# Factors of Haptic Experience across Multiple Haptic Modalities

by

Ahmed Anwar

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Master of Applied Science
in
Management Sciences

Waterloo, Ontario, Canada, 2022

## Author's Declaration

This thesis consists of material all of which I authored or co-authored: see Statement of Contributions included in the thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

## Statement of Contributions

This thesis outlines the process of constructing a novel scale to measure haptic experience. Scale development is a standardized process that involves three main steps, Item Development, Scale Development and Scale Evaluation [7]. More detail about Scale Development process can be obtained from Figure 3.1. The first stage of Item Development was completed by Sathiyamurthy et al. [58] and is summarised in chapter 3.

Using these items, Ahmed completed the next steps of the process, Scale Development (in chapter 4 and chapter 5) and Scale Evaluation (in chapter 6). Tianzheng Shi assisted with the data analysis steps which are presented in chapter 5 and chapter 6. Oliver Schneider, provided assistance throughout the study.

All of the work described in this thesis is a result of collaborative efforts in at least some capacity as described earlier. Even where the author contributed all work, there was often informal feedback from friends and colleagues. As such, this thesis will use the first-person plural, “we”, throughout.

## Abstract

Haptic Experience (HX) is a proposed set of quality criteria useful to haptics, with prior evidence for a 5-factor model with vibrotactile feedback. We report on an ongoing process of scale development to measure HX, and explore whether these criteria hold when applied to more diverse devices, including vibrotactile, force feedback, surface haptics, and mid-air haptics. From an in-person user study with 430 participants, exploratory factor analysis (EFA), and confirmatory factor analysis (CFA), we extract an 11-item and 4-factor model (Realism, Harmony, Involvement, Expressivity) with only a partial overlap to the previous model. We compare this model to the previous vibrotactile model, finding that the new 4-factor model is more generalized and can guide attributes or applications of new haptic systems. This can inform designers about the right quality criteria to use when designing or evaluating haptic devices.

**Acknowledgements**

I thank my supervisor Oliver Schneider for his utmost support and guidance throughout this research. I am grateful to him for providing all the right resources that helped me to complete this project. Moreover, my Masters was quite challenging for me and I am grateful to my supervisor for his support.

I would like to thank Ali Rizvi for helping me with the data collection for my study and Tianzheng Shi for assisting in the data analysis process for this study and writing the conference paper.

I would like to acknowledge the Haptic Lab and the Games Institute community for their support in piloting and providing resources for the user studies.

## Dedication

I dedicate my thesis to my supervisor Dr Oliver Schneider due to his utmost contributions, passion and commitment to the field of haptics. It has been my absolute privilege to work under his supervision that helped me to grow both academically and personally.

# Table of Contents

# List of Figures

# List of Tables

# Chapter 1

# Introduction

Haptic feedback refers to anything that a user feels with the sense of touch. It could be vibrations, force, temperature, pressure, or any other physical sensations. Haptic technology is increasingly used by designers with evidence that it enhances user experience (UX). For instance, mid-air haptic feedback has been shown to make UX more pleasant, creative, and predictable [45] while motion seats incorporating haptic feedback have shown to invoke better experience when measured by EEG and other psychological signals [52]. Similarly, there is evidence that haptic feedback in virtual environment (VR) can lead to increased presence [2]. However, despite the promising adoption of haptic technology, it is difficult to understand how it influences UX, an important consideration for designers to improve their designs.

Haptic designers currently use qualitative methods to understand the influence of haptic feedback on their designs. Schneider et al.'s exploration of haptic experience design found that haptic designers prefer small in-person tests to evaluate their designs, iterating until it just feels right [59]. This approach is time-consuming, costly and not scalable to larger and remote evaluative studies. Although some haptic designers have made use of general scales such as AttrakDiff [15], there is no formal evaluative tool that measures the unique constructs of haptic experience. Therefore, it is pertinent to develop a reliable and scalable instrument that lets haptic designers identify design parameters that require improvement.

