measuring presence: the temple presence inventory. In *Proceedings of the 12th Annual International Workshop on Presence*, pages 1–15.

[118] Lombard, M. and Jones, M. T. (2015). Defining Presence. In Lombard, M., Biocca, F., Freeman, J., IJsselsteijn, W., and Schaevitz, R. J., editors, *Immersed in Media: Telepresence Theory, Measurement & Technology*, pages 13–34. Springer International Publishing, Cham.

[119] Ma, D. and Ma, D. (2011). *Virtual Reality & Augmented Reality in Industry: The 2nd Sino-German Workshop*. Springer, Shanghai.

[120] Malbos, E., Rapee, R. M., and Kavakli, M. (2012). Behavioral Presence Test in Threatening Virtual Environments. *Presence: Teleoperators and Virtual Environments*, 21(3):268–280.

[121] Martin-Gutierrez, J., Luis Saorin, J., Contero, M., Alcaniz, M., Perez-Lopez, D. C., and Ortega, M. (2010). Design and validation of an augmented book for spatial abilities development in engineering students. *Computers and Graphics*, 34(1):77–91.

[122] Mayer, R. (1996). History of instructional psychology. *E., De Corte and FE Weinert (Eds)*.

[123] Meehan, M., Insko, B., Whitton, M., and Brooks, Jr., F. P. (2002). Physiological Measures of Presence in Stressful Virtual Environments. In *Proceedings of the 29th Annual Conference on Computer Graphics and Interactive Techniques*, SIGGRAPH '02, pages 645–652, New York, NY, USA. ACM.

[124] Meehan, M., Razzaque, S., Insko, B., Whitton, M., and Brooks, F. P. (2005). Review of Four Studies on the Use of Physiological Reaction as a Measure of Presence in Stressful Virtual Environments. *Applied Psychophysiology and Biofeedback*, 30(3):239–258.

[125] Melchiori Peruzza, A. P. P. and Zuffo, M. K. (2004). ConstruiRV: Constructing Knowledge Using the Virtual Reality. In *Proceedings of the 2004 ACM SIGGRAPH*

International Conference on Virtual Reality Continuum and Its Applications in Industry, VRCAI '04, pages 180–183, New York, NY, USA. ACM.

[126] Merchant, Z., Goetz, E. T., Cifuentes, L., Keeney-Kennicutt, W., and Davis, T. J. (2014). Effectiveness of virtual reality-based instruction on students' learning outcomes in K-12 and higher education: A meta-analysis. *Computers & Education*, 70:29–40.

[127] Mikropoulos, T. A. and Natsis, A. (2011). Educational virtual environments: A ten-year review of empirical research (1999–2009). *Computers & Education*, 56(3):769–780.

[128] Moore, K., Wiederhold, B. K., Wiederhold, M. D., and Riva, G. (2002). Panic and Agoraphobia in a Virtual World. *CyberPsychology & Behavior*, 5(3):197–202.

[129] Morozov, M., Tanakov, A., Gerasimov, A., Bystrov, D., and Cvirco, E. (2004). Virtual chemistry laboratory for school education. In *IEEE International Conference on Advanced Learning Technologies, 2004. Proceedings.*, pages 605–608. IEEE.

[130] Moshell, M. (1993). Three views of virtual reality: virtual environments in the US military. *Computer*, 26(2):81–82.

[131] Nguyen, C., DiVerdi, S., Hertzmann, A., and Liu, F. (2017). Vremiere: In-Headset Virtual Reality Video Editing. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, CHI '17, pages 5428–5438, New York, NY, USA. ACM.

[132] Nicholson, D. T., Chalk, C., Funnell, W. R. J., and Daniel, S. J. (2006). Can virtual reality improve anatomy education? A randomised controlled study of a computer-generated three-dimensional anatomical ear model. *Medical Education*, 40(11):1081–1087.

[133] Nowak, K. L. and Biocca, F. (2003). The Effect of the Agency and Anthropomorphism on Users' Sense of Telepresence, Copresence, and Social Presence in Virtual Environments. *Presence: Teleoperators and Virtual Environments*, 12(5):481–494.

[134] Peperkorn, H. M., Diemer, J., and Mühlberger, A. (2015). Temporal dynamics in the relation between presence and fear in virtual reality. *Computers in Human Behavior*, 48:542–547.

[135] Perez-Gracia, A. and Thomas, F. (2017). On Cayley's Factorization of 4d Rotations and Applications. *Advances in Applied Clifford Algebras*, 27(1):523–538.

[136] Perkins, D. (2000). *The Eureka Effect: The art and logic of breakthrough thinking*. The Eureka Effect: The art and logic of breakthrough thinking. W W Norton & Co, New York, NY, US.

[137] Pertaub, D.-P., Slater, M., and Barker, C. (2002). An Experiment on Public Speaking Anxiety in Response to Three Different Types of Virtual Audience. *Presence: Teleoperators and Virtual Environments*, 11(1):68–78.

[138] Piaget, J. (1970). *Psychology And Epistemology - Towards A Theory Of Knowledge*. Editions Denoel.

[139] Poh, M., Swenson, N. C., and Picard, R. W. (2010). A Wearable Sensor for Unobtrusive, Long-Term Assessment of Electrodermal Activity. *IEEE Transactions on Biomedical Engineering*, 57(5):1243–1252.

[140] Pouliquen-Lardy, L., Milleville-Pennel, I., Guillaume, F., and Mars, F. (2016). Remote collaboration in virtual reality: asymmetrical effects of task distribution on spatial processing and mental workload. *Virtual Reality*, 20(4):213–220.

[141] Raaen, K. and Kjellmo, I. (2015). Measuring Latency in Virtual Reality Systems. In Chorianopoulos, K., Divitini, M., Baalsrud Hauge, J., Jaccheri, L., and Malaka, R., editors, *Entertainment Computing - ICEC 2015*, Lecture Notes in Computer Science, pages 457–462. Springer International Publishing.

[142] Raja, D., Bowman, D. A., Lucas, J., and North, C. (2004). Exploring the Benefits of Immersion in Abstract Information Visualization. *Proceedings of Immersive Projection Technology Workshop.*, pages 61–69.

[143] Regenbrecht, H., Baratoff, G., and Wilke, W. (2005). Augmented reality projects in the automotive and aerospace industries. *IEEE Computer Graphics and Applications*, 25(6):48–56.

[144] Regenbrecht, H., McGregor, G., Ott, C., Hoermann, S., Schubert, T., Hale, L., Hoermann, J., Dixon, B., and Franz, E. (2011). Out of reach? - A novel AR interface approach for motor rehabilitation. In *2011 10th IEEE International Symposium on Mixed and Augmented Reality*, pages 219–228.

[145] Regenbrecht, H. and Schubert, T. (2002). Real and Illusory Interactions Enhance Presence in Virtual Environments. *Presence: Teleoperators & Virtual Environments*, 11(4):425–434.

[146] Regenbrecht, H. T., Schubert, T. W., and Friedmann, F. (1998). Measuring the Sense of Presence and its Relations to Fear of Heights in Virtual Environments. *Internation Journal of Human-Computer Interaction*, 10(3):233–249.

[147] Rizzo, A., Parsons, T. D., Lange, B., Kenny, P., Buckwalter, J. G., Rothbaum, B., Difede, J., Frazier, J., Newman, B., Williams, J., and Reger, G. (2011). Virtual Reality Goes to War: A Brief Review of the Future of Military Behavioral Healthcare. *Journal of Clinical Psychology in Medical Settings*, 18(2):176–187.

[148] Rothbaum, B., Hodges, L., and Kooper, R. (1997). Virtual reality exposure therapy. *The Journal of Psychotherapy Practice and Research*, 6(3):219–226.

[149] Rothbaum, B. O., Hodges, L., Smith, S., Lee, J. H., and Price, L. (2000). A controlled study of virtual reality exposure therapy for the fear of flying. *Journal of Consulting and Clinical Psychology*, 68(6):1020–1026.

[150] Rothbaum, B. O., Hodges, L. F., Ready, D., Graap, K., and Alarcon, R. D. (2001). Virtual reality exposure therapy for Vietnam veterans with posttraumatic stress disorder. *The Journal of Clinical Psychiatry*, 62(8):617–622.

[151] Roussou, M. (2004). Interactivity and Conceptual Learning in Virtual Environments for Children. In *CHI '04 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '04, pages 1049–1050, New York, NY, USA. ACM.

[152] Roussou, M., Oliver, M., and Slater, M. (2006). The virtual playground: an educational virtual reality environment for evaluating interactivity and conceptual learning. *Virtual Reality*, 10(3):227–240.

[153] Roussou, M. and Slater, M. (2017). Comparison of the Effect of Interactive versus Passive Virtual Reality Learning Activities in Evoking and Sustaining Conceptual Change. *IEEE Transactions on Emerging Topics in Computing*.

[154] Sacks, R., Perlman, A., and Barak, R. (2013). Construction safety training using immersive virtual reality. *Construction Management and Economics*, 31(9):1005–1017.

[155] Sagardia, M., Hertkorn, K., Hulin, T., Wolff, R., Hummell, J., Dodiya, J., and Gerndt, A. (2013). An interactive virtual reality system for on-orbit servicing. In *2013 IEEE Virtual Reality (VR)*, pages 1–1.

[156] Salovaara, H. (2005). An exploration of students' strategy use in inquiry-based computer-supported collaborative learning. *Journal of Computer Assisted Learning*, 21(1):39–52.

[157] Sanchez-Vives, M. V. and Slater, M. (2005). From presence to consciousness through virtual reality. *Nature Reviews Neuroscience*, 6(4):332–339.

[158] Sandkühler, S. and Bhattacharya, J. (2008). Deconstructing Insight: EEG Correlates of Insightful Problem Solving. *PLOS ONE*, 3(1):e1459.

[159] Santos, I. H. F. d., Soares, L. P., Carvalho, F., and Raposo, A. (2012). A Collaborative Virtual Reality Oil & Gas Workflow. *International Journal of Virtual Reality*, 11(1):1–13.

[160] Satava, R. M. (1995). Virtual reality and telepresence for military medicine. *Computers in Biology and Medicine*, 25(2):229–236.

[161] Schubert, T., Friedmann, F., and Regenbrecht, H. (2001). The Experience of Presence: Factor Analytic Insights. *Presence: Teleoperators and Virtual Environments*, 10(3):266–281.

[162] Schuemie, M. J., van der Straaten, P., Krijn, M., and van der Mast, C. A. (2001). Research on Presence in Virtual Reality: A Survey. *CyberPsychology & Behavior*, 4(2):183–201.

[163] Schwienhorst, K. (2002). Why Virtual, Why Environments? Implementing Virtual Reality Concepts in Computer-Assisted Language Learning. *Simulation & Gaming*, 33(2):196–209.

[164] Seymour, N. E., Gallagher, A. G., Roman, S. A., OBrien, M. K., Bansal, V. K., Andersen, D. K., and Satava, R. M. (2002). Virtual Reality Training Improves Operating Room Performance. *Annals of Surgery*, 236(4):458–464.

[165] Sheridan, T. B. (1992). Musings on telepresence and virtual presence. *Presence: Teleoperators & Virtual Environments*, 1(1):120–126.

[166] Shim, K.-C., Park, J.-S., Kim, H.-S., Kim, J.-H., Park, Y.-C., and Ryu, H.-I. (2003). Application of virtual reality technology in biology education. *Journal of Biological Education*, 37(2):71–74.

[167] Silva, M., Freitas, D., Neto, E., Lins, C., Teichrieb, V., and Teixeira, J. M. (2014). Glassist: using augmented reality on Google Glass as an aid to classroom management. In *Virtual and Augmented Reality (SVR), 2014 XVI Symposium on*, pages 37–44. IEEE.

[168] Simon, H. A. (1979). Information processing models of cognition. *Annual review of psychology*, 30(1):363–396.

[169] Skinner, B. F. (1954). The science of learning and the art of teaching. *Cambridge, Mass, USA*, 99:113.

[170] Slater, M. (2004). How colorful was your day? Why questionnaires cannot assess presence in virtual environments. *Presence*, 13(4):484–493.

[171] Slater, M., Pertaub, D., and Steed, A. (1999). Public speaking in virtual reality: facing an audience of avatars. *IEEE Computer Graphics and Applications*, 19(2):6–9.

[172] Slater, M. and Steed, A. (2000). A Virtual Presence Counter. *Presence: Teleoperators and Virtual Environments*, 9(5):413–434.

[173] Slater, M. and Usoh, M. (1993). Representations Systems, Perceptual Position, and Presence in Immersive Virtual Environments. *Presence: Teleoper. Virtual Environ.*, 2(3):221–233.

[174] Slater, M., Usoh, M., and Steed, A. (1994). Depth of presence in virtual environments. *Presence*, 3(2):130–144.

[175] Slater, M. and Wilbur, S. (1995). Through the looking glass world of presence: A framework for immersive virtual environments. In *FIVE*, volume 95, pages 1–20.

[176] Slater, M. and Wilbur, S. (1997). A Framework for Immersive Virtual Environments (FIVE): Speculations on the Role of Presence in Virtual Environments. *Presence: Teleoperators and Virtual Environments*, 6(6):603–616.

[177] Song, K. S. and Lee, W. Y. (2002). A virtual reality application for geometry classes. *Journal of Computer Assisted Learning*, 18(2):149–156.

[178] Sorathia, K., Sharma, K., Bhowmick, S., and Kamidi, P. (2017). Pragati - A Mobile Based Virtual Reality (VR) Platform to Train and Educate Community Health Workers. In *Conference on Human-Computer Interaction*. IFIP.

[179] Steuer, J. (1992). Defining virtual reality: Dimensions determining telepresence. *Journal of communication*, 42(4):73–93.

[180] Stone, R. J., Panfilov, P. B., and Shukshunov, V. E. (2011). Evolution of aerospace simulation: From immersive Virtual Reality to serious games. In *Proceedings of 5th International Conference on Recent Advances in Space Technologies - RAST2011*, pages 655–662.

[181] Stork, A., Sevilmis, N., Weber, D., Gorecky, D., Stahl, C., Loskyll, M., and Michel, F. (2012). Enabling virtual assembly training in and beyond the automotive industry. In *2012 18th International Conference on Virtual Systems and Multimedia*, pages 347–352.

[182] Stroud, K. J., Harm, D. L., and Klaus, D. M. (2005). Preflight virtual reality training as a countermeasure for space motion sickness and disorientation. *Aviation, space, and environmental medicine*, 76(4):352–356.

[183] Sun, K.-T., Chan, H.-T., and Meng, K.-C. (2010). Research on the application of virtual reality on arts core curricula. In *5th International Conference on Computer Sciences and Convergence Information Technology*, pages 234–239.

[184] Sveistrup, H. (2004). Motor rehabilitation using virtual reality. *Journal of Neuro-Engineering and Rehabilitation*, 1(1):10.

[185] Szalma, J. L., Hancock, P. A., Warm, J. S., Dember, W. N., and Parsons, K. S. (2006). Training for vigilance: Using predictive power to evaluate feedback effectiveness. *Human factors*, 48(4):682–692.

[186] Talaba, D. and Amditis, A. (2008). *Product engineering: tools and methods based on virtual reality*. Springer Science & Business Media.

[187] Taylor, S., Jaques, N., Chen, W., Fedor, S., Sano, A., and Picard, R. (2015). Automatic identification of artifacts in electrodermal activity data. In *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 1934–1937.

[188] Thorndike, E. L. (1922). *The psychology of arithmetic*. Macmillan.

[189] Topolinski, S. and Reber, R. (2010). Gaining Insight Into the “Aha” Experience. *Current Directions in Psychological Science*, 19(6):402–405.

[190] Tussyadiah, I. P., Wang, D., Jung, T. H., and tom Dieck, M. C. (2018). Virtual reality, presence, and attitude change: Empirical evidence from tourism. *Tourism Management*, 66:140–154.

[191] Urhahne, D., Nick, S., and Schanze, S. (2009). The Effect of Three-Dimensional Simulations on the Understanding of Chemical Structures and Their Properties. *Research in Science Education*, 39(4):495–513.

[192] Usoh, M., Catena, E., Arman, S., and Slater, M. (2000). Using Presence Questionnaires in Reality. *Presence: Teleoperators and Virtual Environments*, 9(5):497–503.

[193] Vollmer, M. (2015). A robust, simple and reliable measure of heart rate variability using relative RR intervals. In *2015 Computing in Cardiology Conference (CinC)*, pages 609–612.

[194] Von Foerster, H., von Glasersfeld, E., and Hejl, P. M. (1992). *Einführung in den Konstruktivismus*. Piper.

[195] Von Glasersfeld, E. (1991). An exposition of constructivism: Why some like it radical. In *Facets of systems science*, pages 229–238. Springer.

[196] Vygotsky, L. S. (1980). *Mind in society: The development of higher psychological processes*. Harvard university press.

[197] Welch, R. B. (1999). How Can We Determine if the Sense of Presence Affects Task Performance? *Presence: Teleoperators & Virtual Environments*, 8(5):574–577.

[198] Welch, R. B., Blackmon, T. T., Liu, A., Mellers, B. A., and Stark, L. W. (1996). The Effects of Pictorial Realism, Delay of Visual Feedback, and Observer Interactivity on the Subjective Sense of Presence. *Presence: Teleoperators and Virtual Environments*, 5(3):263–273.

[199] Wiederhold, B. K., Gevirtz, R., and Wiederhold, M. D. (1998). Fear of Flying: A Case Report Using Virtual Reality Therapy with Physiological Monitoring. *CyberPsychology & Behavior*, 1(2):97–103.

[200] Wiederhold, B. K., Jang, D. P., Kim, S. I., and Wiederhold, M. D. (2002). Physiological Monitoring as an Objective Tool in Virtual Reality Therapy. *CyberPsychology & Behavior*, 5(1):77–82.

[201] Winn, W. (1995). The Virtual Reality Roving Vehicle Project. *THE Journal (Technological Horizons In Education)*, 23(5):70.

[202] Winn, W. and Bricken, W. (1992). Designing Virtual Worlds for Use in Mathematics Education: The Example of Experiential Algebra. *Educational Technology*, 32(12):12–19.

[203] Winterbottom, C. and Blake, E. (2004). Designing a VR interaction authoring tool using constructivist practices. In *Proceedings of the 3rd international conference on Computer graphics, virtual reality, visualisation and interaction in Africa*, pages 67–71. ACM.

[204] Winterbottom, C. and Blake, E. (2008). Constructivism, virtual reality and tools to support design. In *Proceedings of the 7th ACM conference on Designing interactive systems*, pages 230–239. ACM.

[205] Witmer, B. G. and Singer, M. J. (1998). Measuring presence in virtual environments: A presence questionnaire. *Presence: Teleoperators and virtual environments*, 7(3):225–240.

[206] Wu, H.-K., Lee, S. W.-Y., Chang, H.-Y., and Liang, J.-C. (2013). Current status, opportunities and challenges of augmented reality in education. *Computers & Education*, 62:41–49.

[207] Wyk, E. v. (2006). Improving Mine Safety Training Using Interactive Simulations. pages 2454–2459. Association for the Advancement of Computing in Education (AACE).

[208] Yang, J. C., Chen, C. H., and Chang Jeng, M. (2010). Integrating video-capture virtual reality technology into a physically interactive learning environment for English learning. *Computers & Education*, 55(3):1346–1356.

[209] Yeh, A. (2004). VRMath: Knowledge Construction of 3d Geometry in Virtual Reality Microworlds. In *CHI '04 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '04, pages 1061–1062, New York, NY, USA. ACM.

[210] Youngblut, C. (1998). Educational Uses of Virtual Reality Technology. Technical Report IDA-D-2128, Institue For Defense Analyses Alexandria VA.

[211] Yuen, S., Yaoyuneyong, G., and Johnson, E. (2011). Augmented Reality: An Overview and Five Directions for AR in Education. *Journal of Educational Technology Development and Exchange (JETDE)*, 4(1).

[212] Zacharia, Z. C. (2007). Comparing and combining real and virtual experimentation: an effort to enhance students' conceptual understanding of electric circuits. *Journal of Computer Assisted Learning*, 23(2):120–132.

Appendix A

Igroup Presence Questionnaire Items

The following pages contain a copy of the IPQ questionnaire as presented to participants.

Igroup Presence Questionnaire (IPQ)

Q1. In the computer generated world I had a sense of “being there”

	-3	-2	-1	0	1	2	3	
Not at all	◯	◯	◯	◯	◯	◯	◯	Very much

Q2. Somehow I felt that the virtual world surrounded me.

	-3	-2	-1	0	1	2	3	
Fully disagree	◯	◯	◯	◯	◯	◯	◯	Fully agree

Q3. I felt like I was just perceiving pictures.

	-3	-2	-1	0	1	2	3	
Fully disagree	◯	◯	◯	◯	◯	◯	◯	Fully agree

Q4. I did not feel present in the virtual space.

	-3	-2	-1	0	1	2	3	
Did not feel	◯	◯	◯	◯	◯	◯	◯	Felt present

Q5. I had a sense of acting in the virtual space, rather than operating something from outside.

	-3	-2	-1	0	1	2	3	
Fully disagree	◯	◯	◯	◯	◯	◯	◯	Fully agree

Q6. I felt present in the virtual space

	-3	-2	-1	0	1	2	3	
Fully disagree	◯	◯	◯	◯	◯	◯	◯	Fully agree

Q7. How aware were you of the real world surrounding while navigating in the virtual world (i.e. sounds, room temperature, other people, etc.)?

	-3	-2	-1	0	1	2	3	
Extremely aware	○	○	○	○	○	○	○	Not aware at all

Q8. I was not aware of my real environment.

	-3	-2	-1	0	1	2	3	
Fully disagree	○	○	○	○	○	○	○	Fully agree

Q9. I still paid attention to the real environment.

	-3	-2	-1	0	1	2	3	
Fully disagree	○	○	○	○	○	○	○	Fully agree

Q10. I was completely captivated by the virtual world.

	-3	-2	-1	0	1	2	3	
Fully disagree	○	○	○	○	○	○	○	Fully agree

Q11. How real did the virtual world seem to you?

	-3	-2	-1	0	1	2	3	
Completely real	○	○	○	○	○	○	○	Not real at all

Q12. How much did your experience in the virtual environment seem consistent with your real world experience?

	-3	-2	-1	0	1	2	3	
Not consistent	○	○	○	○	○	○	○	Very consistent

Q13. How real did the virtual world seem to you?

	-3	-2	-1	0	1	2	3	
About as real as an imagined world	○	○	○	○	○	○	○	Indistinguishable from the real world

Q14. The virtual world seemed more realistic than the real world.

	-3	-2	-1	0	1	2	3	
Fully disagree	○	○	○	○	○	○	○	Fully agree

References:

1) Regenbrecht, H., & Schubert, T. (2002). Real and illusory interaction enhance presence in virtual environments. Presence: Teleoperators and Virtual Environments, 11(4), 425-434.
2) Schubert, T. (2003). The sense of presence in virtual environments: A three-component scale measuring spatial presence, involvement, and realness. Zeitschrift für Medienpsychologie, 15, 69-71.
3) Schubert, T., Friedmann, F., & Regenbrecht, H. (1999). Embodied Presence in Virtual Environments. In Ray Paton & Irene Neilson (Eds.), Visual Representations and Interpretations (pp. 269-278). London: Springer-Verlag.
4) Schubert, T., Friedmann, F., & Regenbrecht, H. (2001). The experience of presence: Factor analytic insights. Presence: Teleoperators and Virtual Environments, 10(3), 266-281.
5) Schubert, T., & Regenbrecht, H. (2002). Wer hat Angst vor virtueller Realität? Phobie, Therapie und Präsenz in virtuellen Welten [Who is afraid of virtual reality? Phobia, therapy and presence and virtual worlds]. In Bente, G. (Hrsg.), Virtuelle Realitäten [Virtual Realities], pp. 255-274. Göttingen: Hogrefe.

Appendix B

Simulator Sickness Questionnaire Items

The following page contains a copy of the SSQ items as presented to participants.

No______________ Date____________________

SIMULATOR SICKNESS QUESTIONNAIRE

Kennedy, Lane, Berbaum, & Lilienthal (1993)***

Instructions : Circle how much each symptom below is affecting you right now.

1. General discomfort	None	Slight	Moderate	Severe
2. Fatigue	None	Slight	Moderate	Severe
3. Headache	None	Slight	Moderate	Severe
4. Eye strain	None	Slight	Moderate	Severe
5. Difficulty focusing	None	Slight	Moderate	Severe
6. Salivation increasing	None	Slight	Moderate	Severe
7. Sweating	None	Slight	Moderate	Severe
8. Nausea	None	Slight	Moderate	Severe
9. Difficulty concentrating	None	Slight	Moderate	Severe
10. « Fullness of the Head »	None	Slight	Moderate	Severe
11. Blurred vision	None	Slight	Moderate	Severe
12. Dizziness with eyes open	None	Slight	Moderate	Severe
13. Dizziness with eyes closed	None	Slight	Moderate	Severe
14. *Vertigo	None	Slight	Moderate	Severe
15. **Stomach awareness	None	Slight	Moderate	Severe
16. Burping	None	Slight	Moderate	Severe

* Vertigo is experienced as loss of orientation with respect to vertical upright.