In this paper, we report on the development of a novel scale to measure haptic experience (HX) proposed by Kim and Schneider [39]. The HX model outlines the important parameters that encompass the user experience of haptic interaction. It consists of four design parameters, four usability requirements and five experiential factors (Autotelics, Harmony, Immersion, Expressivity, Realism) also shown in Figure 1.1. Sathiyamurthy et

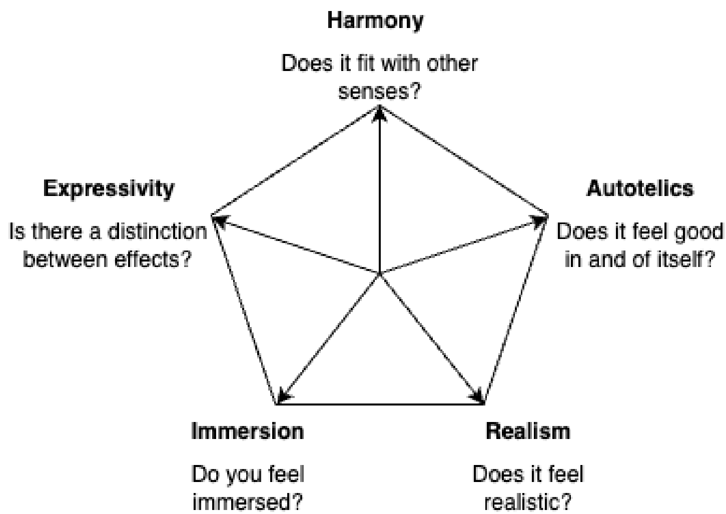


Figure 1.1: Five dimensional Haptic Experience model [39] that encapsulates the unique constructs of HX

al. [58] utilized these five dimensions to explore the HX model through a remote study incorporating only vibrotactile feedback and only conducted exploratory factor analysis.

These five dimensions also serve as guiding principles for designing our proposed instrument. However, we aim to explore the HX model with an in-person study incorporating different haptic feedback (vibrotactile, mid-air, force feedback, surface-haptics) and evaluate it with a confirmatory factor analysis. Therefore, this study is different from the one done by Sathiyamurthy, as it involves in-person survey administration, use of multiple haptic modalities and confirmatory factor analysis

Our aim is to support the development of a measurement instrument that can be used with any type of haptic device, and guide development of novel devices or experiences that work with multiple devices.

## 1.1 Research Question

### 1.1.1 Can haptic experience be effectively measured using a scale ?

The long-term goal of this research is to explore the feasibility of measuring and evaluating haptic experience across multiple modalities using a questionnaire. Using the existing literature, we aim to create a novel scale to effectively measure HX. This scale intends to provide haptic designers with a quantitative and scalable instrument to evaluate and improve their designs. In addition, it aims to support end users to effectively describe their interactions with haptic devices.

### 1.1.2 How different is generalized HX different from vibrotactile alone ?

Haptic feedback is rendered through various modalities. These include touch, mid-air, force, temperature and vibrotactile feedback. Therefore, it is important that an evaluative instrument should be able to encompass different modalities. In this regard, Sathiyamurthy et al. [58] obtained a 5-factor and 22-item model using only vibrotactile haptic feedback. However, in our study we aim to use multiple haptic feedback (force, touch, vibrotactile, mid-air) so that our scale could be used with wide range of haptic devices. To test the generalizability of our proposed scale, we aim to test whether HX is generalizable across multiple modalities or specific to any particular haptic feedback.

## 1.2 Contribution

Based on our results, we propose an 11-item model of HX with four experiential dimensions of “Harmony”, “Expressivity”, “Involvement” and “Realism”. Although the obtained model was confirmed using CFA, it needs further validation for use in practice. Our contributions are:

1. Evidence of a four factor HX [39] model built using different haptic modalities
2. A comparison of vibrotactile model (22 items and five factors) with the obtained multi-device model (11 items and four factors)
3. Guidelines for creating evaluative instruments to measure haptic experiences

## 1.3 Outline

This work is organized in the following sections:

- Chapter chapter 2 summarises the related work pertaining to measuring HX. It starts by differentiating haptic experience from user experience followed by scale development approaches in Human Computer Interaction (HCI) studies. Next, related instruments, their limitations and the need for developing a quantitative and scalable instrument to measure HX are discussed.

- Chapter 3 outlines the scale development approach, construct definition, item generation, face validity and construct validity.

- Chapter 4 explains the selection of appropriate haptic devices, in-person study design, survey administration to test the generated items and summary statistics of collected data (N=428).

- Chapter 5 provides detailed analysis for sampling adequacy, exploratory factor analysis, item refinement and factor extraction. We extract a 4-factor model consisting of 11 items. These four factors include “Harmony”, “Expressivity”, “Involvement” and “Realism”.

- Chapter 6 discusses the results of confirmatory factor analysis to test the appropriateness of the obtained 11-item model.

- Chapter 7 outlines the interpretation of the obtained factors, practical considerations for hapticians and limitations.

- Chapter 8 provides the conclusion for the thesis.