** Stomach awareness is usually used to indicate a feeling of discomfort which is just short of nausea.

Last version : March 2013

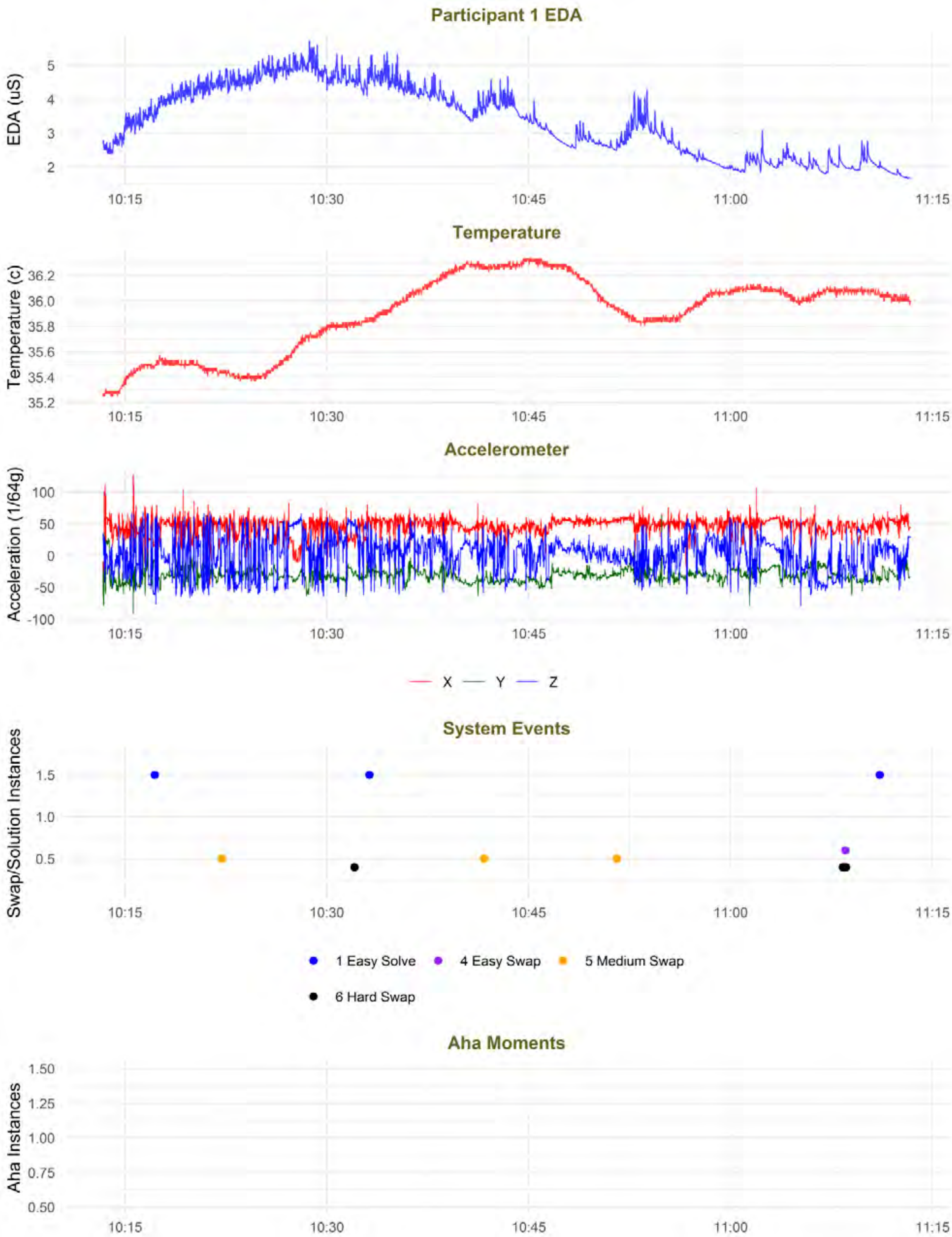
***Original version : Kennedy, R.S., Lane, N.E., Berbaum, K.S., & Lilienthal, M.G. (1993). Simulator Sickness Questionnaire: An enhanced method for quantifying simulator sickness. *International Journal of Aviation Psychology, 3*(3), 203-220.

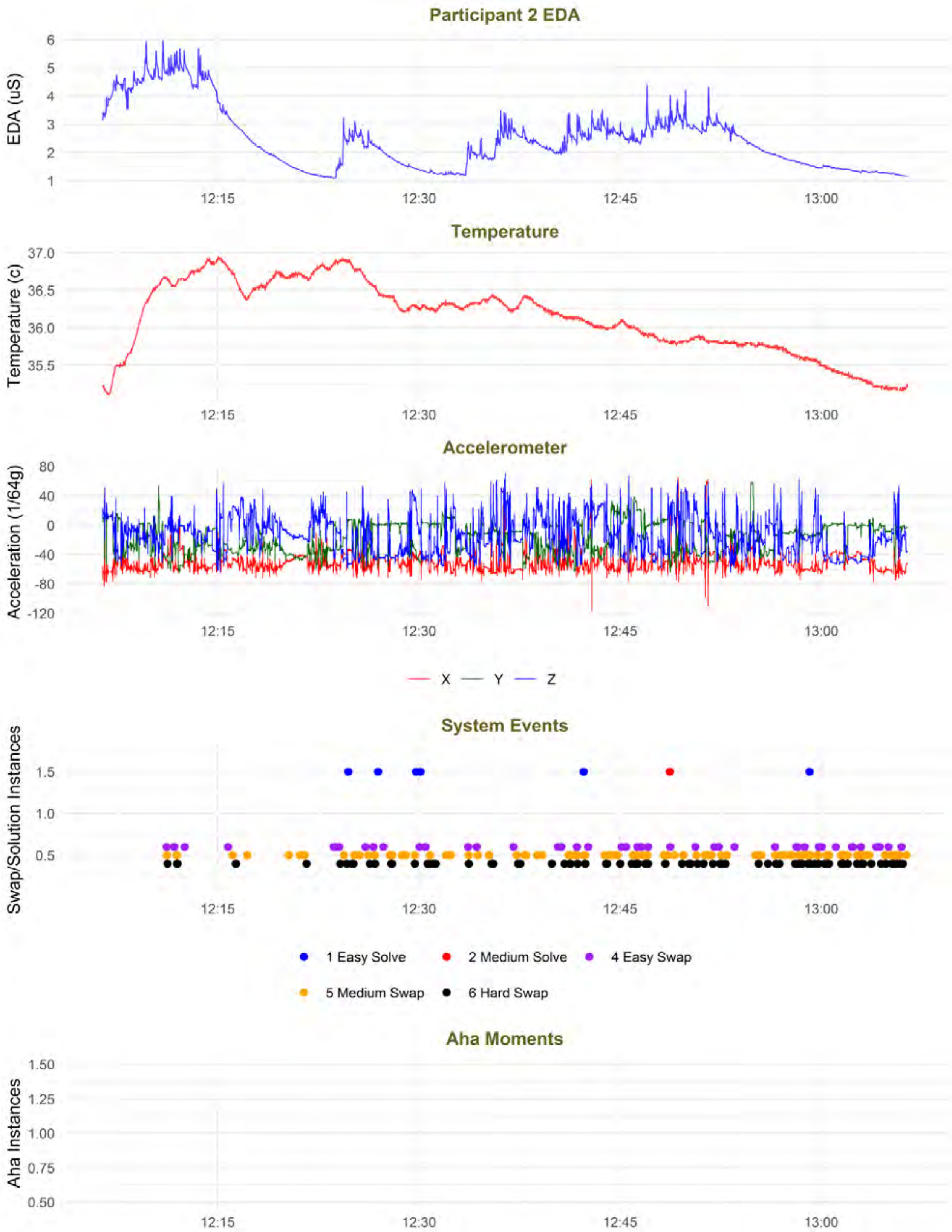
Appendix C

All Data Dimension Plots - Study Four

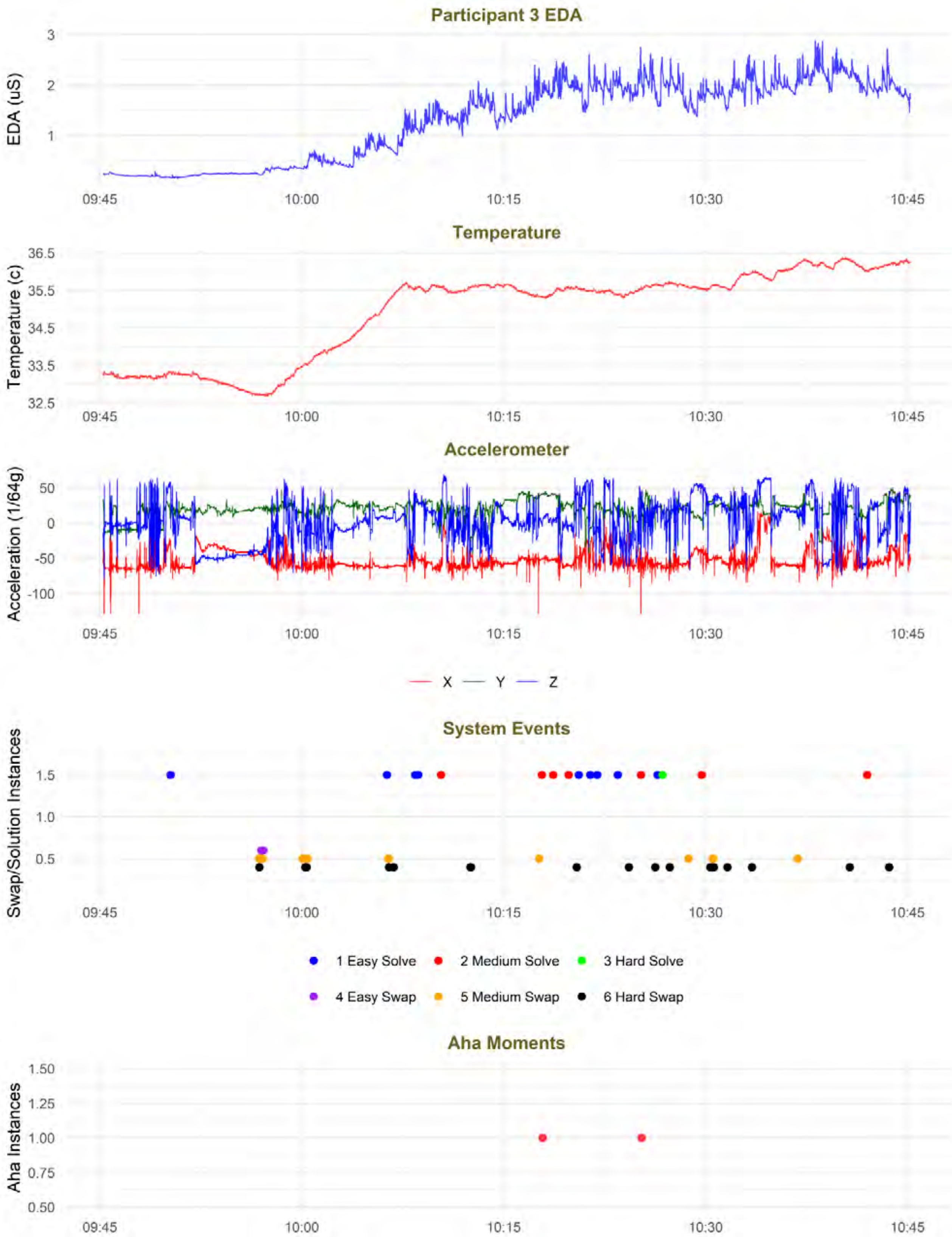
Below are the 24 plots generated for all participants' immersive conditions from the fourth study. The plots visualise the following data along a shared x-axis measured in time (timestamps):

- Electrodermal Activity (EDA) (E4 Wristband)
- Peripheral Skin Temperature (E4 Wristband)
- Accelerometer Data (E4 Wristband)
- Solution Events (Apparatus Recording)
- Ghostcube Switch Events (Apparatus Recording)
- Reported Aha! Moments (Apparatus Recording)

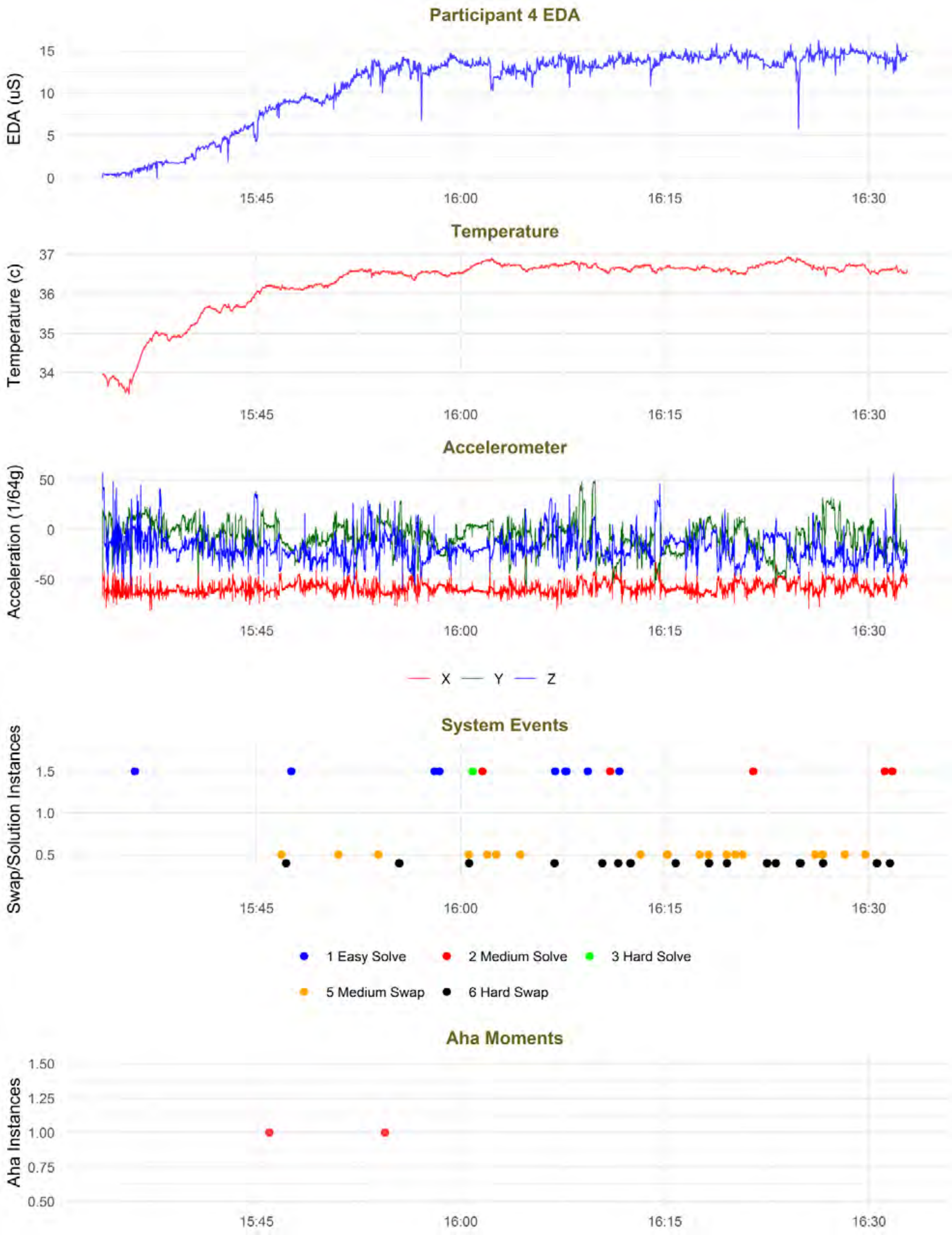

Participant 1 EDA
EDA (uS)
Temperature
Temperature (c)
Accelerometer
Acceleration (1/64g)
X
Y
Z
System Events
Swap/Solution Instances
1 Easy Solve
4 Easy Swap
5 Medium Swap
6 Hard Swap
Aha Moments
Aha Instances
10:15
10:30
10:45
11:00
11:15



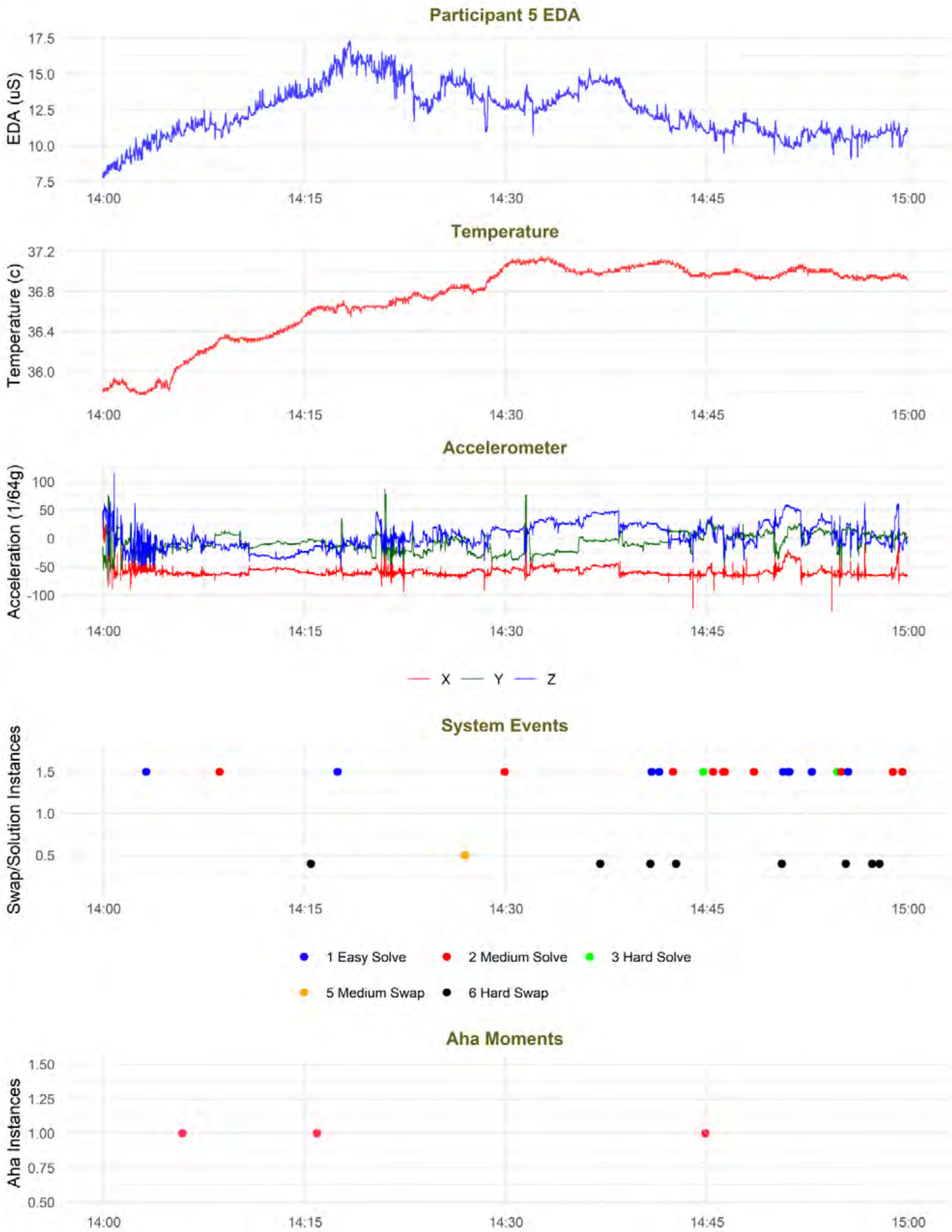
Participant 2 EDA
EDA (uS)
6
5
4
3
2
1
12:15
12:30
12:45
13:00
Temperature
Temperature (c)
37.0
36.5
36.0
35.5
12:15
12:30
12:45
13:00
Accelerometer
Acceleration (1/64g)
80
40
0
-40
-80
-120
12:15
12:30
12:45
13:00
X
Y
Z
System Events
Swap/Solution Instances
1.5
1.0
0.5
12:15
12:30
12:45
13:00
1 Easy Solve
2 Medium Solve
4 Easy Swap
5 Medium Swap
6 Hard Swap
Aha Moments
Aha Instances
1.50
1.25
1.00
0.75
0.50
12:15
12:30
12:45
13:00



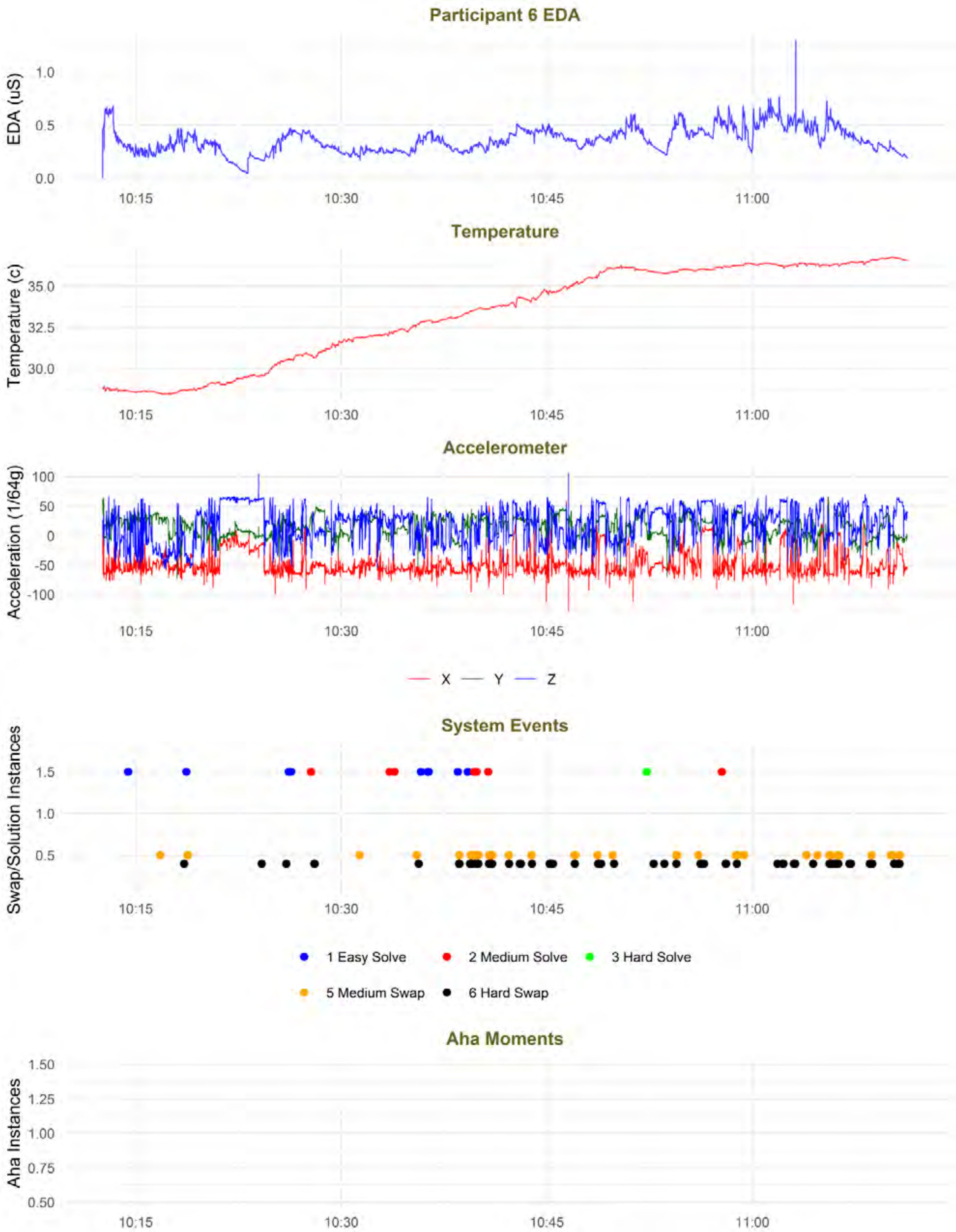
Participant 3 EDA
EDA (uS)
3
2
1
09:45
10:00
10:15
10:30
10:45
Temperature
Temperature (c)
36.5
35.5
34.5
33.5
32.5
Accelerometer
Acceleration (1/64g)
50
0
-50
-100
X
Y
Z
System Events
Swap/Solution Instances
1.5
1.0
0.5
1 Easy Solve
2 Medium Solve
3 Hard Solve
4 Easy Swap
5 Medium Swap
6 Hard Swap
Aha Moments
Aha Instances
1.50
1.25
1.00
0.75
0.50

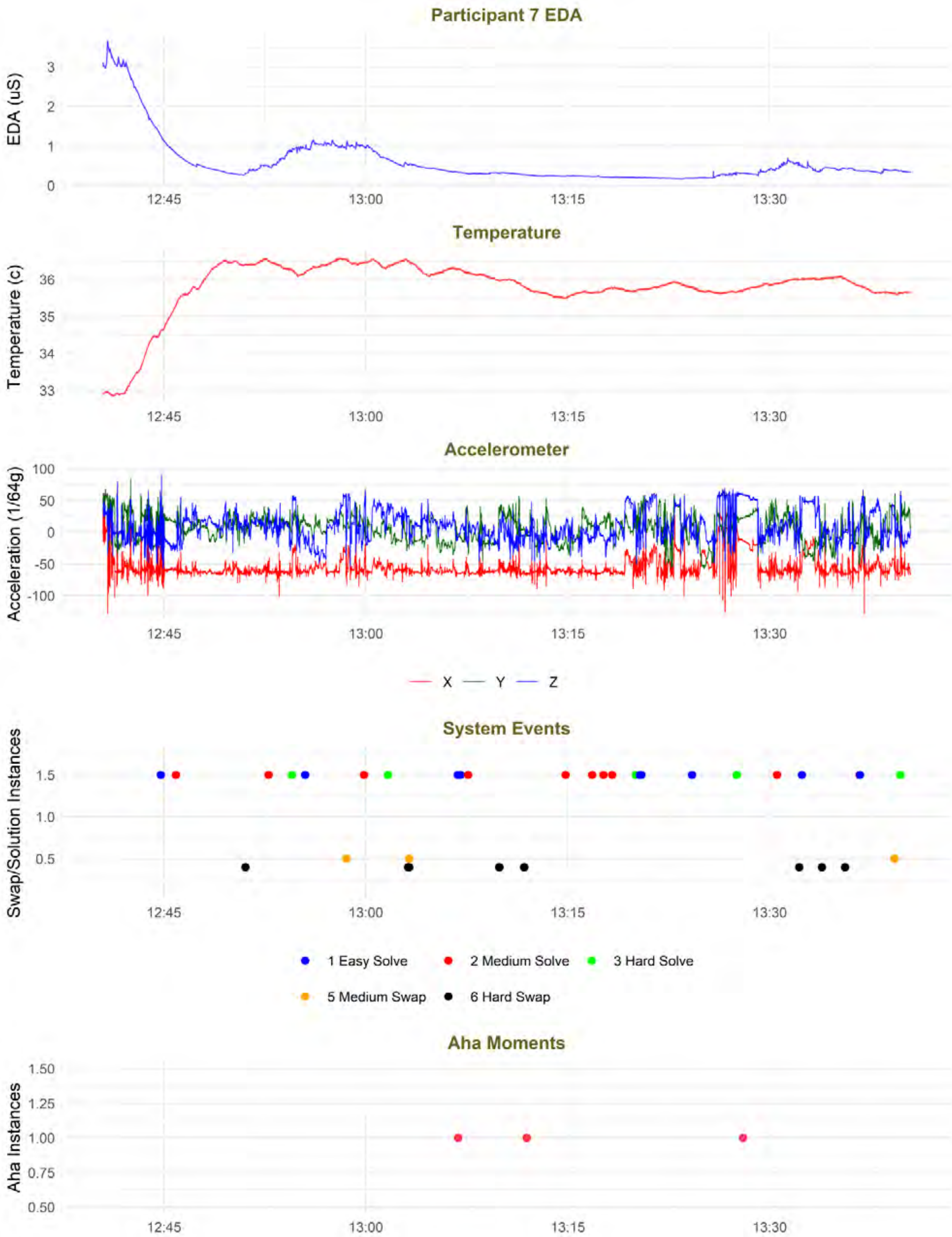


Participant 4 EDA
EDA (uS)
Temperature
Temperature (c)
Accelerometer
Acceleration (1/64g)
X
Y
Z
System Events
Swap/Solution Instances
1 Easy Solve
2 Medium Solve
3 Hard Solve
5 Medium Swap
6 Hard Swap
Aha Moments
Aha Instances
15:45
16:00
16:15
16:30

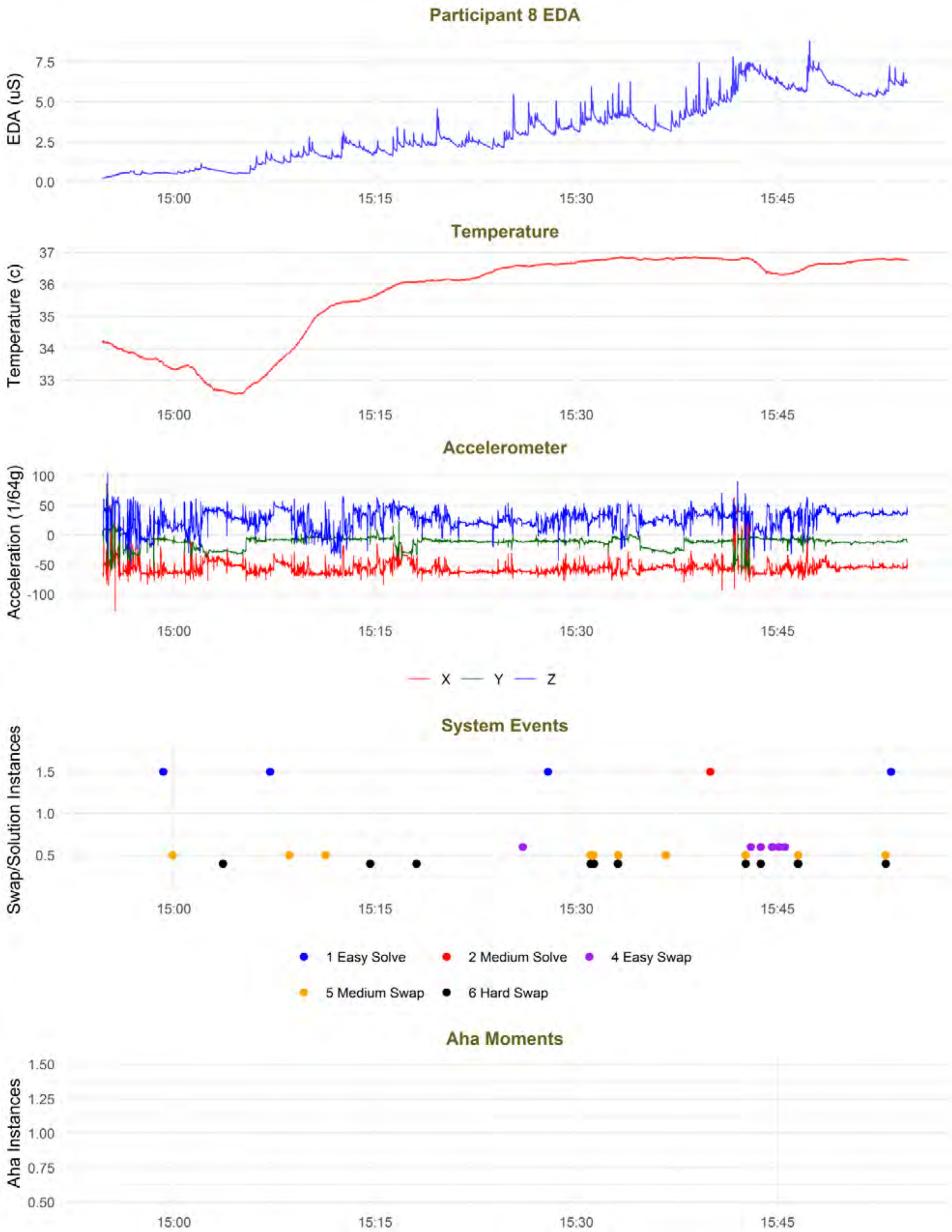


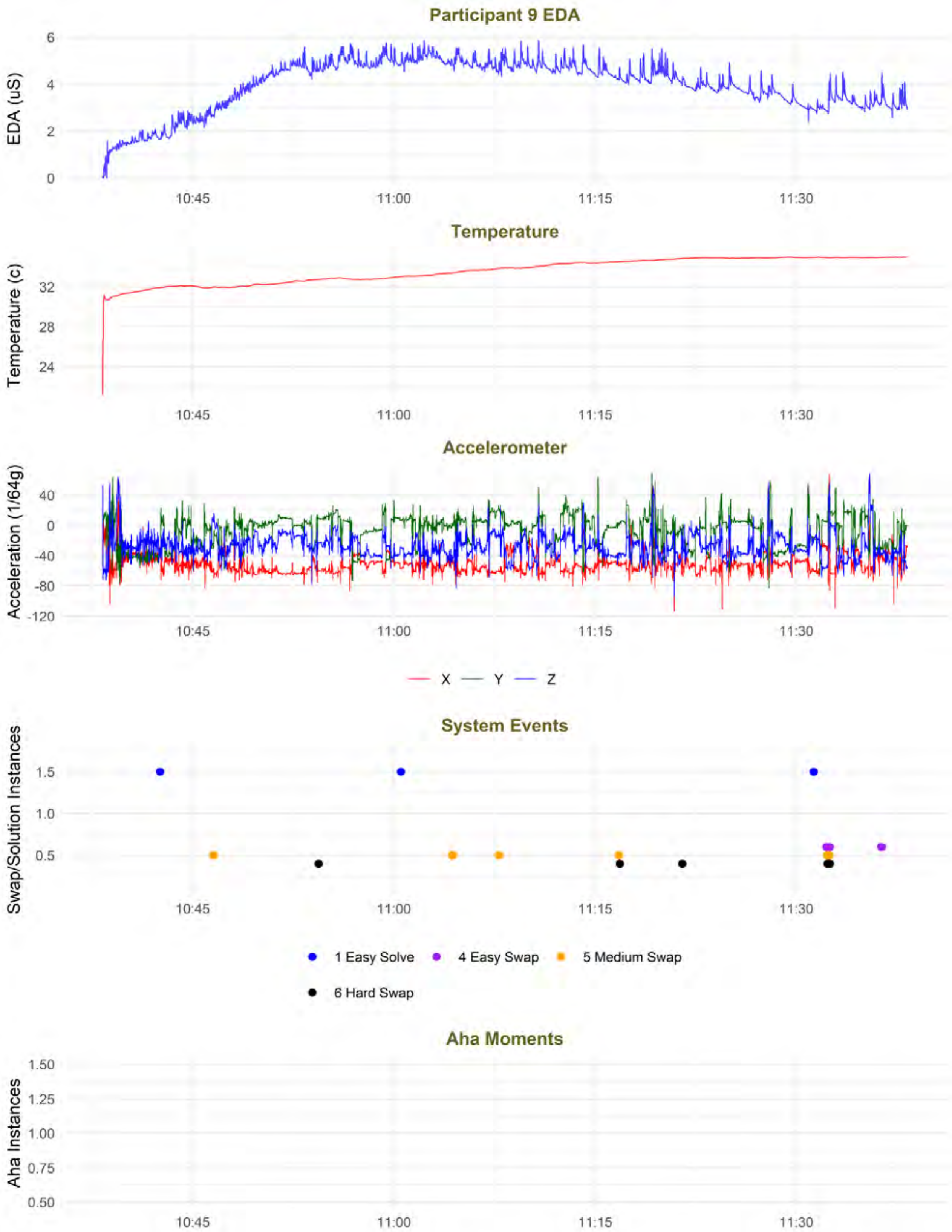
Participant 5 EDA
EDA (uS)
17.5
15.0
12.5
10.0
7.5
14:00
14:15
14:30
14:45
15:00
Temperature
Temperature (c)
37.2
36.8
36.4
36.0
Accelerometer
Acceleration (1/64g)
100
50
0
-50
-100
X
Y
Z
System Events
Swap/Solution Instances
1.5
1.0
0.5
1 Easy Solve
2 Medium Solve
3 Hard Solve
5 Medium Swap
6 Hard Swap
Aha Moments
Aha Instances
1.50
1.25
1.00
0.75
0.50

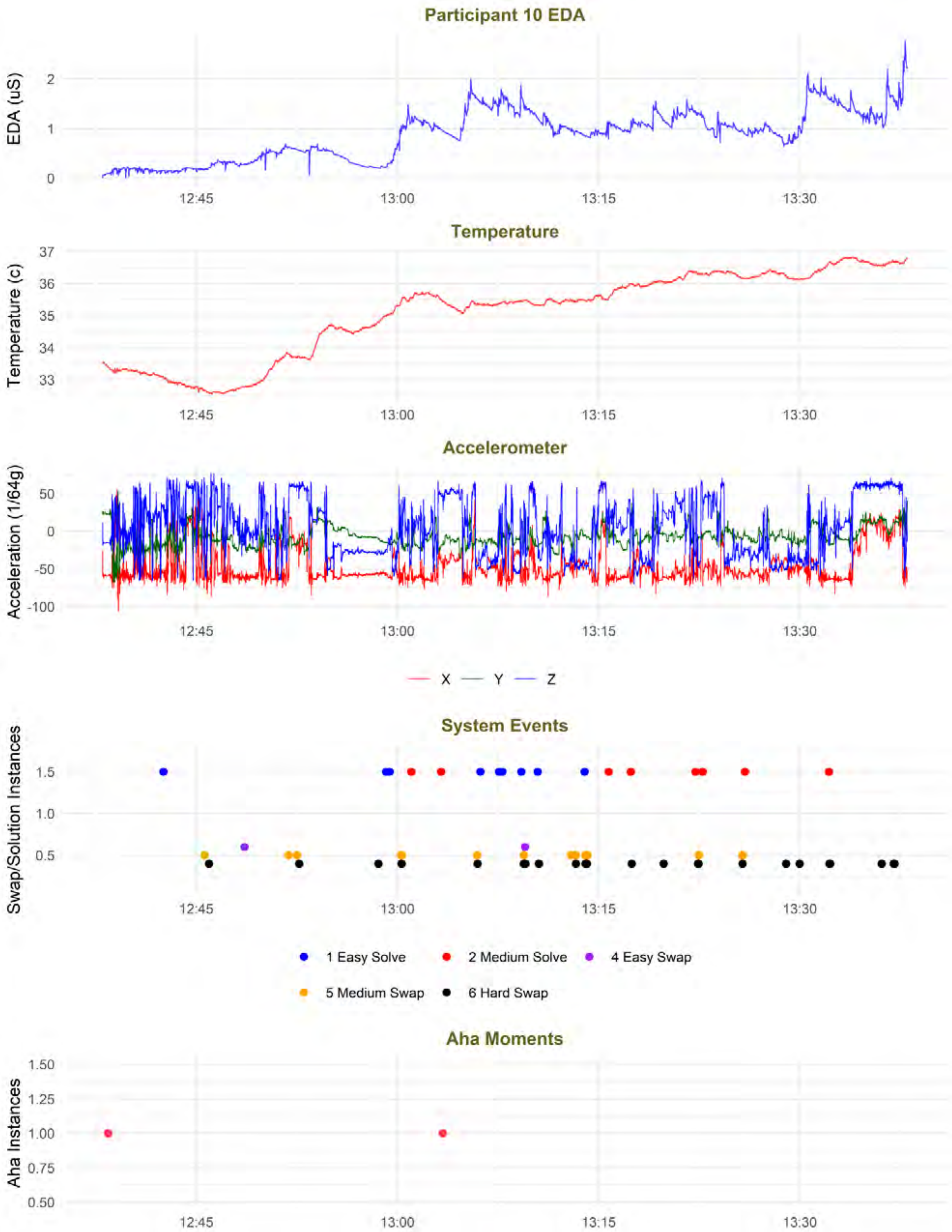

Participant 6 EDA
EDA (uS)
Temperature
Temperature (c)
Accelerometer
Acceleration (1/64g)
X
Y
Z
System Events
Swap/Solution Instances
1 Easy Solve
2 Medium Solve
3 Hard Solve
5 Medium Swap
6 Hard Swap
Aha Moments
Aha Instances



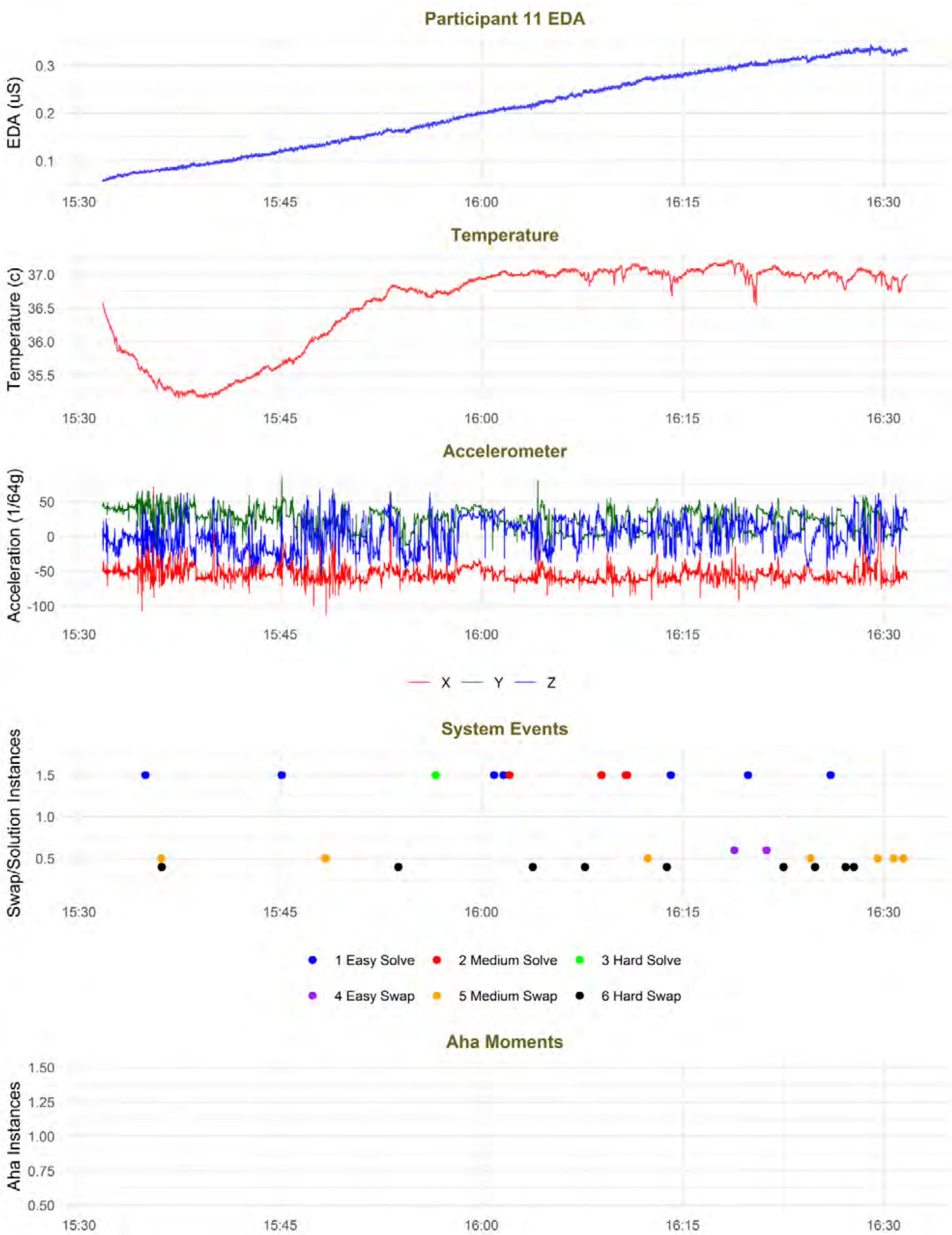
Participant 7 EDA
EDA (uS)
3
2
1
0
12:45
13:00
13:15
13:30
Temperature
Temperature (c)
36
35
34
33
12:45
13:00
13:15
13:30
Accelerometer
Acceleration (1/64g)
100
50
0
-50
-100
12:45
13:00
13:15
13:30
X
Y
Z
System Events
Swap/Solution Instances
1.5
1.0
0.5
12:45
13:00
13:15
13:30
1 Easy Solve
2 Medium Solve
3 Hard Solve
5 Medium Swap
6 Hard Swap
Aha Moments
Aha Instances
1.50
1.25
1.00
0.75
0.50
12:45
13:00
13:15
13:30


Participant 8 EDA
EDA (uS)
7.5
5.0
2.5
0.0
15:00
15:15
15:30
15:45
Temperature
Temperature (c)
37
36
35
34
33
15:00
15:15
15:30
15:45
Accelerometer
Acceleration (1/64g)
100
50
0
-50
-100
15:00
15:15
15:30
15:45
X
Y
Z
System Events
Swap/Solution Instances
1.5
1.0
0.5
15:00
15:15
15:30
15:45
1 Easy Solve
2 Medium Solve
4 Easy Swap
5 Medium Swap
6 Hard Swap
Aha Moments
Aha Instances
1.50
1.25
1.00
0.75
0.50
15:00
15:15
15:30
15:45

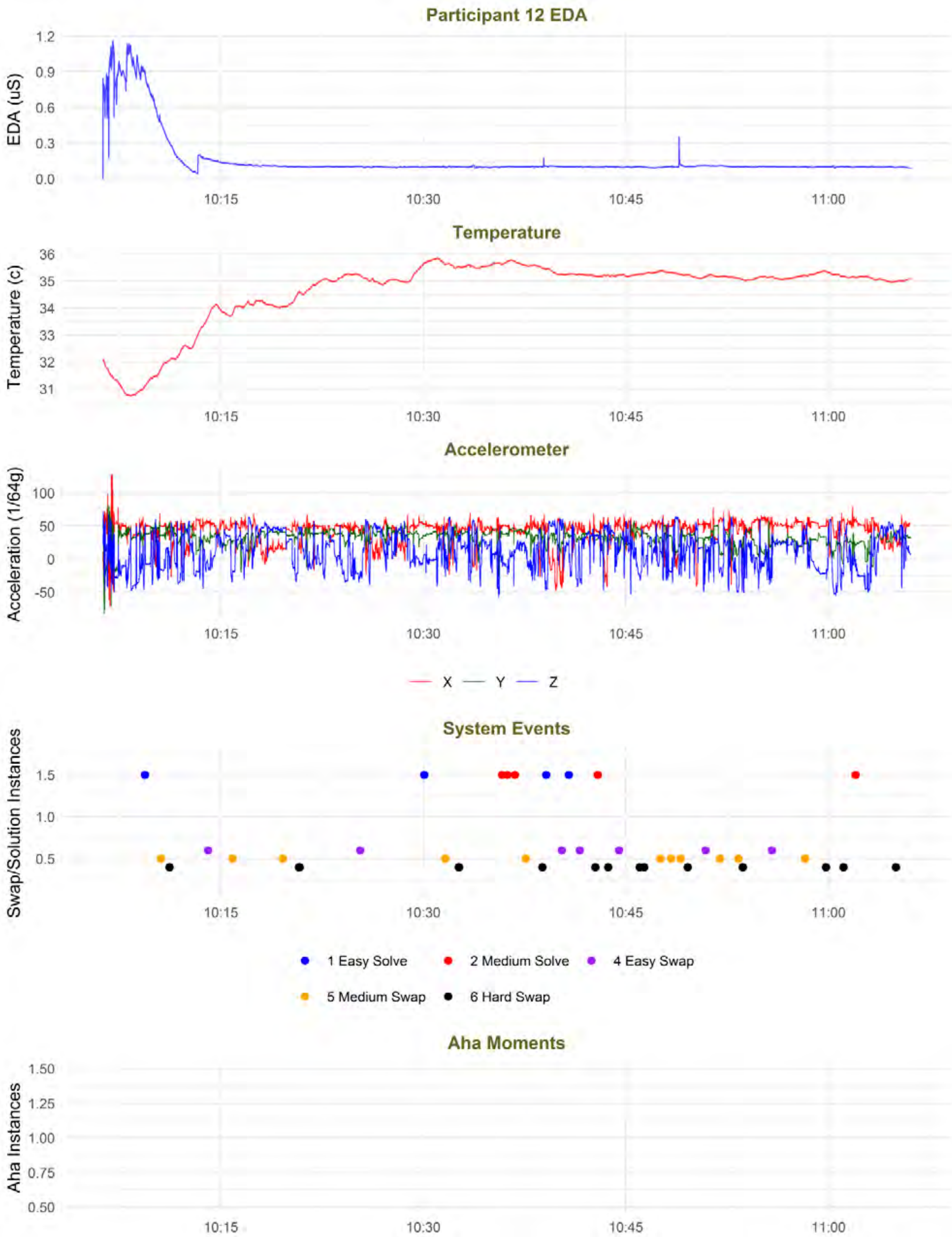

Participant 9 EDA
EDA (uS)
6
4
2
0
10:45
11:00
11:15
11:30
Temperature
Temperature (c)
32
28
24
10:45
11:00
11:15
11:30
Accelerometer
Acceleration (1/64g)
40
0
-40
-80
-120
10:45
11:00
11:15
11:30
X
Y
Z
System Events
Swap/Solution Instances
1.5
1.0
0.5
10:45
11:00
11:15
11:30
1 Easy Solve
4 Easy Swap
5 Medium Swap
6 Hard Swap
Aha Moments
Aha Instances
1.50
1.25
1.00
0.75
0.50
10:45
11:00
11:15
11:30