# Chapter 2

# Related Work

In this section we discuss the related work in existing literature that serves as the motivation for the construction of a scale to measure HX. We identify the research gaps and limitations of these works and stress the need for development of a novel scale to measure HX. We begin by outlining the differences between HX and UX. Then, we discuss the scale development process used in various HCI studies. Next, we outline existing overlapping scales used to measure UX, and their inadequacy to measure the unique constructs of HX.

## 2.1 Haptics and UX

UX has been defined by Schrepp et al. [61] as: *"a set of distinct quality criteria that includes classical usability criteria, like efficiency, controllability or learnability, and non-goal directed or hedonic quality criteria, like stimulation, fun-of-use, novelty, emotions, or aesthetics"*. User experience evaluation can be categorized with respect to pragmatic quality and hedonic quality. Pragmatic quality refers to the effectiveness and efficiency by which the user is able to achieve their goals [35]. On the other hand, hedonic quality means the non-pragmatic quality aspects of the product such as enjoyment [36]. UEQ [43] has been used in previous studies to evaluate UX based on three pragmatic factors, Perspicuity (easy to learn, easy to understand), Efficiency (fast, organized) and Dependability (predictable, secure) and two hedonic factors, Novelty (creative, innovative) and Stimulation (exciting, interesting). Similarly AttrakDiff2 [35] is also used for UX evaluation and consists of four factors, Attractiveness, Pragmatic Quality, Identity and Stimulation. Another evaluation framework called meCUE questionnaire [47] also evaluates UX based on instrumental and non-instrumental product qualities with a focus on a user's emotional response.

However, these existing scales are insufficient to measure the unique constructs of HX. HX is highly dependent on the context of the interaction and feedback from other modalities such as visual and auditory. In addition, existing UX models evaluate the product as a whole and are not able to measure the HX in isolation which is important to understand its efficacy in the system. Moreover, HX is pertinent to the moment of touch but existing UX evaluative instruments measure the experience before, during and after an interaction. This presents a need for an evaluative instrument that is able to measure the unique constructs of HX.

## 2.2 Scale development in HCI

Scale development is often used in HCI studies not just used to produce a quantitative evaluative instrument, but to provide insight into the construct of interest being studied. For instance, Baumgartner et al. developed a pictorial multi-item scale, called PSUS (Pictorial System Usability Scale), which aims to measure the perceived usability of mobile devices [3]. Bentvelzen et al. developed a scale called Technology-Supported Reflection Inventory (TSRI) that evaluates how effectively a system supports reflection [6]. TSRI enables researchers and practitioners to compare prototypes designed to support reflection. Brühlmann et al. developed User Motivation Inventory (UMI) to measure a user's motivation to engage with an interactive system [14]. Suh et al. created User Burden Scale (UBS) to measure the level of burden faced by a user while interacting with a computing system [66]. Similarly, Votipka et al. created and validated secure software development self-efficacy (SSD-SES) scale to measure software developers' belief in ability to perform vulnerability identification and mitigation as well as security communication tasks [71].

These scales are developed through a systematic scale development process. Boateng et al. [7] as outlined the best practices for scale development which involves three stages (Item development, scale development and scale evaluation). Item development step involves generating items for the intended scale from the theoretical construct. The scale development step includes administering the generated questions and extracting a model using Exploratory Factor Analysis (EFA). The final step i.e. scale evaluation involves testing the extracted model structure using Confirmatory Factor Analysis (CFA) and validity studies.

## 2.3 Haptics and related scales

There are existing scales in literature that have partial overlap with the experiential dimensions of HX. These are Need for Touch Scale (NFT) [53], Presence Questionnaire (PQ), Immersive Tendencies Questionnaire (ITQ) [72] and Haptic Fidelity Framework [48]. Need for Touch Scale is designed to measure the user's need of obtaining product information by Haptic feedback with respect to two dimensions - instrumental factors and autotelic factors. Instrumental factors utilize haptics to reflect product's textural properties and measure purpose-driven evaluations based on consumer preferences and the product using haptics. In contrast, autotelic factors are not purpose driven but mainly for hedonic purposes like enjoyment or sensory stimulation of touching a product. However, NFT is user-centered and focuses only on measuring the user's desire to touch, not the quality of the HX provided through touch.

Haptic feedback has been shown to increase presence in virtual environments [2] and two instruments exist to measure presence - Immersive Tendencies Questionnaire (ITQ) and Presence Questionnaire (PQ) [72]. Presence questionnaire measures the degree to which an individual experiences presence in a virtual environment (VE). PQ consists of 32 items constructed using four experiential factors i.e. control, sensory, distraction and realism. Control factors include the degree of control provided by the system to the user, the immediacy of control, anticipation, mode of control and physical environmental modifiability. Sensory factors include the sensory modality, environmental richness, multimodal presentation, consistency of multimodal information, degree of multimodal perception and active search. Distraction factors include isolation, selective attention and interface awareness. Realism factors include scene realism, consistency of information with the objective world, meaningfulness of experience and separation anxiety. These four factors are then further categorized into six different subscales of involvement/control, natural, auditory, haptic, resolution and interface quality. PQ overlaps with one of the experiential dimensions of HX, realism, but does not cover autotelics, harmony or expressivity. In addition, PQ is mostly restricted to measuring presence in VE so it is not suitable for haptic devices used without VE such as smartphones.

Similarly, ITQ is a 29 item scale that measures individuals's inclination to feel presence in a virtual environment. It consists of three subscales i.e. involvement, focus and games. Involvement subscale measures the individual's tendency to become involved in activities, focus subscale measures the user's tendency to maintain focus on current activities, and games subscale measures the user's tendency to play video games. However, ITQ is user centered and is intended to measures immersion at an individual level rather than of the haptic system.

Additionally, Muender et al. developed the Haptic Fidelity Framework to define the factors encapsulating realistic haptic feedback for virtual reality [48]. The framework constituted two dimensions; haptic fidelity and versatility. The Haptic Fidelity dimension measures the haptic feedback provided by the system with respect to 14 factors categorized into three categories (Sensing, Hardware and Software). The Versatility dimension, analyzes the specificity of the haptic feedback with respect to a particular application by rating the haptic feedback on a Likert-scale from specific to generic. For instance, systems that provide generic haptic feedback, could be repurposed to other applications such as vibrotactile feedback could be used to represent contact force as well as weight. However, this framework does not fully encapsulate the actual HX of a user as it does not take into account the individualist factors such as state of mind and past experiences. Moreover, this framework is not generalizable to all types of haptic modalities (e.g. force feedback, vibrotactile, mid-air, surface haptics) as it was constructed specifically for haptic feedback in virtual environments (VE).

Boos et al. [9] presented a modified version of UEQ scale by adding two additional sub-scales of acoustics and haptics. The acoustic sub-scale rates the product on the following four items: quiet-noisy, melodious-discordant, booming-muffled and shrilled-gentle. Similarly, the haptics sub-scale lets the users rate the product on the following four items: stable-unstable, pleasant-uncomfortable to touch, smooth-rough and slippery-non-slip. However, the scale was developed using an online study involving household appliances alone which limits its generalizability. Additionally, it does not encapsulate items pertaining to the five experiential dimensions of HX [39].

## 2.4 Haptics and Gaming Scales

Haptic feedback is utilized in gaming applications to enhance the user experience and there are many instruments used to measure the game user experience. Games Engagement Questionnaire [11] measures the impact of playing games based on engagement with respect to four factors - presence, flow, psychological absorption and immersion. However, this scale focuses on measuring gaming experience as a whole but not any specific haptic feedback rendered during gameplay. Moreover, this study has not been validated [44].

Similarly, another scale Player Traits Model [69] measures gaming experience based on player traits using five dimensions - challenge orientation, goal orientation, aesthetic orientation, narrative orientation and social orientation. However, this scale is user-centered and measures user preferences for game design but does not assess the HX vis-à-vis the game quality.

Another instrument called Player Experience Inventory [1] also exists that measures player experience with respect to functional consequences (audiovisual appeal, progress feedback, ease of control, challenge and goals) and psychological consequences (mastery, curiosity, immersion, autonomy, meaning). However, this scale also does not measure the constructs of HX as it is aimed for understanding and improving game design parameters based on the user's game experiences.

## 2.5 Chapter Summary

In this section we, outlined the existing instruments used to measure user experience. We outlined the scale development process used in various HCI studies and also identified relevant overlapping scales from literature review. Next, we distinguished HX from traditional UX and the inadequacy of existing instruments to encapsulate the unique constructs of HX. This serves as the motivation for the construction of a new scale to measure HX.

# Chapter 3

# Scale Development Approach

Scale development is a systematic process that is used to develop questionnaires in research studies. Various research questionnaires such as Pictorial System Usability Scale [3], Technology-Supported Reflection Inventory [6], User Motivation Inventory [14] and User Burden Scale [66] were built using this systematic process. Throughout this study we followed the scale development approach outlined by Boateng et al. [7]. The approach is summarized in Figure 3.1. It involves three main steps i.e. item development, scale development and scale evaluation. Item development stage involves, generating items for the intended scale from the theoretical construct. The Scale development step incorporates administering the generated questions and extracting a model using factor analysis. The final step i.e. Scale evaluation involves testing the extracted model structure using confirmatory factor analysis and validity studies.