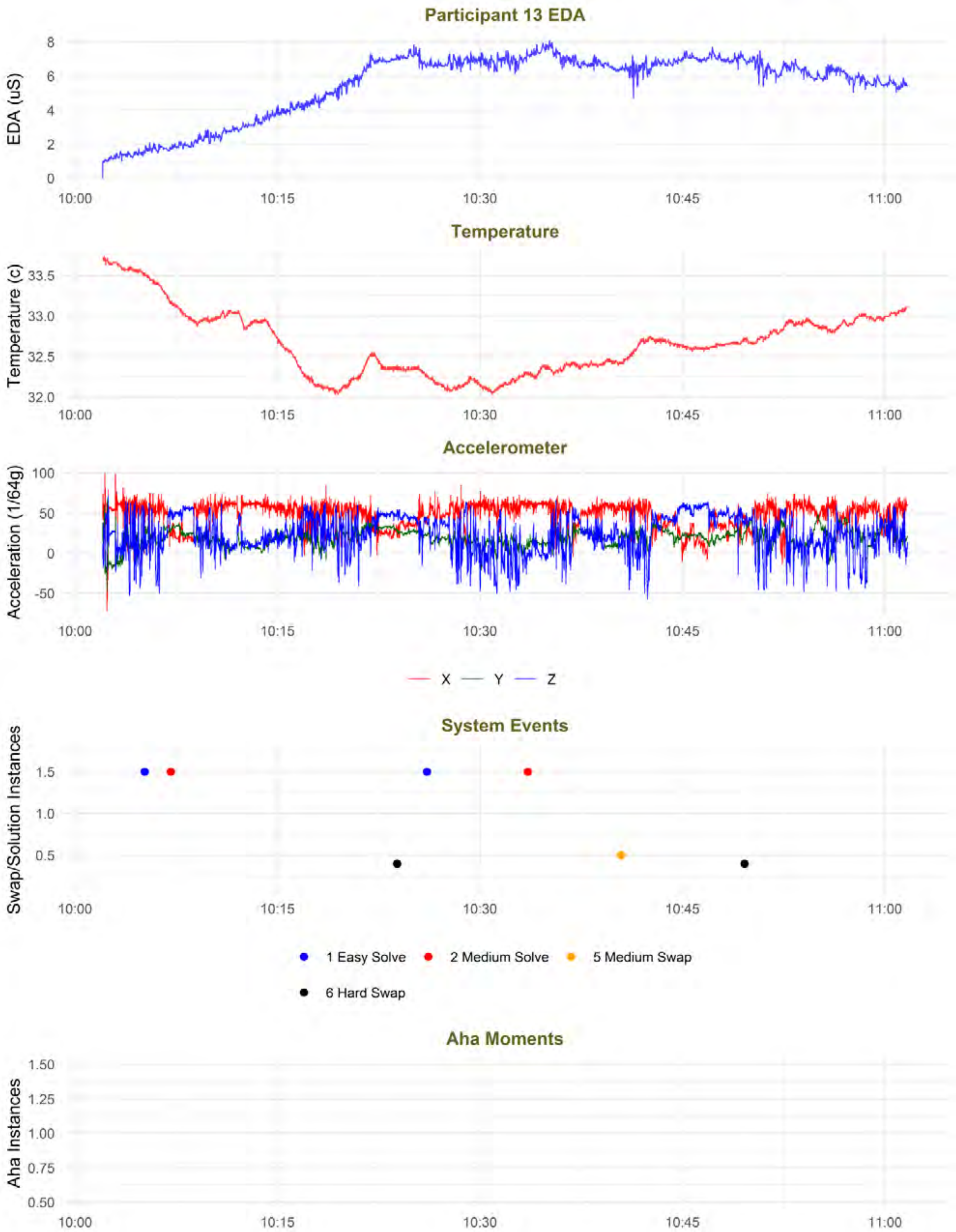
Participant 10 EDA
EDA (uS)
2
1
0
12:45
13:00
13:15
13:30
Temperature
Temperature (c)
37
36
35
34
33
12:45
13:00
13:15
13:30
Accelerometer
Acceleration (1/64g)
50
0
-50
-100
12:45
13:00
13:15
13:30
X
Y
Z
System Events
Swap/Solution Instances
1.5
1.0
0.5
12:45
13:00
13:15
13:30
1 Easy Solve
2 Medium Solve
4 Easy Swap
5 Medium Swap
6 Hard Swap
Aha Moments
Aha Instances
1.50
1.25
1.00
0.75
0.50
12:45
13:00
13:15
13:30

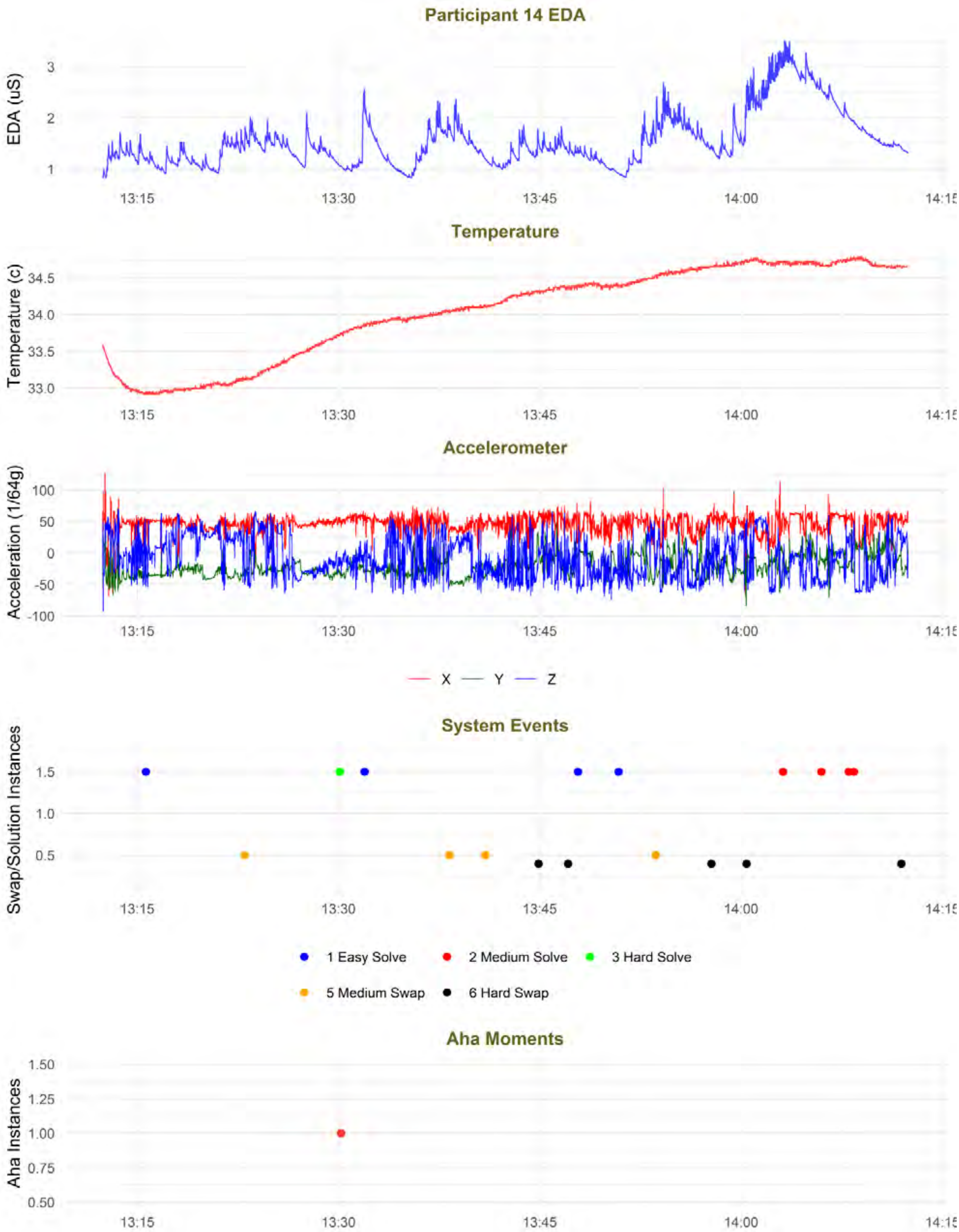


Participant 11 EDA
EDA (uS)
0.3
0.2
0.1
15:30
15:45
16:00
16:15
16:30
Temperature
Temperature (c)
37.0
36.5
36.0
35.5
Accelerometer
Acceleration (1/64g)
50
0
-50
-100
X
Y
Z
System Events
Swap/Solution Instances
1.5
1.0
0.5
1 Easy Solve
2 Medium Solve
3 Hard Solve
4 Easy Swap
5 Medium Swap
6 Hard Swap
Aha Moments
Aha Instances
1.50
1.25
1.00
0.75
0.50

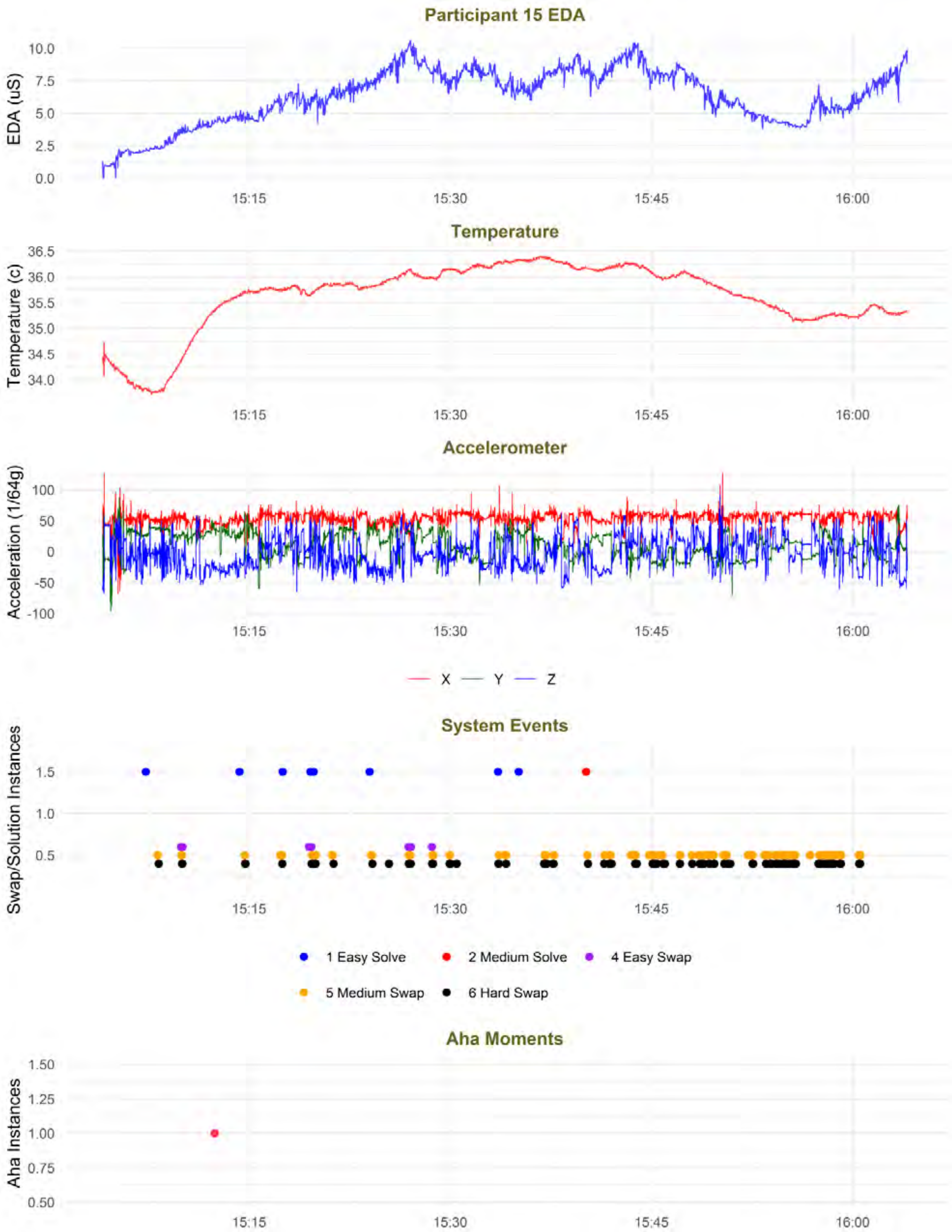


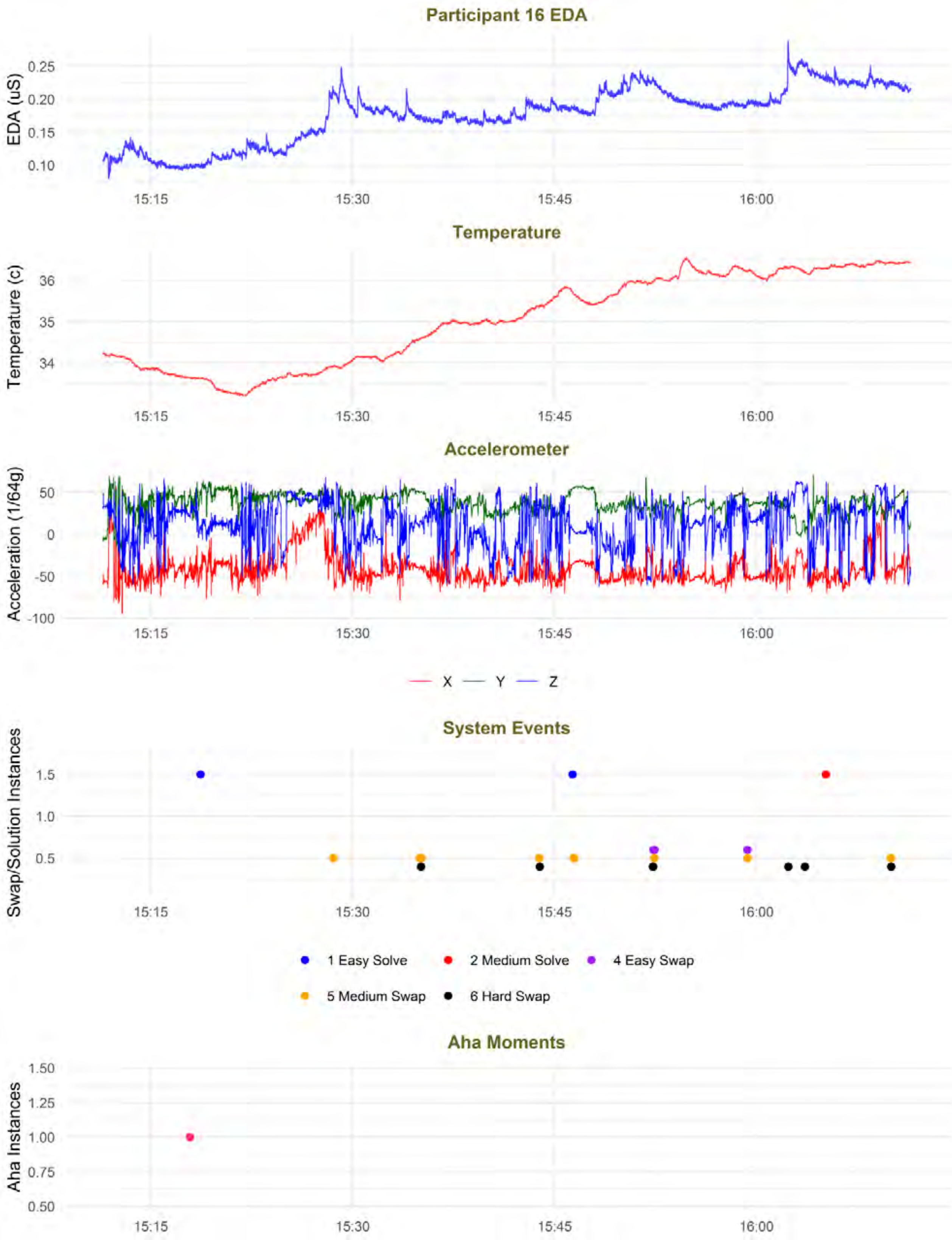
Participant 12 EDA
EDA (uS)
1.2
0.9
0.6
0.3
0.0
10:15
10:30
10:45
11:00
Temperature
Temperature (c)
36
35
34
33
32
31
Accelerometer
Acceleration (1/64g)
100
50
0
-50
X
Y
Z
System Events
Swap/Solution Instances
1.5
1.0
0.5
1 Easy Solve
2 Medium Solve
4 Easy Swap
5 Medium Swap
6 Hard Swap
Aha Moments
Aha Instances
1.50
1.25
1.00
0.75
0.50

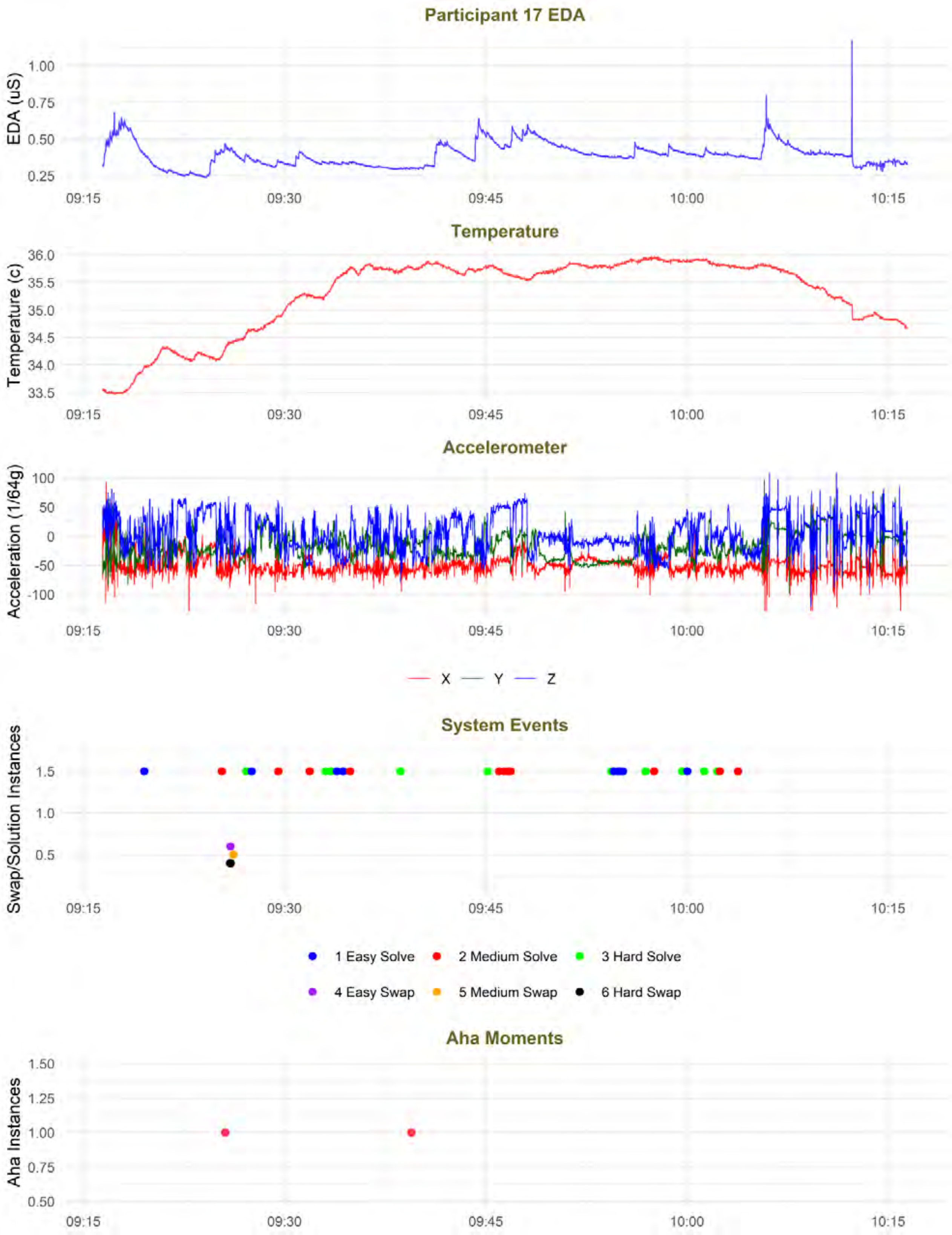

Participant 13 EDA
EDA (uS)
Temperature
Temperature (c)
Accelerometer
Acceleration (1/64g)
X
Y
Z
System Events
Swap/Solution Instances
1 Easy Solve
2 Medium Solve
5 Medium Swap
6 Hard Swap
Aha Moments
Aha Instances
10:00
10:15
10:30
10:45
11:00

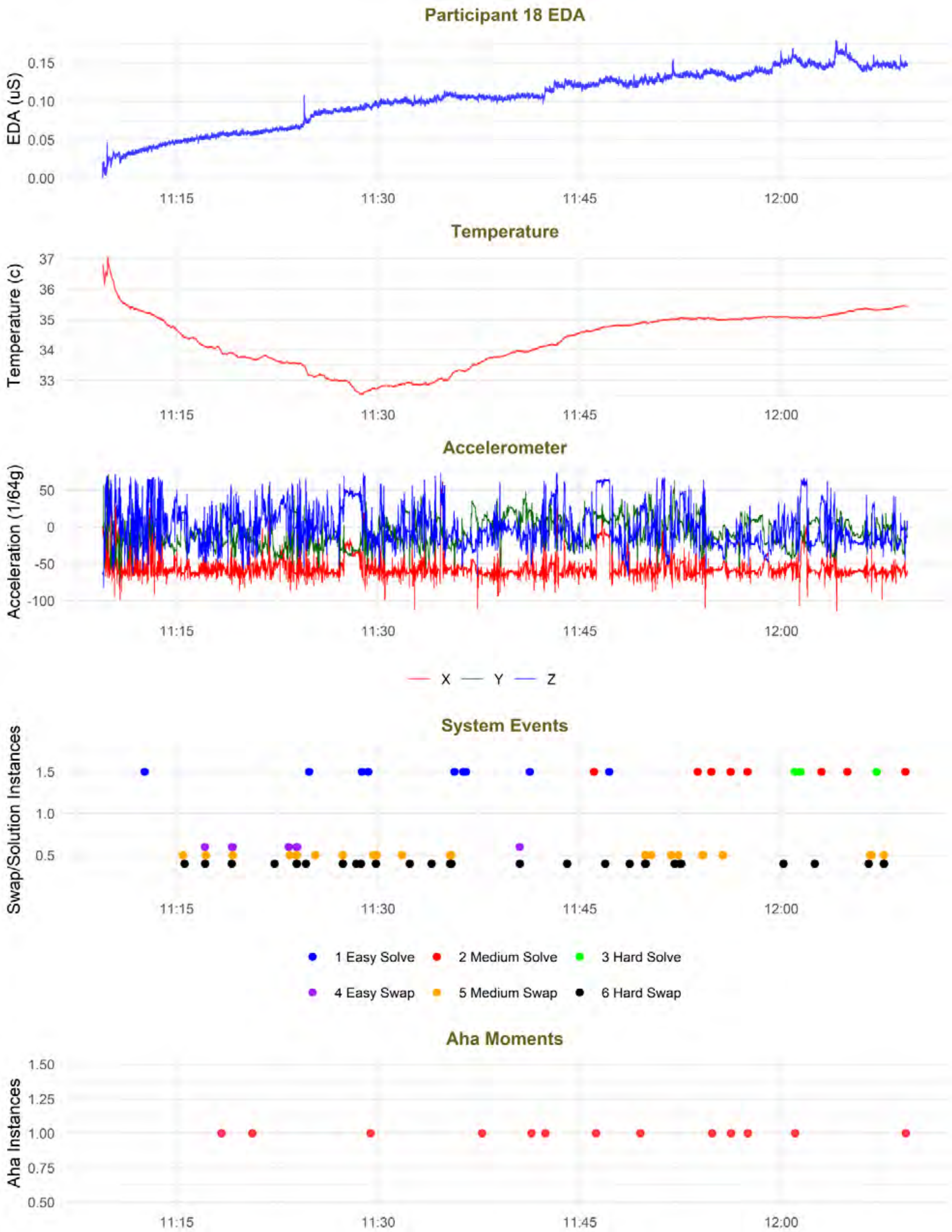


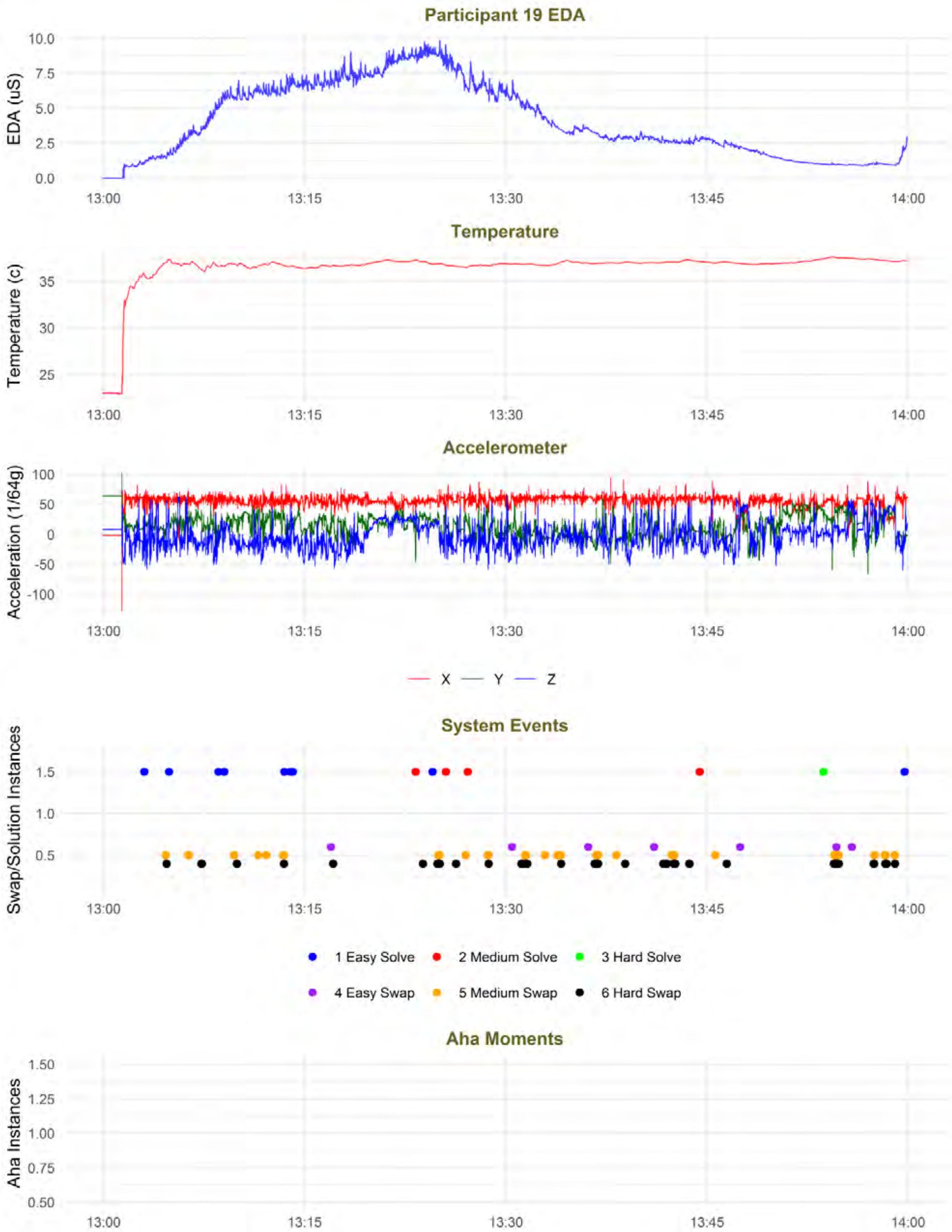
Participant 14 EDA
EDA (uS)
3
2
1
13:15
13:30
13:45
14:00
14:15
Temperature
Temperature (c)
34.5
34.0
33.5
33.0
Accelerometer
Acceleration (1/64g)
100
50
0
-50
-100
X
Y
Z
System Events
Swap/Solution Instances
1.5
1.0
0.5
1 Easy Solve
2 Medium Solve
3 Hard Solve
5 Medium Swap
6 Hard Swap
Aha Moments
Aha Instances
1.50
1.25
1.00
0.75
0.50


Participant 15 EDA
EDA (uS)
10.0
7.5
5.0
2.5
0.0
15:15
15:30
15:45
16:00
Temperature
Temperature (c)
36.5
36.0
35.5
35.0
34.5
34.0
Accelerometer
Acceleration (1/64g)
100
50
0
-50
-100
X
Y
Z
System Events
Swap/Solution Instances
1.5
1.0
0.5
1 Easy Solve
2 Medium Solve
4 Easy Swap
5 Medium Swap
6 Hard Swap
Aha Moments
Aha Instances
1.50
1.25
1.00
0.75
0.50

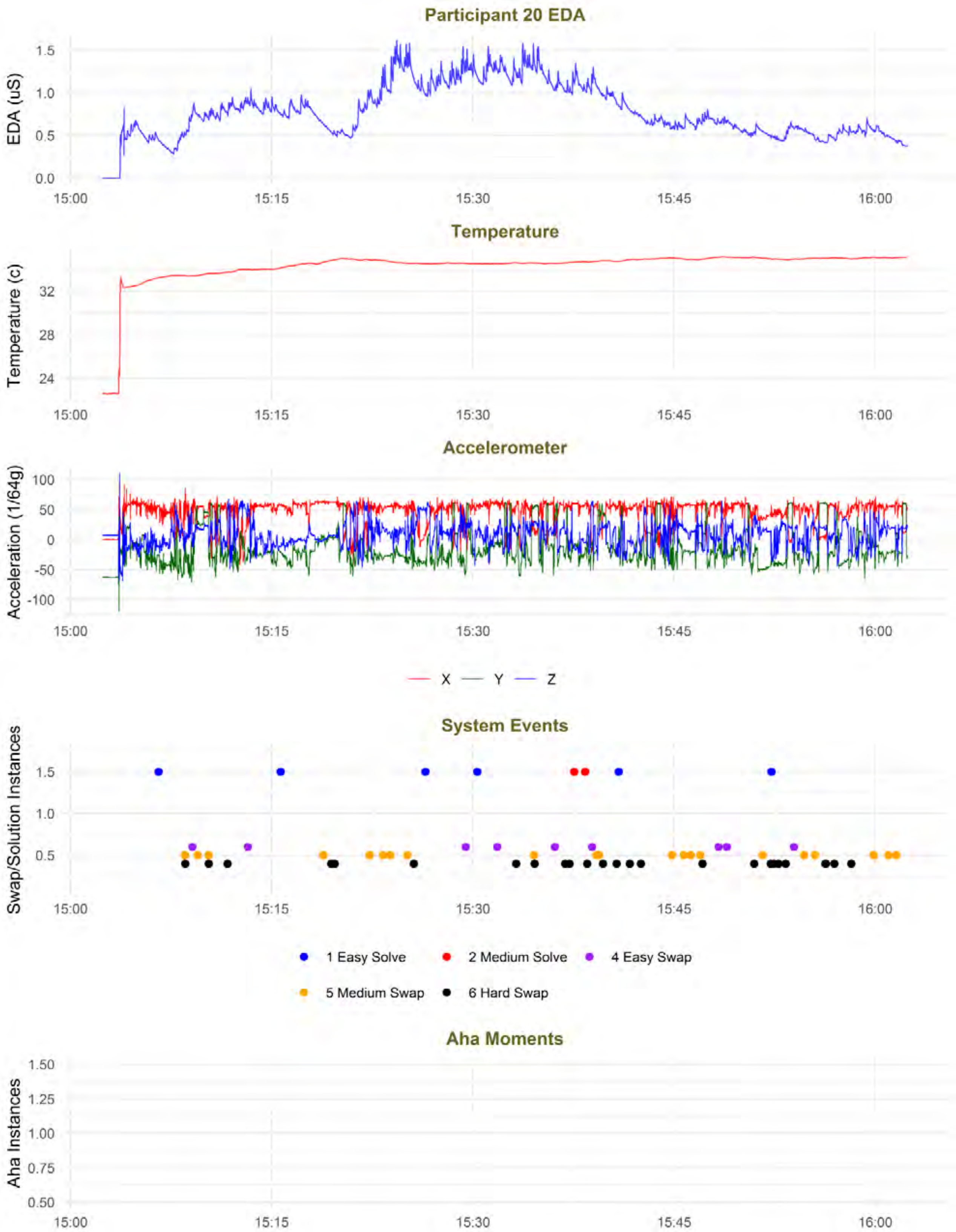

Participant 16 EDA
EDA (uS)
0.25
0.20
0.15
0.10
15:15
15:30
15:45
16:00
Temperature
Temperature (c)
36
35
34
15:15
15:30
15:45
16:00
Accelerometer
Acceleration (1/64g)
50
0
-50
-100
15:15
15:30
15:45
16:00
X
Y
Z
System Events
Swap/Solution Instances
1.5
1.0
0.5
15:15
15:30
15:45
16:00
1 Easy Solve
2 Medium Solve
4 Easy Swap
5 Medium Swap
6 Hard Swap
Aha Moments
Aha Instances
1.50
1.25
1.00
0.75
0.50
15:15
15:30
15:45
16:00

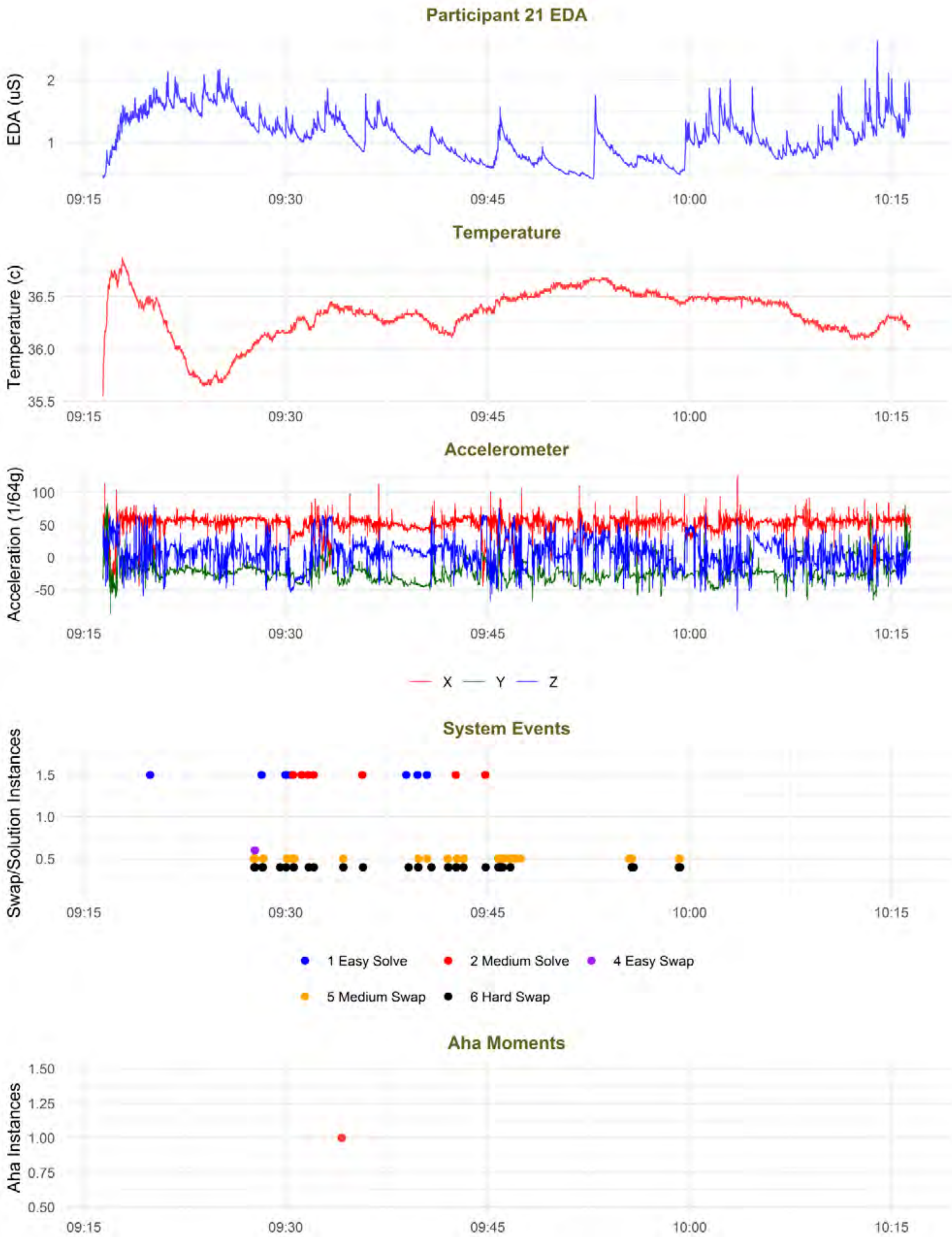

Participant 17 EDA
EDA (uS)
1.00
0.75
0.50
0.25
09:15
09:30
09:45
10:00
10:15
Temperature
Temperature (c)
36.0
35.5
35.0
34.5
34.0
33.5
09:15
09:30
09:45
10:00
10:15
Accelerometer
Acceleration (1/64g)
100
50
0
-50
-100
09:15
09:30
09:45
10:00
10:15
X
Y
Z
System Events
Swap/Solution Instances
1.5
1.0
0.5
09:15
09:30
09:45
10:00
10:15
1 Easy Solve
2 Medium Solve
3 Hard Solve
4 Easy Swap
5 Medium Swap
6 Hard Swap
Aha Moments
Aha Instances
1.50
1.25
1.00
0.75
0.50
09:15
09:30
09:45
10:00
10:15


Participant 18 EDA
EDA (uS)
0.15
0.10
0.05
0.00
11:15
11:30
11:45
12:00
Temperature
Temperature (c)
37
36
35
34
33
11:15
11:30
11:45
12:00
Accelerometer
Acceleration (1/64g)
50
0
-50
-100
11:15
11:30
11:45
12:00
X
Y
Z
System Events
Swap/Solution Instances
1.5
1.0
0.5
11:15
11:30
11:45
12:00
1 Easy Solve
2 Medium Solve
3 Hard Solve
4 Easy Swap
5 Medium Swap
6 Hard Swap
Aha Moments
Aha Instances
1.50
1.25
1.00
0.75
0.50
11:15
11:30
11:45
12:00

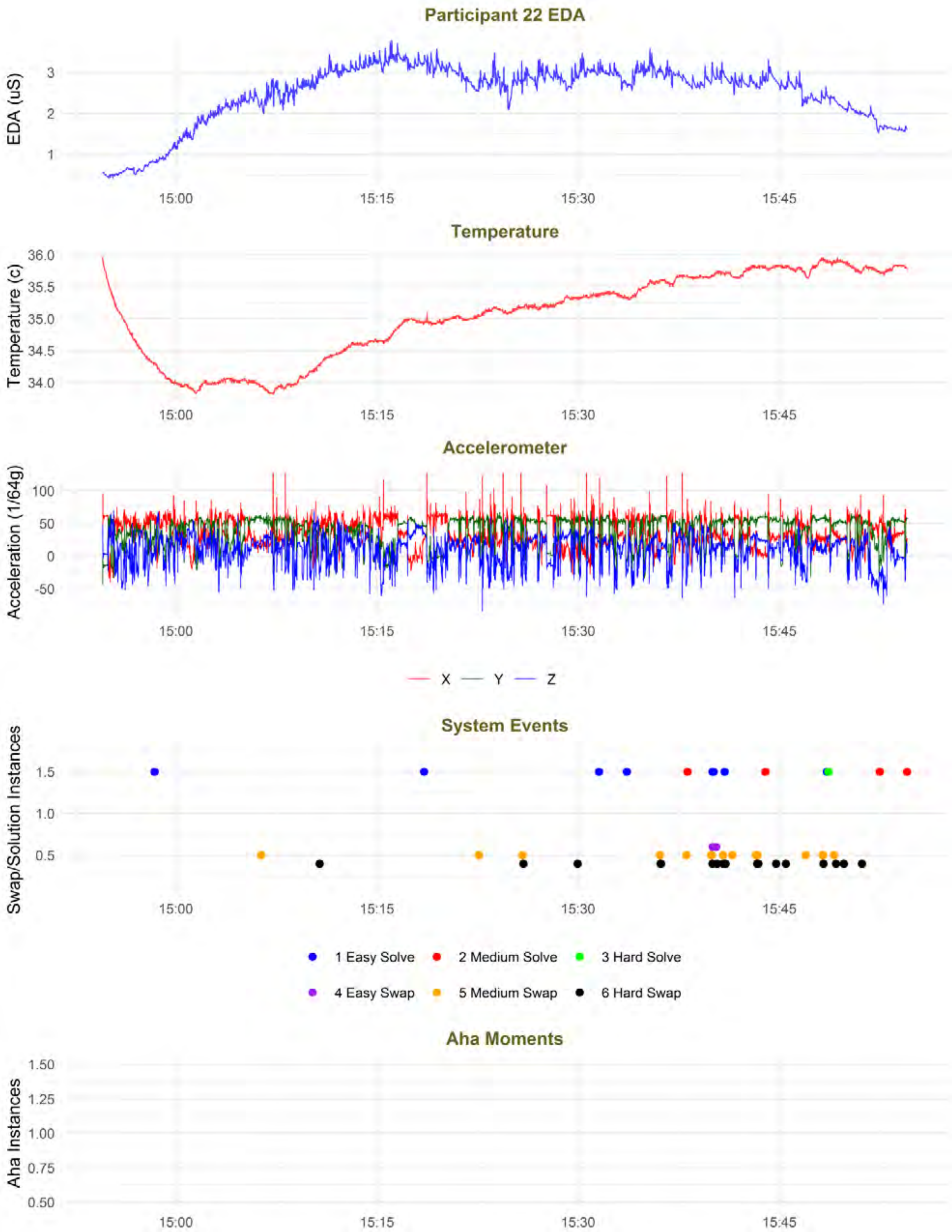


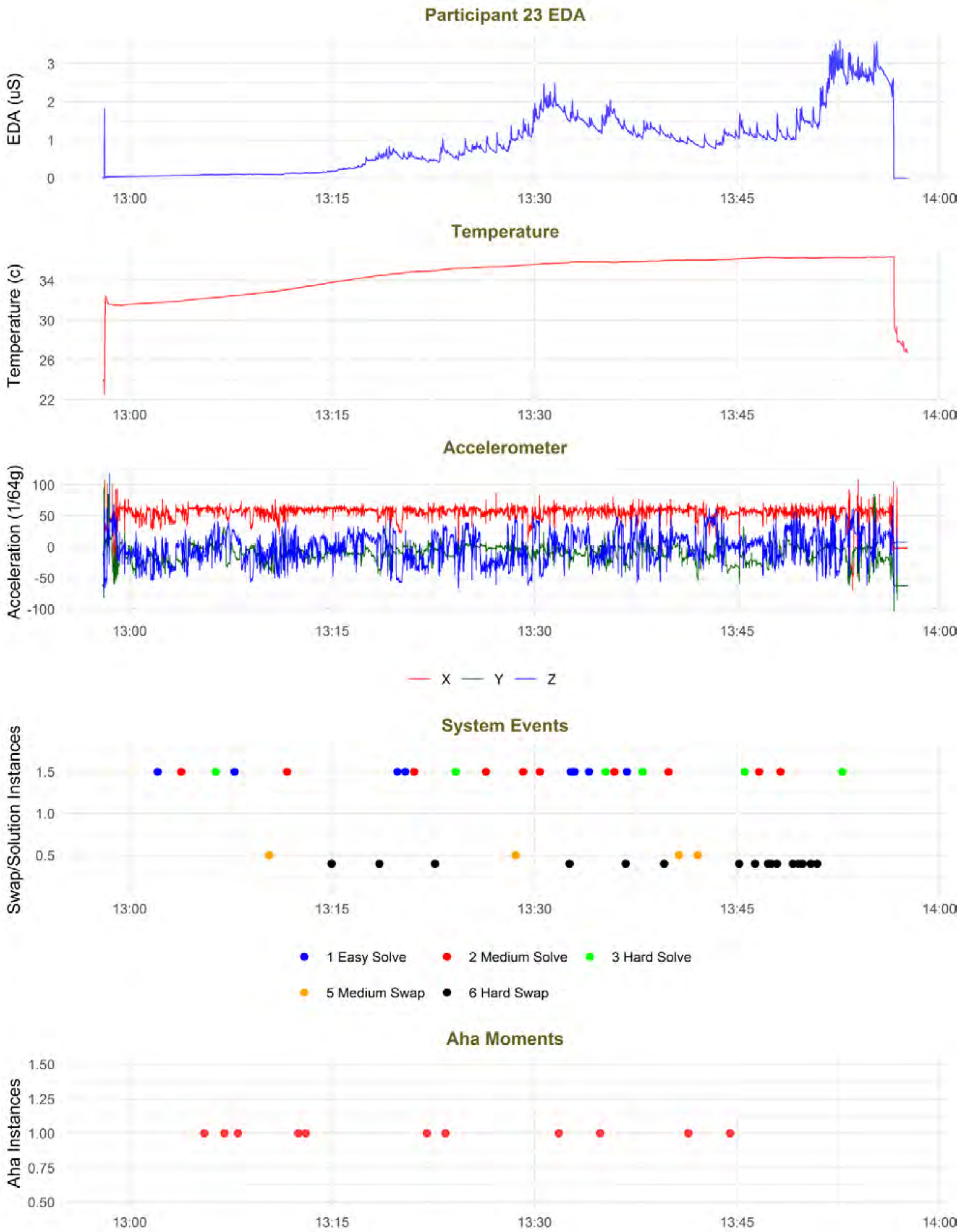
Participant 19 EDA
EDA (uS)
10.0
7.5
5.0
2.5
0.0
13:00
13:15
13:30
13:45
14:00
Temperature
Temperature (c)
35
30
25
Accelerometer
Acceleration (1/64g)
100
50
0
-50
-100
X
Y
Z
System Events
Swap/Solution Instances
1.5
1.0
0.5
1 Easy Solve
2 Medium Solve
3 Hard Solve
4 Easy Swap
5 Medium Swap
6 Hard Swap
Aha Moments
Aha Instances
1.50
1.25
1.00
0.75
0.50

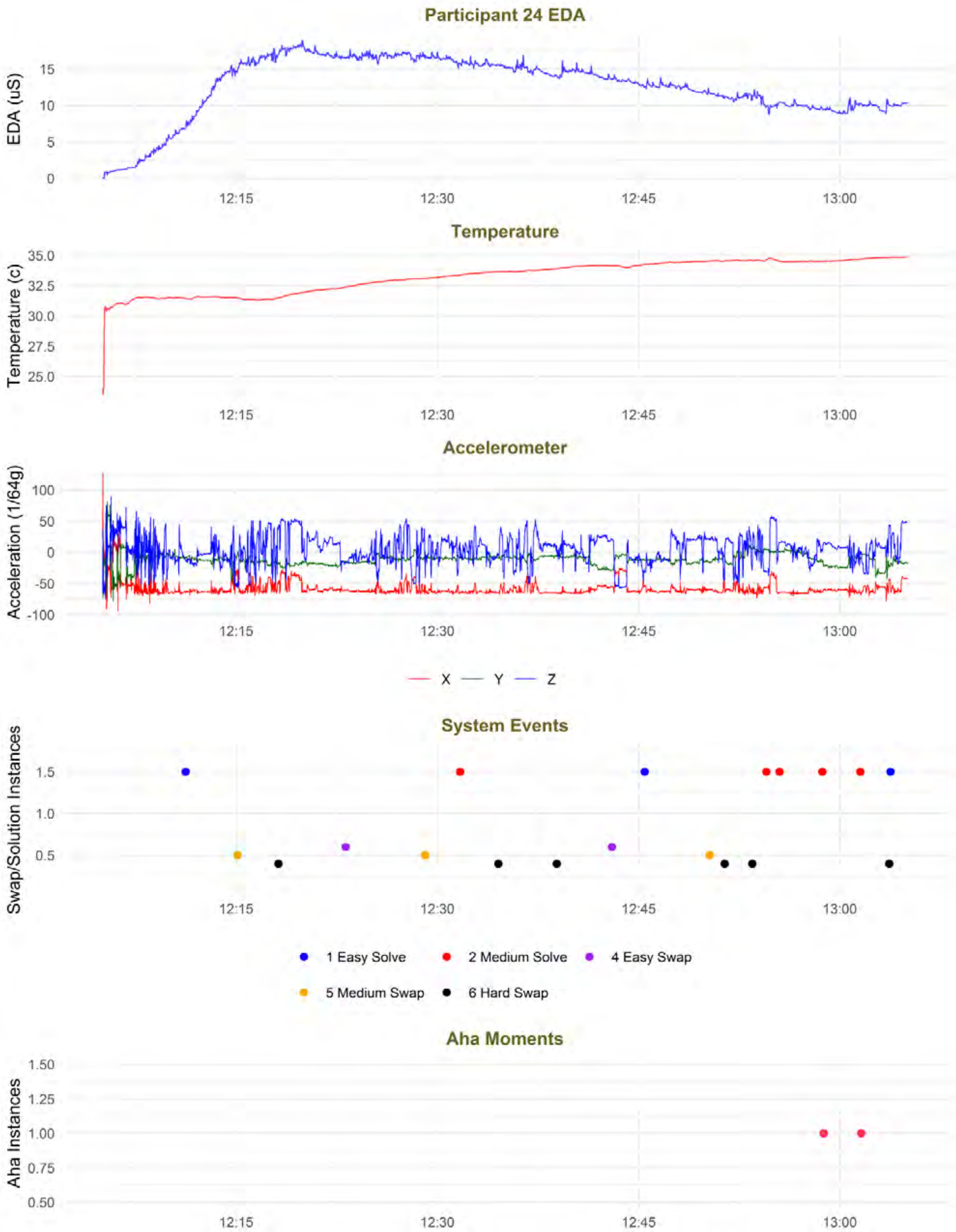

Participant 20 EDA
EDA (uS)
1.5
1.0
0.5
0.0
15:00
15:15
15:30
15:45
16:00
Temperature
Temperature (c)
32
28
24
15:00
15:15
15:30
15:45
16:00
Accelerometer
Acceleration (1/64g)
100
50
0
-50
-100
15:00
15:15
15:30
15:45
16:00
X
Y
Z
System Events
Swap/Solution Instances
1.5
1.0
0.5
15:00
15:15
15:30
15:45
16:00
1 Easy Solve
2 Medium Solve
4 Easy Swap
5 Medium Swap
6 Hard Swap
Aha Moments
Aha Instances
1.50
1.25
1.00
0.75
0.50
15:00
15:15
15:30
15:45
16:00



Participant 21 EDA
EDA (uS)
2
1
09:15
09:30
09:45
10:00
10:15
Temperature
Temperature (c)
36.5
36.0
35.5
Accelerometer
Acceleration (1/64g)
100
50
0
-50
X
Y
Z
System Events
Swap/Solution Instances
1.5
1.0
0.5
1 Easy Solve
2 Medium Solve
4 Easy Swap
5 Medium Swap
6 Hard Swap
Aha Moments
Aha Instances
1.50
1.25
1.00
0.75
0.50


Participant 22 EDA
EDA (uS)
3
2
1
15:00
15:15
15:30
15:45
Temperature
Temperature (c)
36.0
35.5
35.0
34.5
34.0
15:00
15:15
15:30
15:45
Accelerometer
Acceleration (1/64g)
100
50
0
-50
15:00
15:15
15:30
15:45
X
Y
Z
System Events
Swap/Solution Instances
1.5
1.0
0.5
15:00
15:15
15:30
15:45
1 Easy Solve
2 Medium Solve
3 Hard Solve
4 Easy Swap
5 Medium Swap
6 Hard Swap
Aha Moments
Aha Instances
1.50
1.25
1.00
0.75
0.50
15:00
15:15
15:30
15:45


Participant 23 EDA
EDA (uS)
Temperature
Temperature (c)
Accelerometer
Acceleration (1/64g)
X
Y
Z
System Events
Swap/Solution Instances
1 Easy Solve
2 Medium Solve
3 Hard Solve
5 Medium Swap
6 Hard Swap
Aha Moments
Aha Instances
13:00
13:15
13:30
13:45
14:00

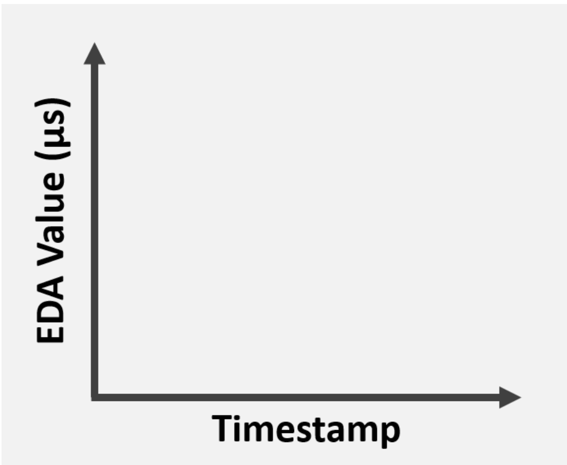

Participant 24 EDA
EDA (uS)
0
5
10
15
12:15
12:30
12:45
13:00
Temperature
Temperature (c)
25.0
27.5
30.0
32.5
35.0
12:15
12:30
12:45
13:00
Accelerometer
Acceleration (1/64g)
-100
-50
0
50
100
12:15
12:30
12:45
13:00
X
Y
Z
System Events
Swap/Solution Instances
0.5
1.0
1.5
12:15
12:30
12:45
13:00
1 Easy Solve
2 Medium Solve
4 Easy Swap
5 Medium Swap
6 Hard Swap
Aha Moments
Aha Instances
0.50
0.75
1.00
1.25
1.50
12:15
12:30
12:45
13:00

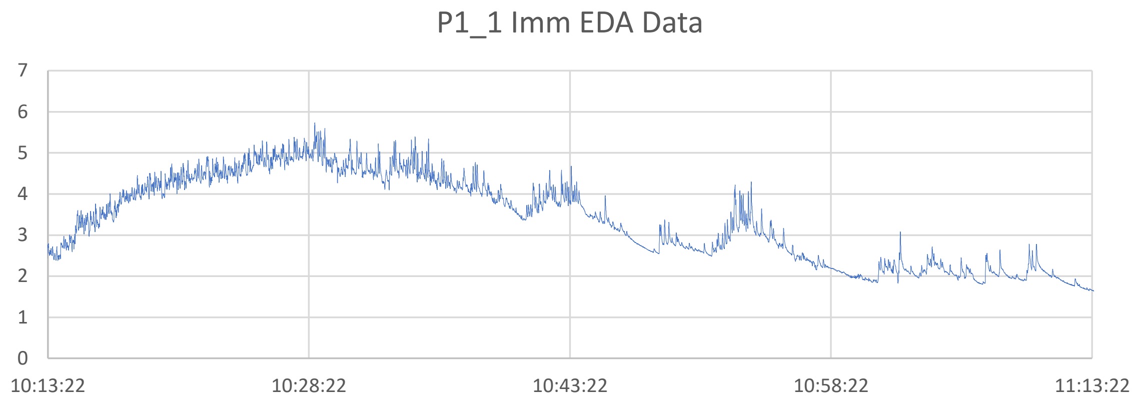
Appendix D

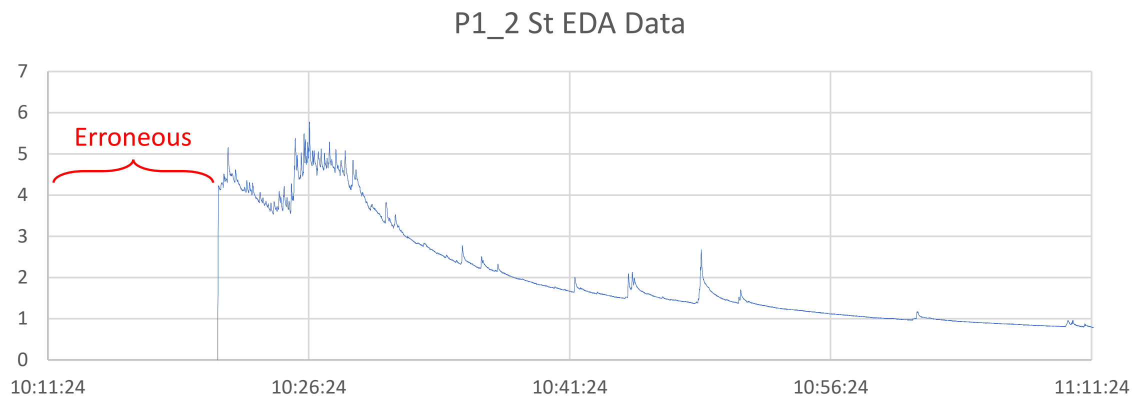
Electrodermal Activity Data - Study Four

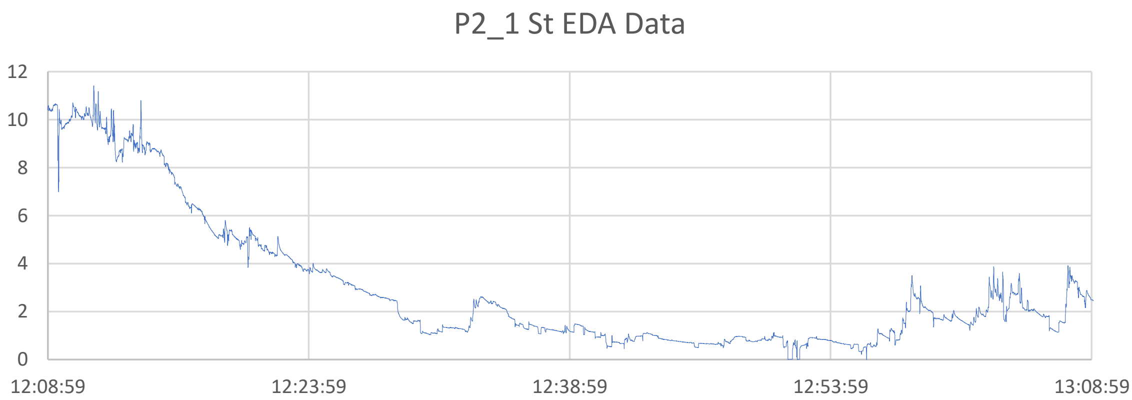
This appendix contains the plots for all electrodermal activity (EDA) data collected in our fourth experiment. Each participant has two plots, one for each condition. Each plot has a title above it of the form "Px_y Imm EDA Data" where x is the participant number, and y is the session (i.e. first or second). If the heading says "Imm", it was an immersive condition. If it says "St", it was the stereo condition.

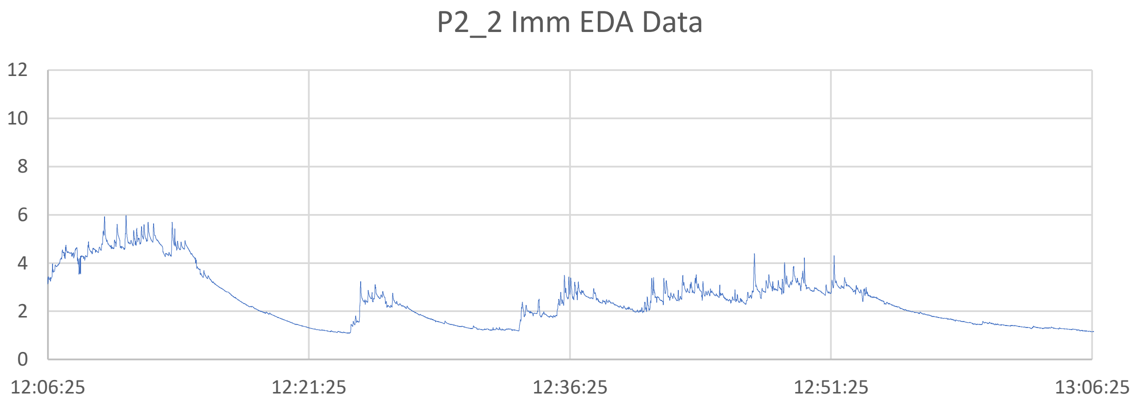
The EDA readings are not always plotted within the same EDA range due to each participants' unique physiology. Even within participant readings can vary. The figure below demonstrates the axes values for all following plots. In all cases the vertical axis is the EDA reading in microsiemens, and the horizontal axis is the timestamp.

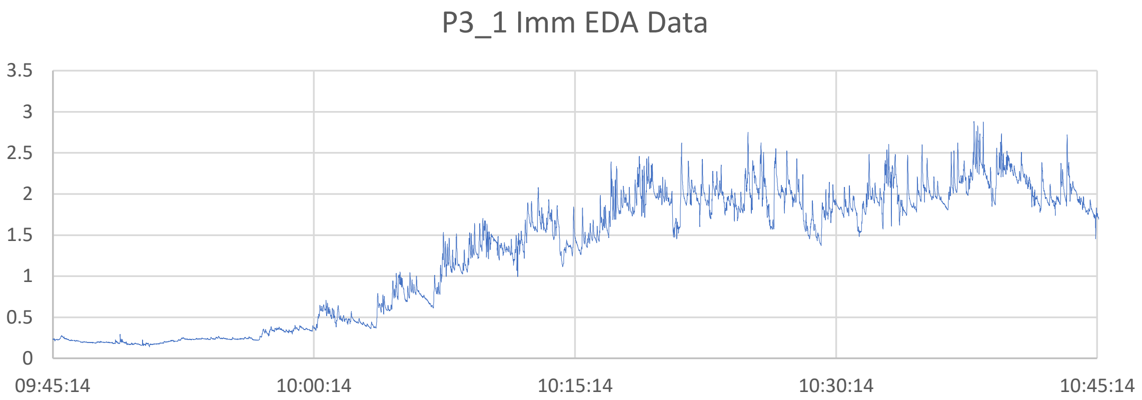


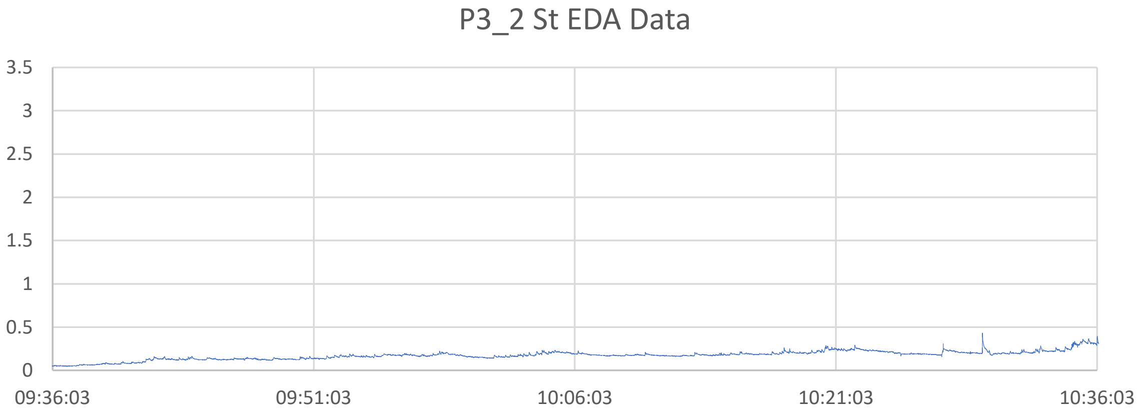

P1_1 Imm EDA Data
7
6
5
4
3
2
1
0
10:13:22
10:28:22
10:43:22
10:58:22
11:13:22

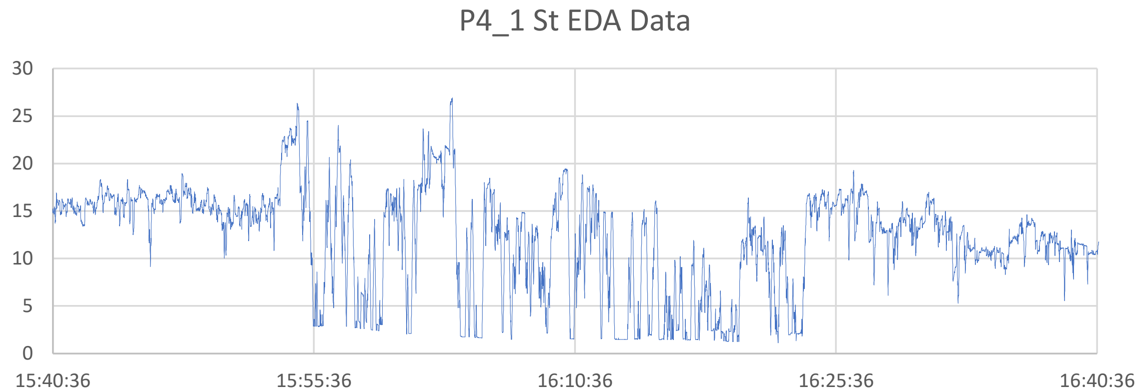

P1_2 St EDA Data
Erroneous
7
6
5
4
3
2
1
0
10:11:24
10:26:24
10:41:24
10:56:24
11:11:24

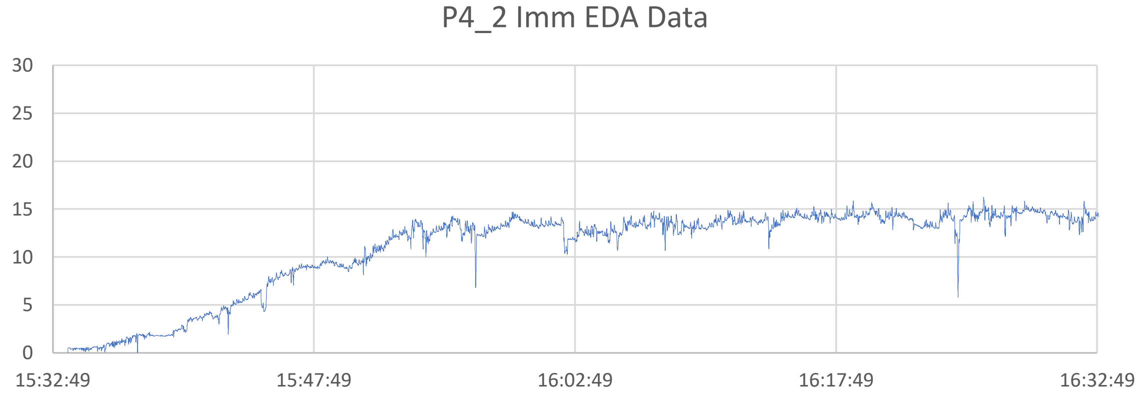

P2_1 St EDA Data
12
10
8
6
4
2
0
12:08:59
12:23:59
12:38:59
12:53:59
13:08:59


P2_2 Imm EDA Data
12
10
8
6
4
2
0
12:06:25
12:21:25
12:36:25
12:51:25
13:06:25

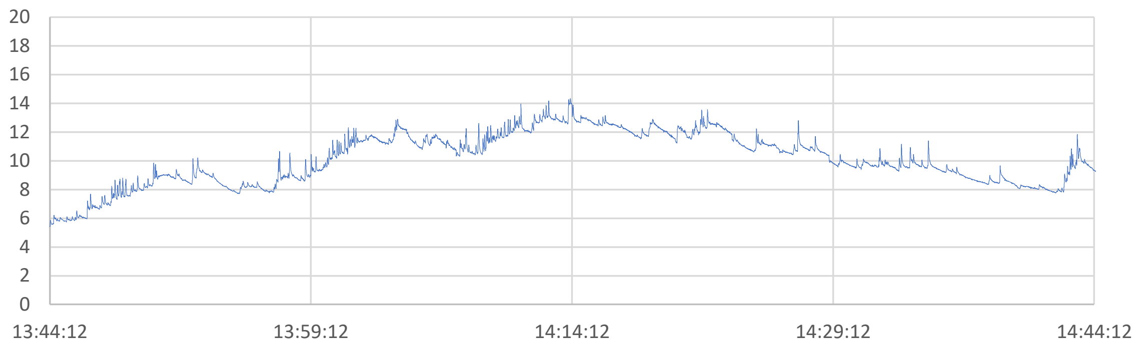

P3_1 Imm EDA Data
3.5
3
2.5
2
1.5
1
0.5
0
09:45:14
10:00:14
10:15:14
10:30:14
10:45:14


P3_2 St EDA Data
3.5
3
2.5
2
1.5
1
0.5
0
09:36:03
09:51:03
10:06:03
10:21:03
10:36:03

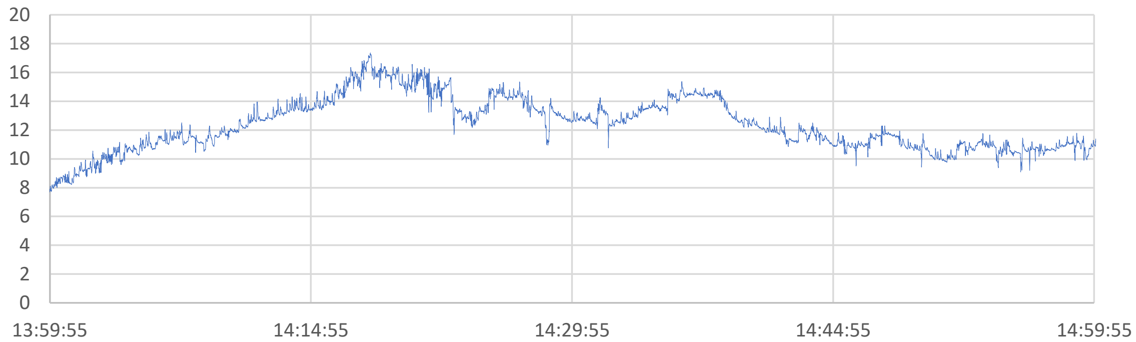

P4_1 St EDA Data
30
25
20
15
10
5
0
15:40:36
15:55:36
16:10:36
16:25:36
16:40:36


P4_2 Imm EDA Data
30
25
20
15
10
5
0
15:32:49
15:47:49
16:02:49
16:17:49
16:32:49

P5_1 St EDA Data

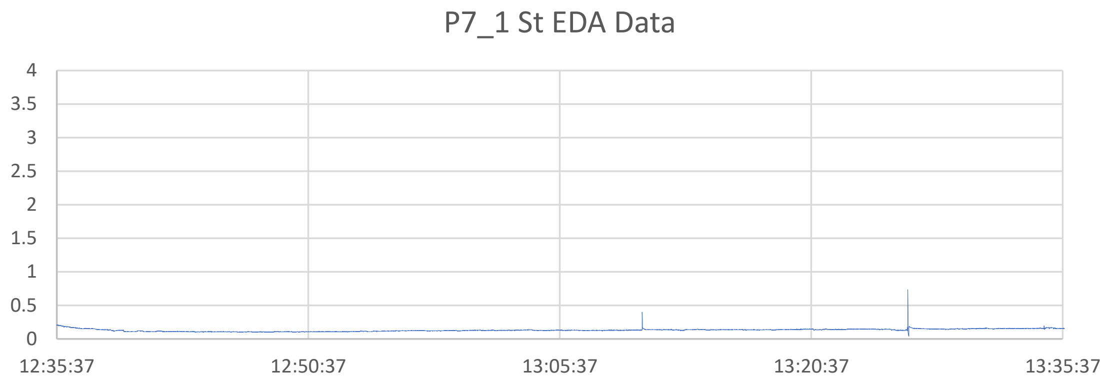


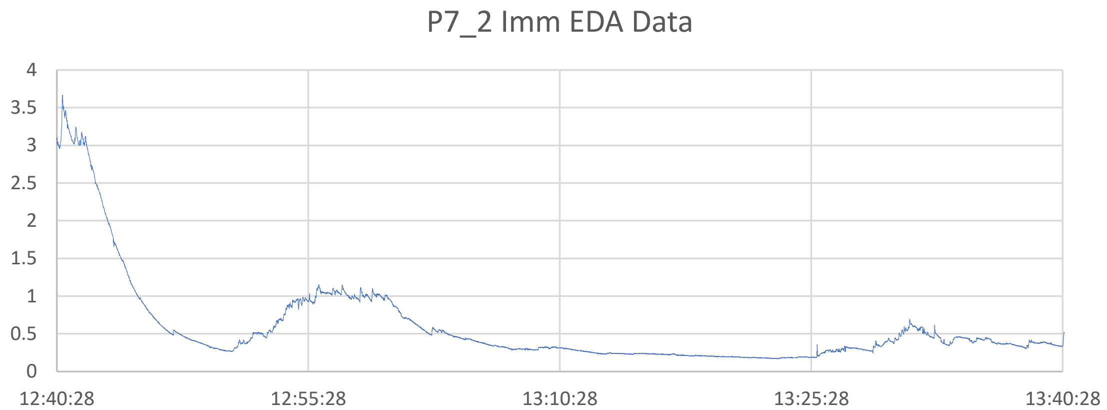
P5_2 Imm EDA Data

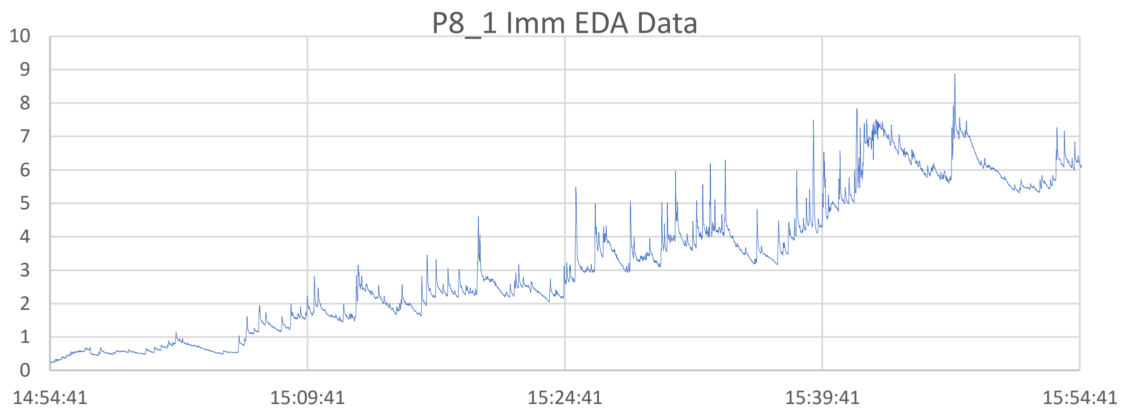


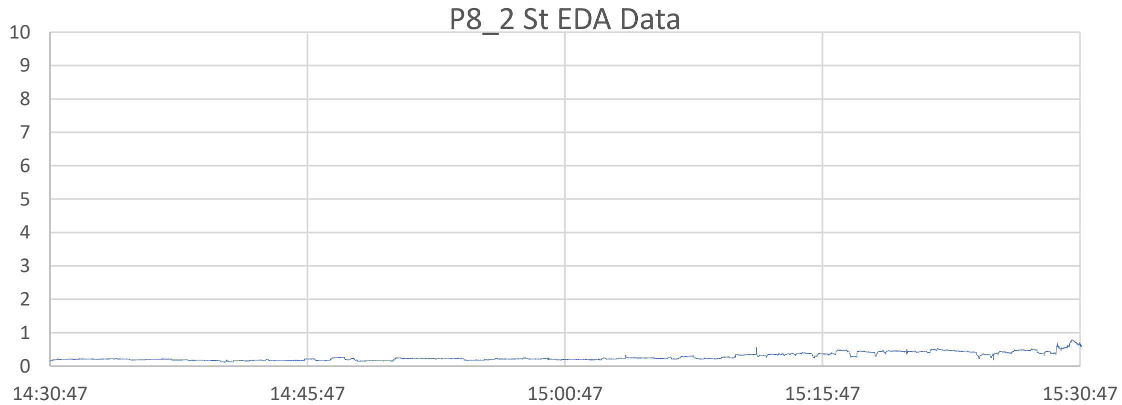
P6_1 Imm EDA Data

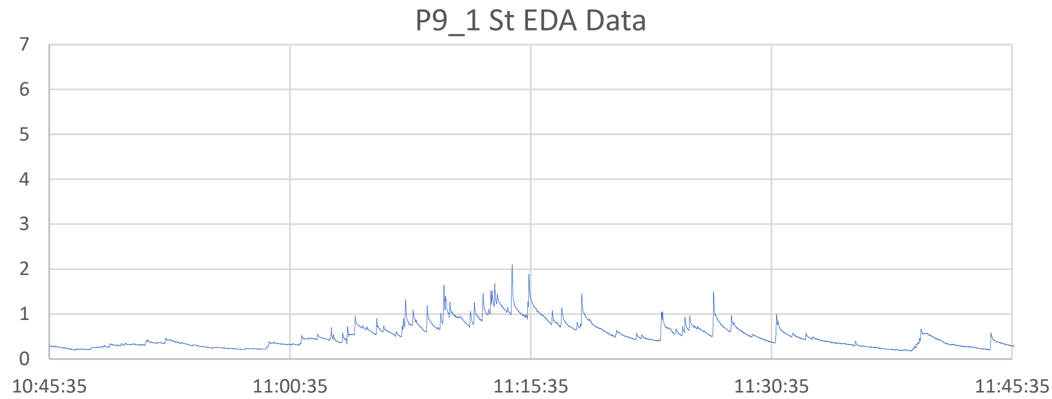
P6_2 St EDA Data

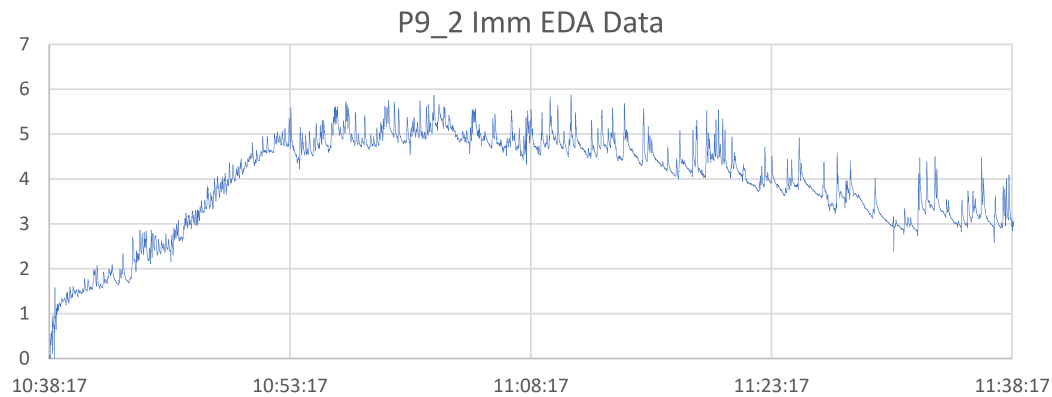

P7_1 St EDA Data
4
3.5
3
2.5
2
1.5
1
0.5
0
12:35:37
12:50:37
13:05:37
13:20:37
13:35:37

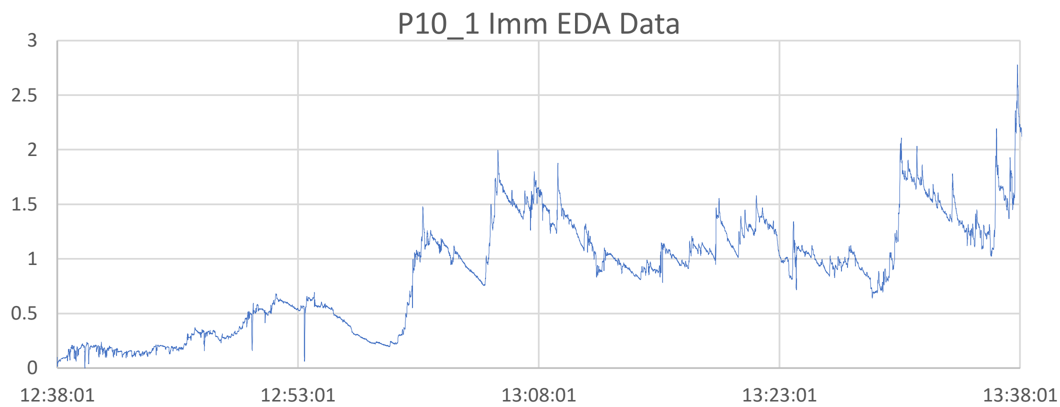

P7_2 Imm EDA Data
4
3.5
3
2.5
2
1.5
1
0.5
0
12:40:28
12:55:28
13:10:28
13:25:28
13:40:28

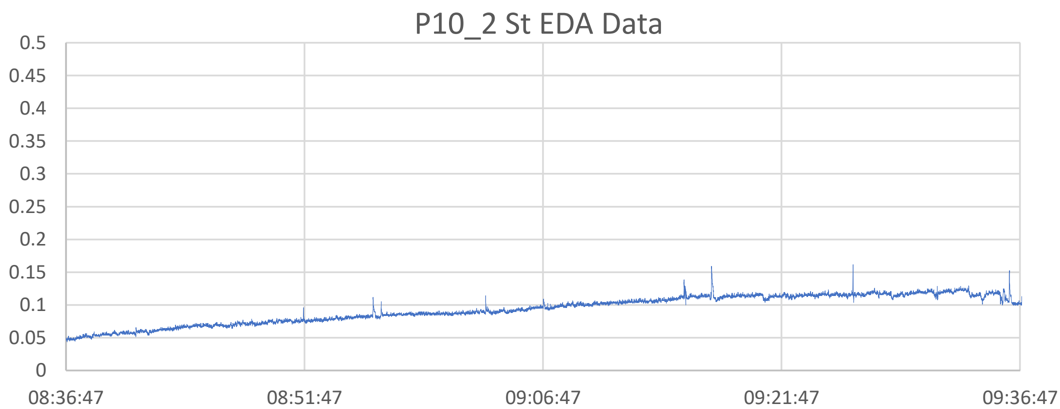

P8_1 Imm EDA Data
10
9
8
7
6
5
4
3
2
1
0
14:54:41
15:09:41
15:24:41
15:39:41
15:54:41

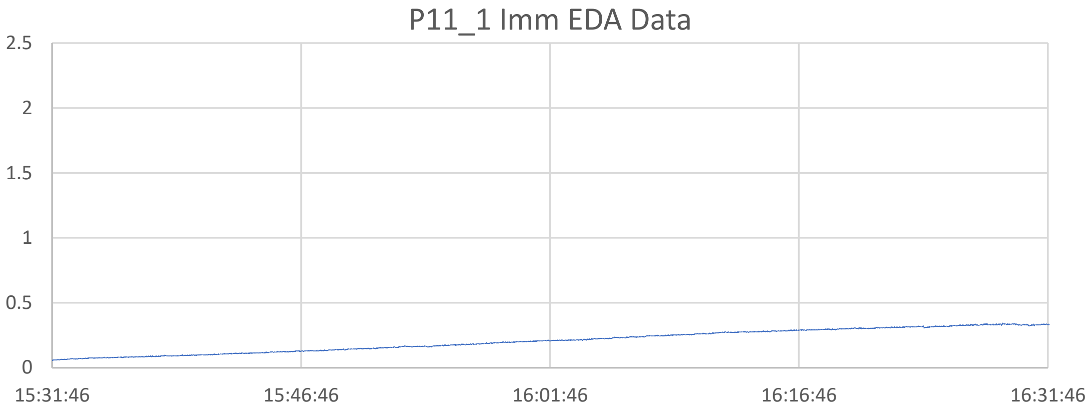

P8_2 St EDA Data
10
9
8
7
6
5
4
3
2
1
0
14:30:47
14:45:47
15:00:47
15:15:47
15:30:47


P9_1 St EDA Data
7
6
5
4
3
2
1
0
10:45:35
11:00:35
11:15:35
11:30:35
11:45:35

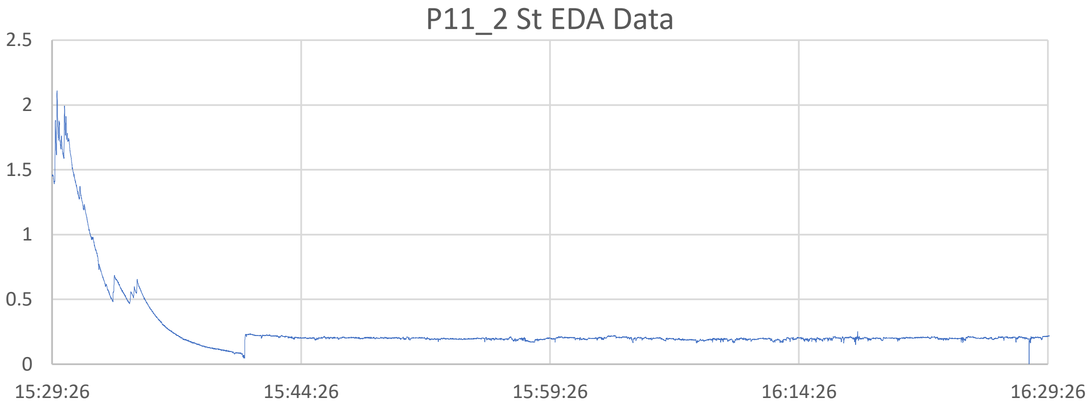

P9_2 Imm EDA Data
7
6
5
4
3
2
1
0
10:38:17
10:53:17
11:08:17
11:23:17
11:38:17


P10_1 Imm EDA Data
3
2.5
2
1.5
1
0.5
0
12:38:01
12:53:01
13:08:01
13:23:01
13:38:01

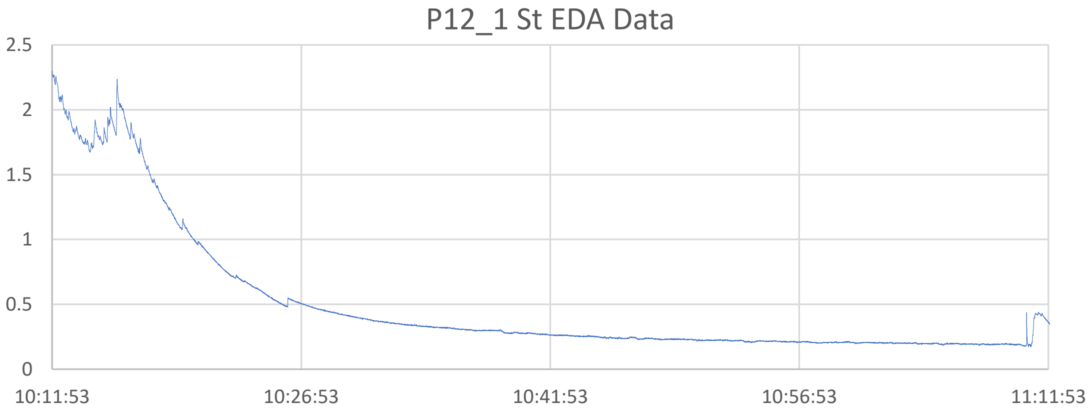

P10_2 St EDA Data
0.5
0.45
0.4
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
08:36:47
08:51:47
09:06:47
09:21:47
09:36:47



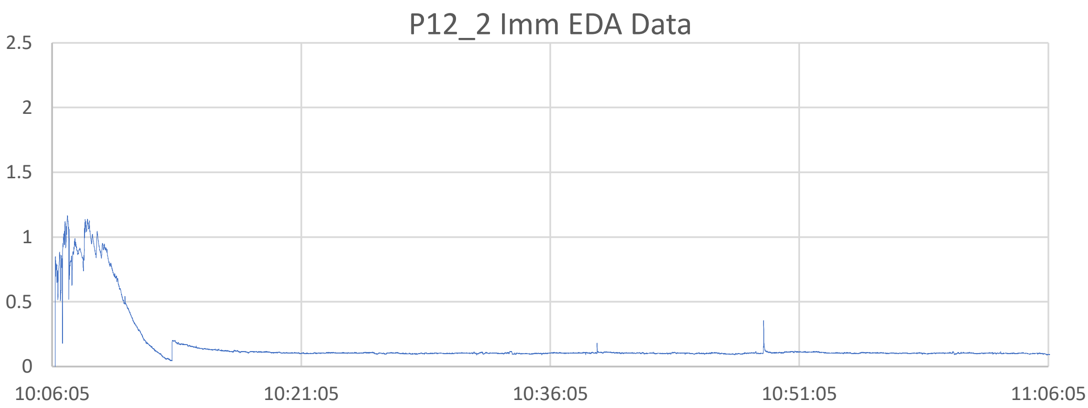
P11_1 Imm EDA Data
2.5
2
1.5
1
0.5
0
15:31:46
15:46:46
16:01:46
16:16:46
16:31:46



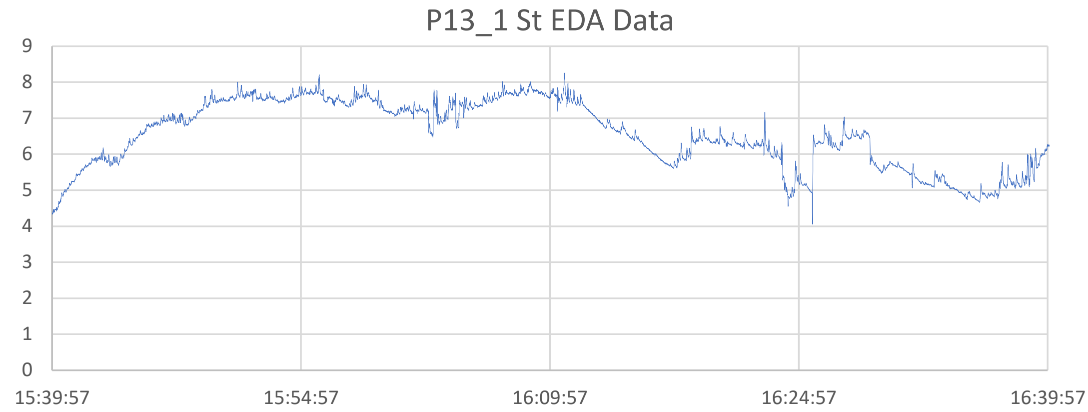
P11_2 St EDA Data
2.5
2
1.5
1
0.5
0
15:29:26
15:44:26
15:59:26
16:14:26
16:29:26

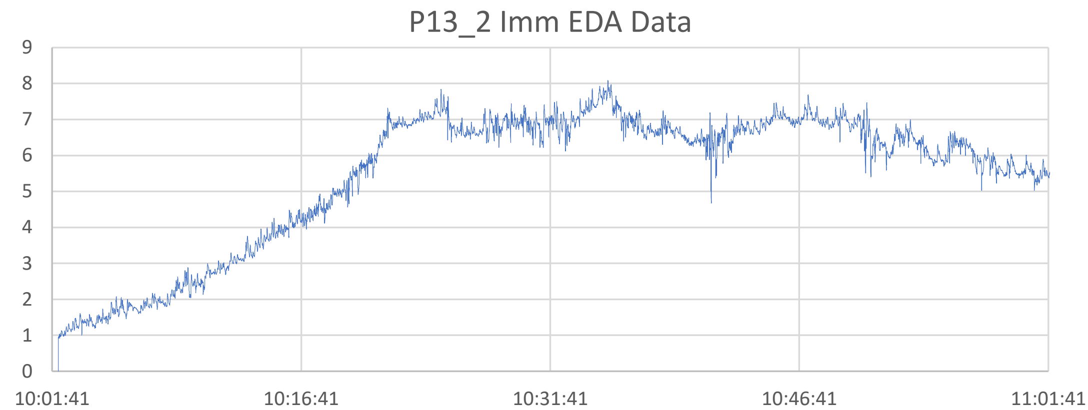


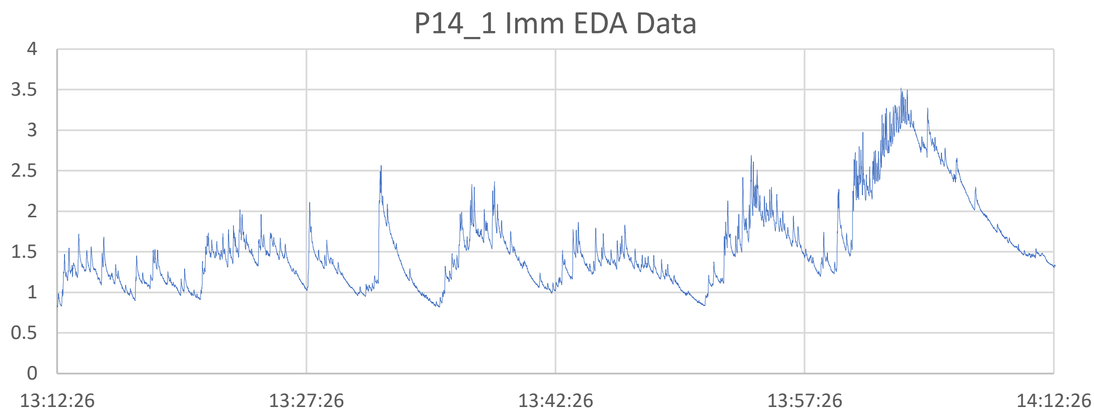
P12_1 St EDA Data
2.5
2
1.5
1
0.5
0
10:11:53
10:26:53
10:41:53
10:56:53
11:11:53

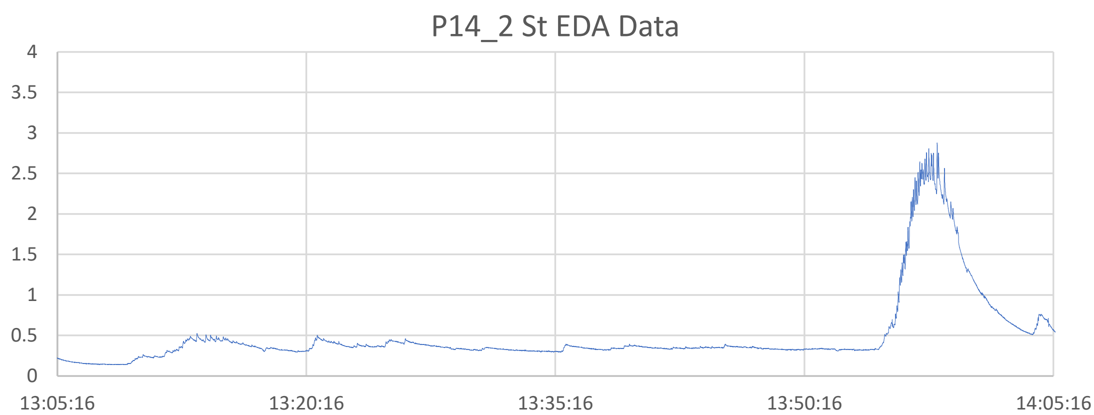


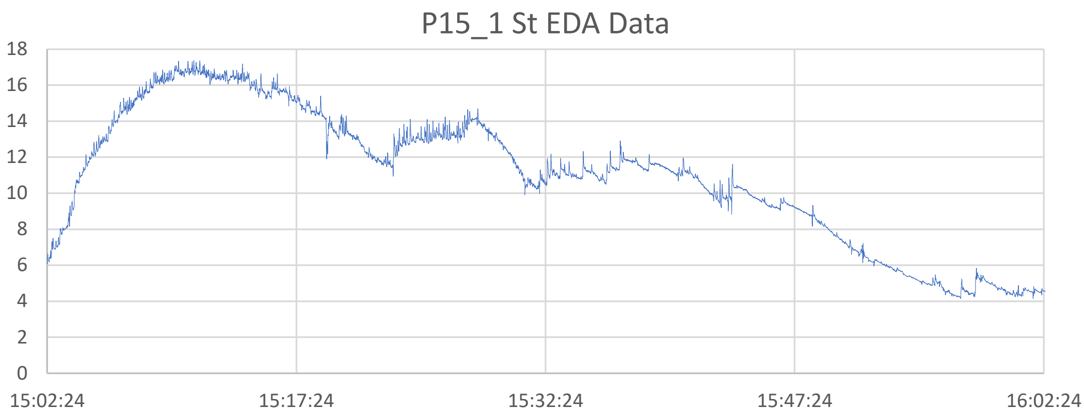
P12_2 Imm EDA Data
2.5
2
1.5
1
0.5
0
10:06:05
10:21:05
10:36:05
10:51:05
11:06:05

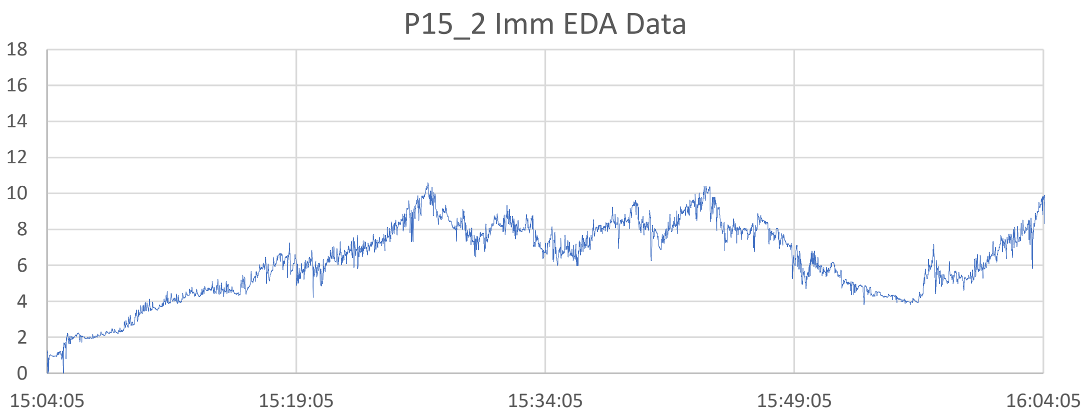

P13_1 St EDA Data
9
8
7
6
5
4
3
2
1
0
15:39:57
15:54:57
16:09:57
16:24:57
16:39:57

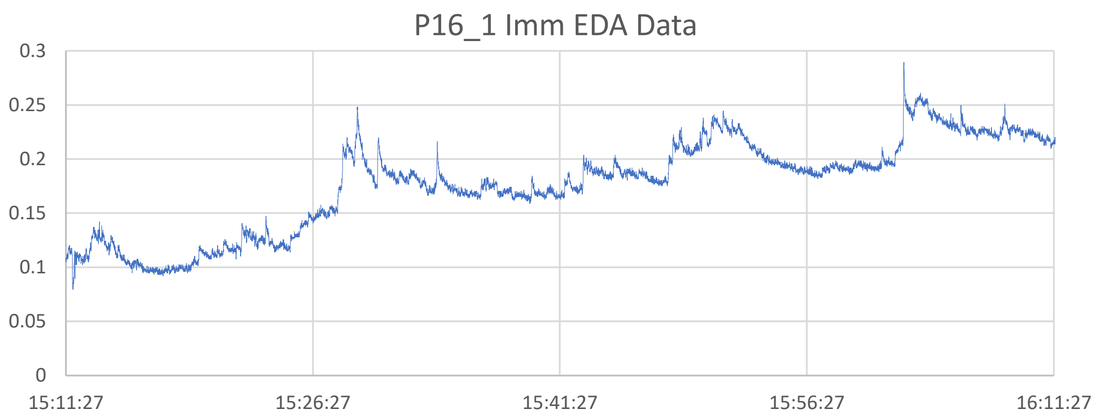

P13_2 Imm EDA Data
9
8
7
6
5
4
3
2
1
0
10:01:41
10:16:41
10:31:41
10:46:41
11:01:41

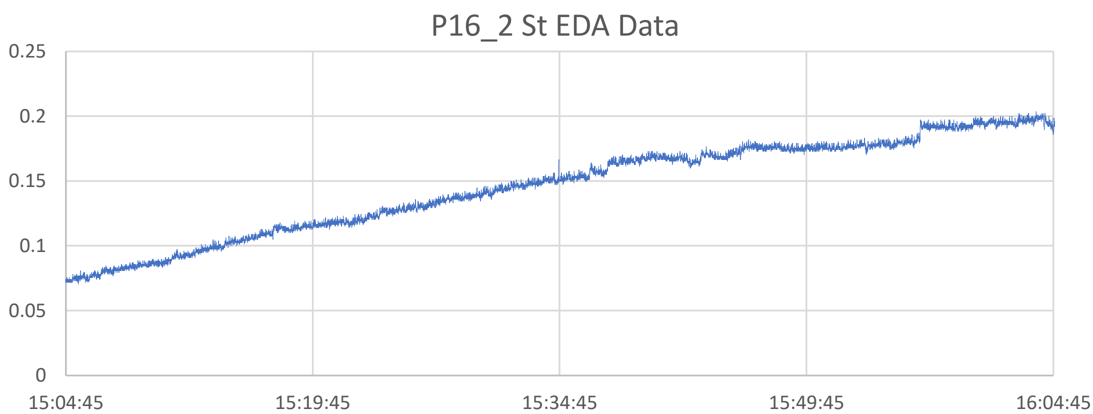

P14_1 Imm EDA Data
4
3.5
3
2.5
2
1.5
1
0.5
0
13:12:26
13:27:26
13:42:26
13:57:26
14:12:26

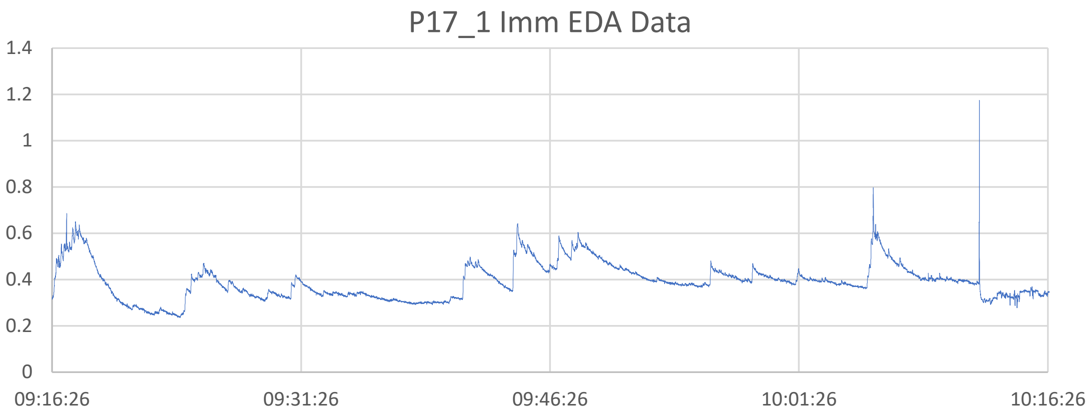

P14_2 St EDA Data
4
3.5
3
2.5
2
1.5
1
0.5
0
13:05:16
13:20:16
13:35:16
13:50:16
14:05:16

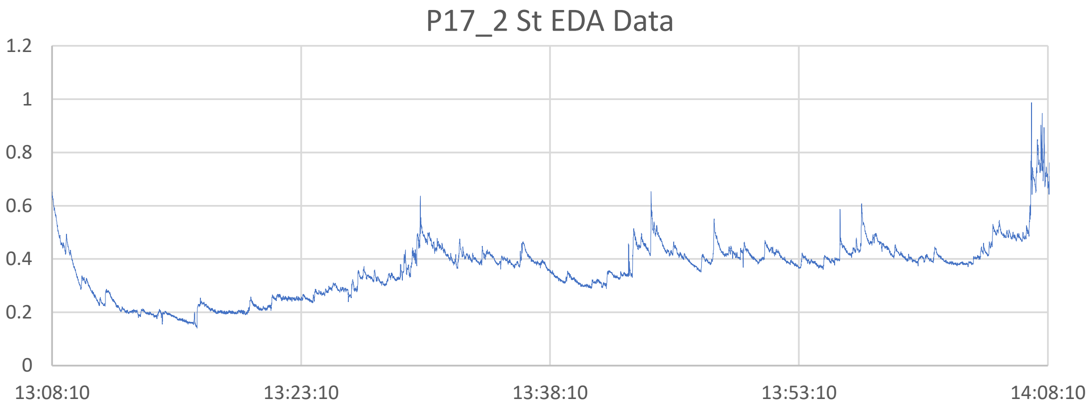

P15_1 St EDA Data
18
16
14
12
10
8
6
4
2
0
15:02:24
15:17:24
15:32:24
15:47:24
16:02:24

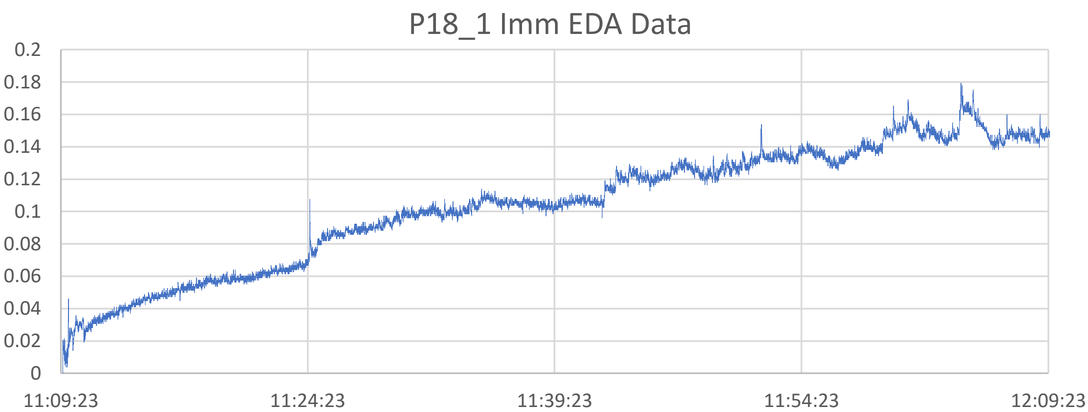

P15_2 Imm EDA Data
18
16
14
12
10
8
6
4
2
0
15:04:05
15:19:05
15:34:05
15:49:05
16:04:05

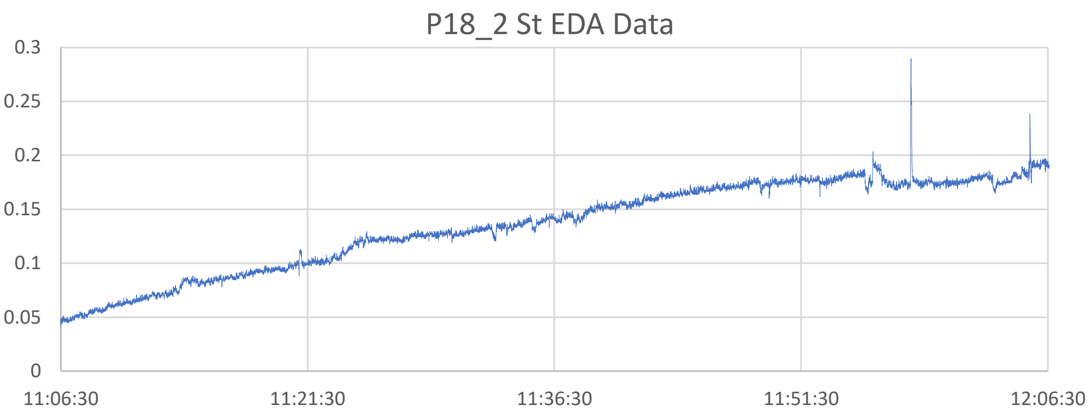

P16_1 Imm EDA Data
0.3
0.25
0.2
0.15
0.1
0.05
0
15:11:27
15:26:27
15:41:27
15:56:27
16:11:27

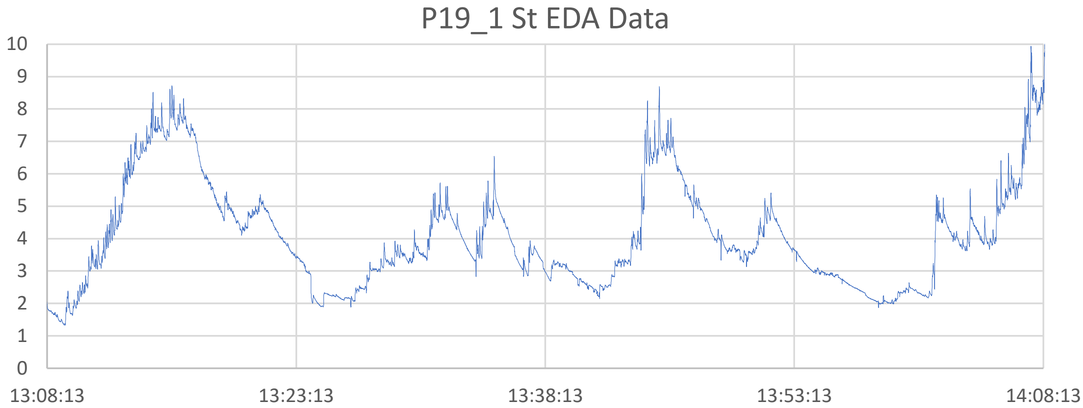

P16_2 St EDA Data
0.25
0.2
0.15
0.1
0.05
0
15:04:45
15:19:45
15:34:45
15:49:45
16:04:45

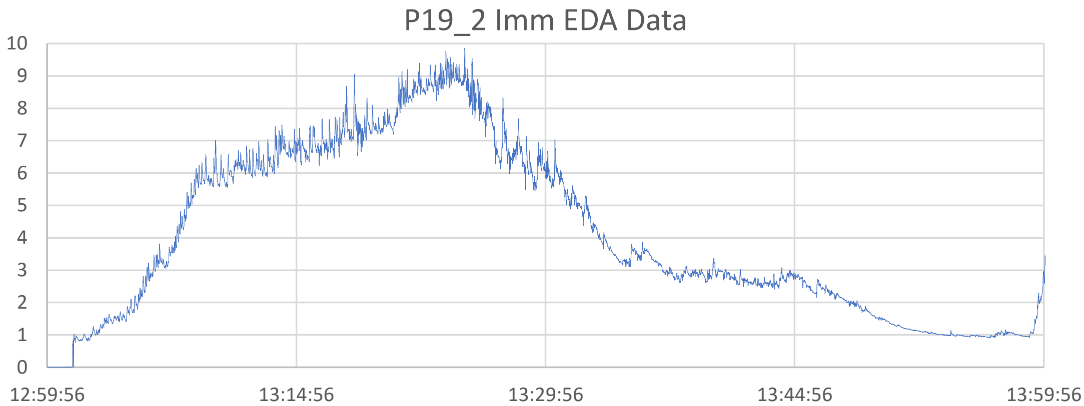

P17_1 Imm EDA Data
1.4
1.2
1
0.8
0.6
0.4
0.2
0
09:16:26
09:31:26
09:46:26
10:01:26
10:16:26

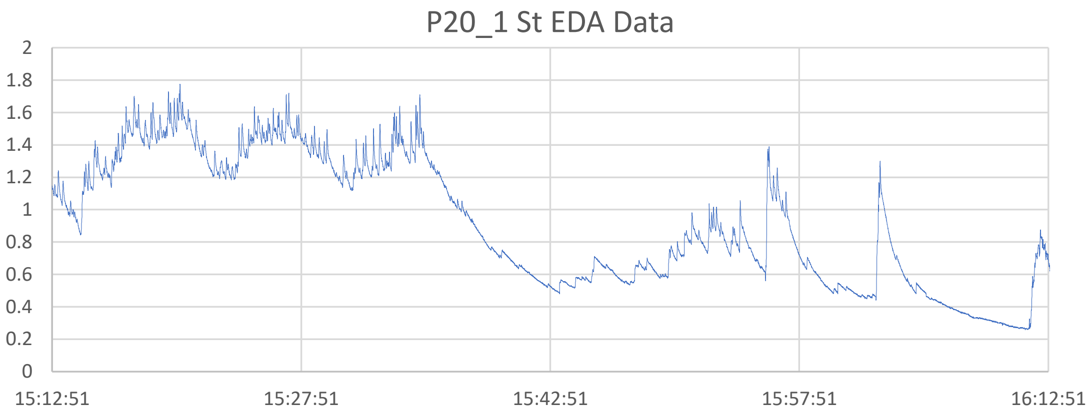

P17_2 St EDA Data
1.2
1
0.8
0.6
0.4
0.2
0
13:08:10
13:23:10
13:38:10
13:53:10
14:08:10

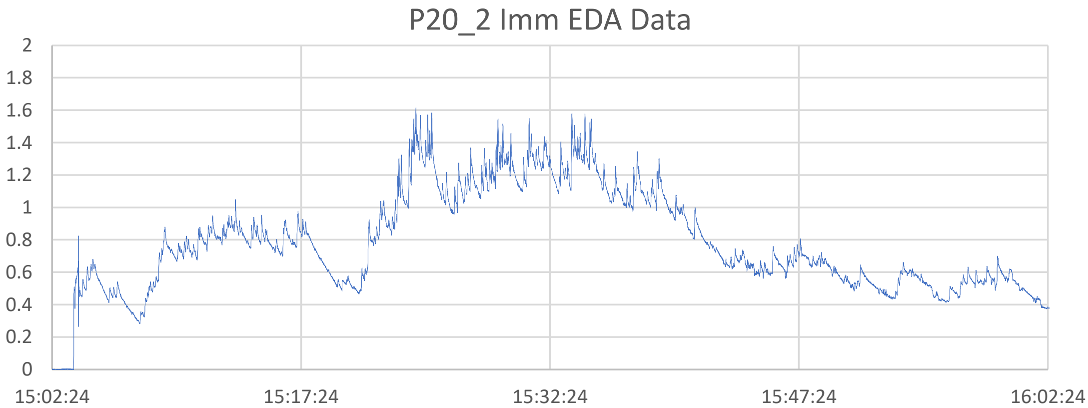

P18_1 Imm EDA Data
0.2
0.18
0.16
0.14
0.12
0.1
0.08
0.06
0.04
0.02
0
11:09:23
11:24:23
11:39:23
11:54:23
12:09:23

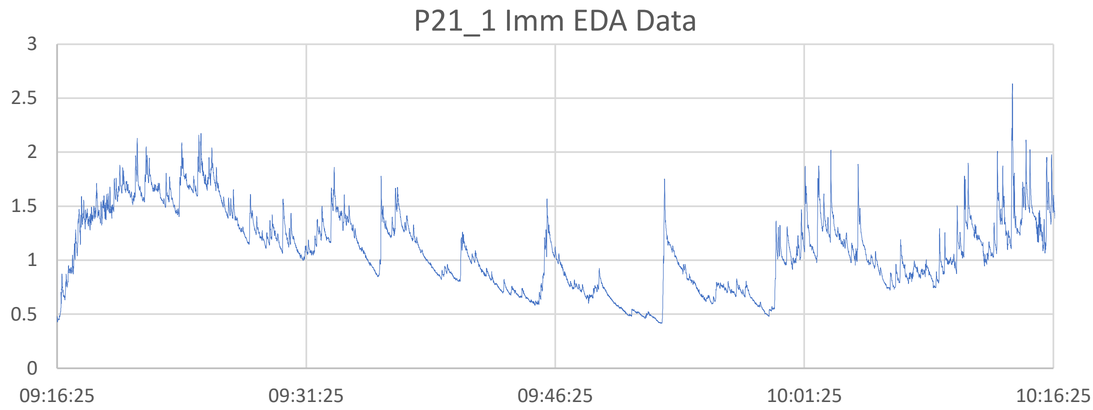

P18_2 St EDA Data
0.3
0.25
0.2
0.15
0.1
0.05
0
11:06:30
11:21:30
11:36:30
11:51:30
12:06:30

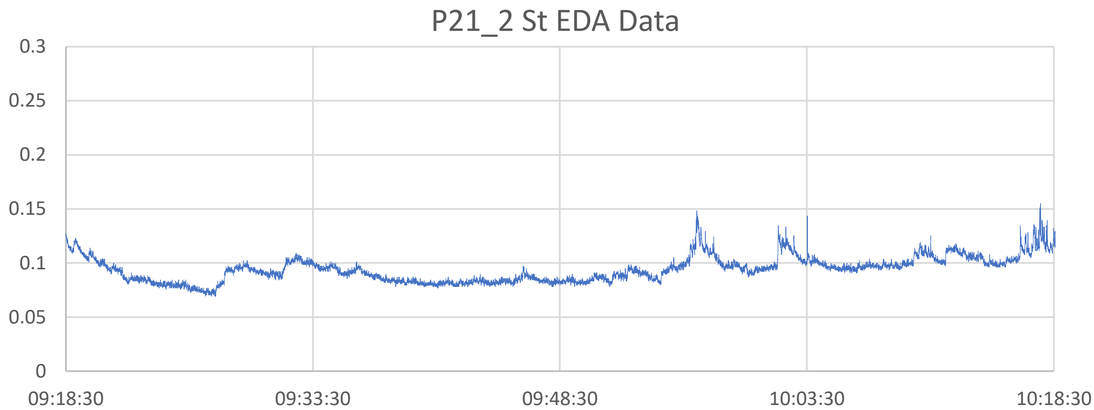

P19_1 St EDA Data
10
9
8
7
6
5
4
3
2
1
0
13:08:13
13:23:13
13:38:13
13:53:13
14:08:13

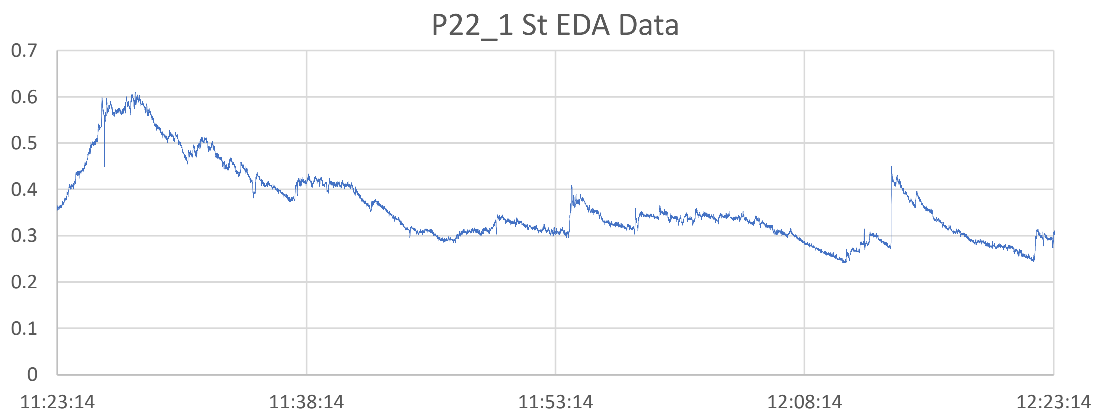

P19_2 Imm EDA Data
10
9
8
7
6
5
4
3
2
1
0
12:59:56
13:14:56
13:29:56
13:44:56
13:59:56

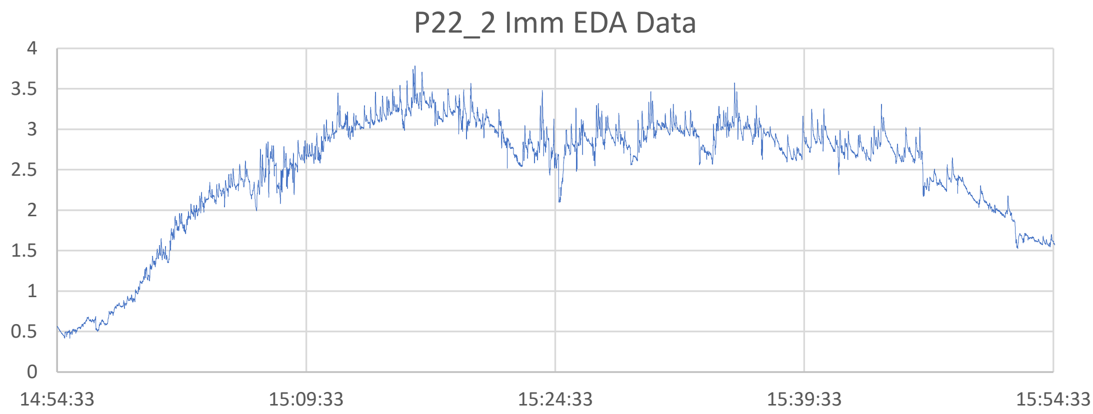

P20_1 St EDA Data
2
1.8
1.6
1.4
1.2
1
0.8
0.6
0.4
0.2
0
15:12:51
15:27:51
15:42:51
15:57:51
16:12:51

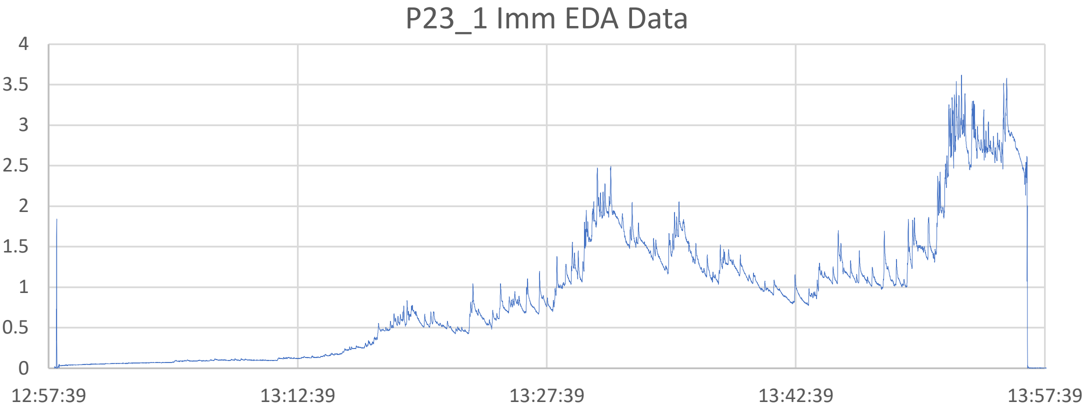

P20_2 Imm EDA Data
2
1.8
1.6
1.4
1.2
1
0.8
0.6
0.4
0.2
0
15:02:24
15:17:24
15:32:24
15:47:24
16:02:24

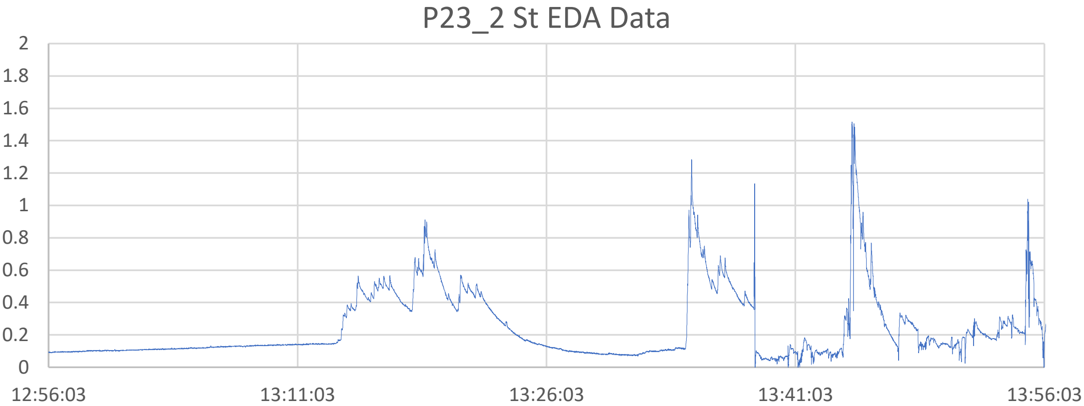

P21_1 Imm EDA Data
3
2.5
2
1.5
1
0.5
0
09:16:25
09:31:25
09:46:25
10:01:25
10:16:25

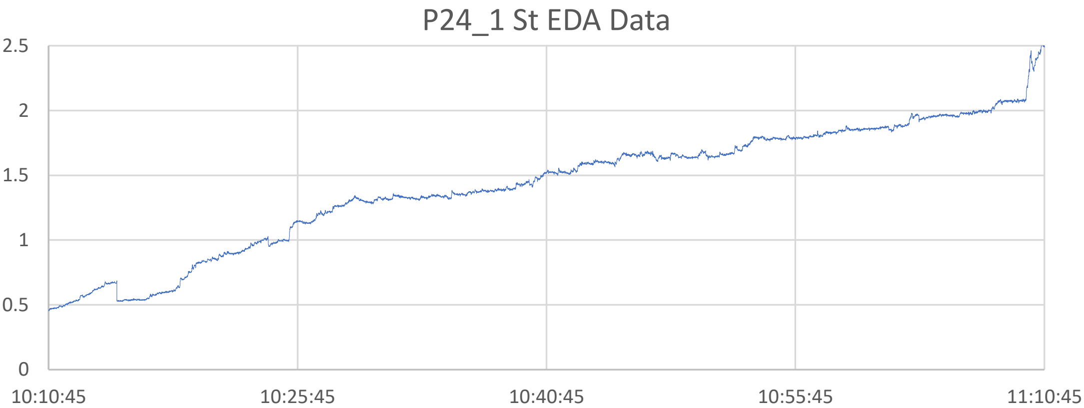

P21_2 St EDA Data
0.3
0.25
0.2
0.15
0.1
0.05
0
09:18:30
09:33:30
09:48:30
10:03:30
10:18:30

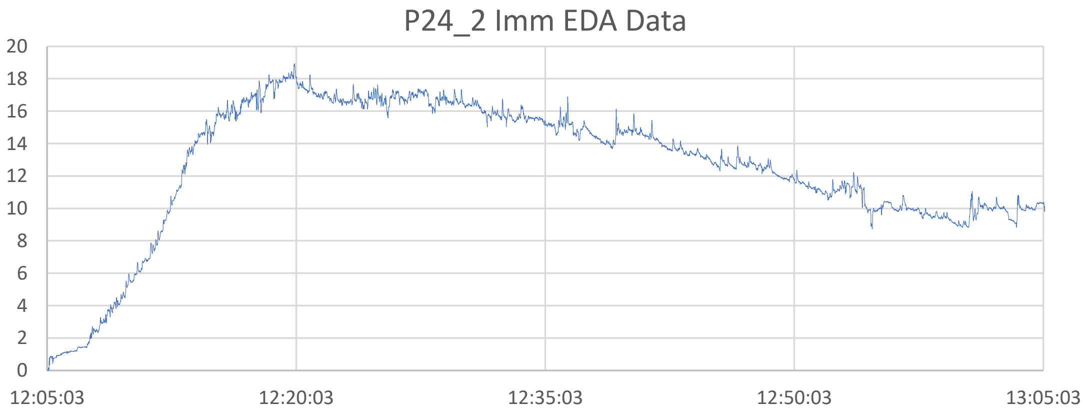

P22_1 St EDA Data
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
11:23:14
11:38:14
11:53:14
12:08:14
12:23:14


P22_2 Imm EDA Data
4
3.5
3
2.5
2
1.5
1
0.5
0
14:54:33
15:09:33
15:24:33
15:39:33
15:54:33


P23_1 Imm EDA Data
4
3.5
3
2.5
2
1.5
1
0.5
0
12:57:39
13:12:39
13:27:39
13:42:39
13:57:39


P23_2 St EDA Data
2
1.8
1.6
1.4
1.2
1
0.8
0.6
0.4
0.2
0
12:56:03
13:11:03
13:26:03
13:41:03
13:56:03


P24_1 St EDA Data
2.5
2
1.5
1
0.5
0
10:10:45
10:25:45
10:40:45
10:55:45
11:10:45


P24_2 Imm EDA Data
20
18
16
14
12
10
8
6
4
2
0
12:05:03
12:20:03
12:35:03
12:50:03
13:05:03