In this chapter, we will discuss item development stage. These steps are:

- Construct Definition to determine the theoretical boundaries of the construct
- Item Generation which involves generating items from the theoretical construct.
- Content validity to assess if generated items adequately measure the domain of interest
- Face validity to ascertain if generated items aptly reflect the domain of interest and that answers produce valid measurements

All of the above mentioned steps (Construct Definition, Item Generation, Content Validity and Face Validity) were completed by Sathiyamurthy et al. [58] in an earlier study.

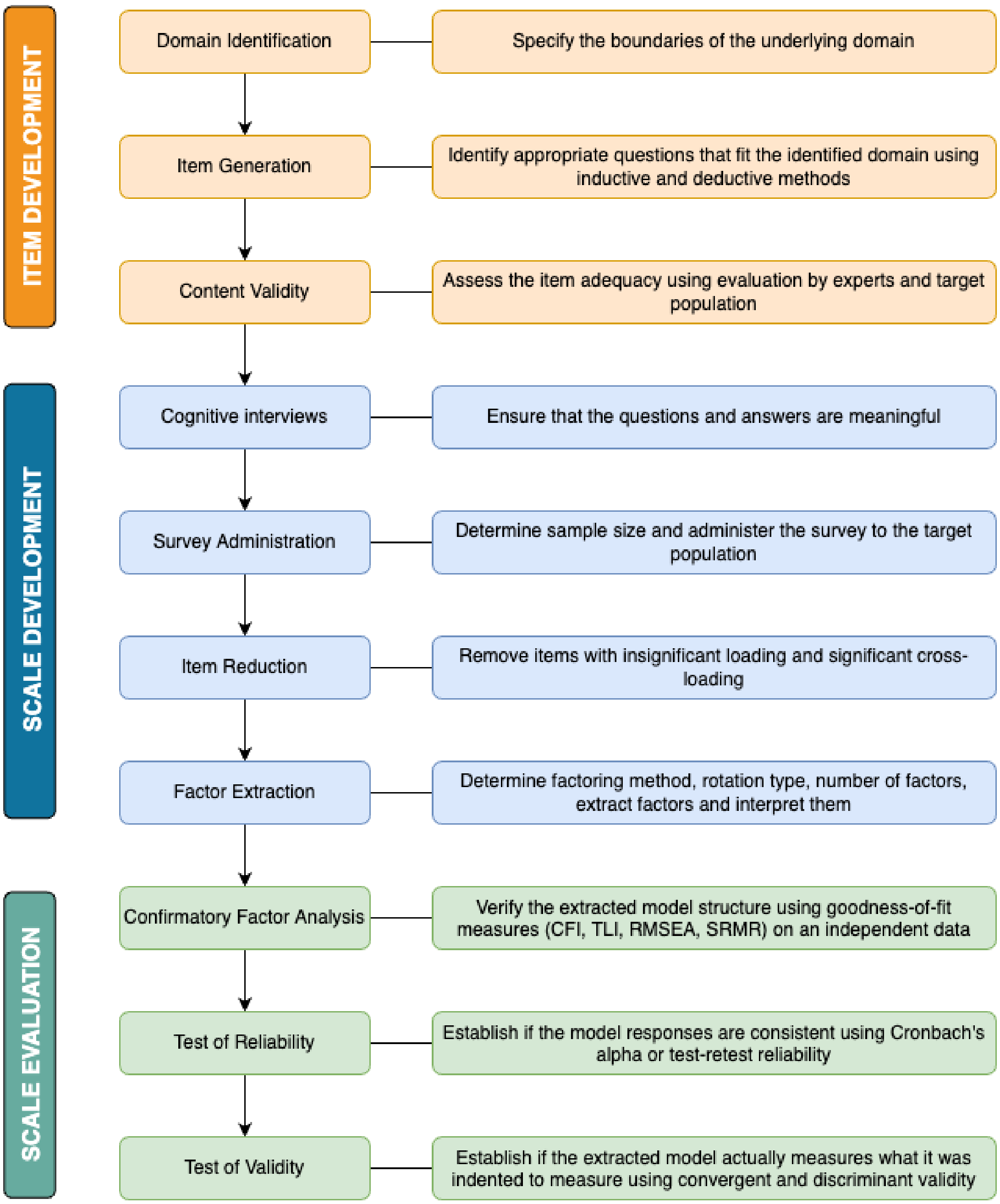


Figure 3.1: Scale Development Approach outlined by Boateng et al. [7]. For item development, we utilize the work of Sathiyamurthy et al. and complete the scale development and confirmatory factor analysis steps [58]

## 3.1 Construct Definition

The first step in scale construction is construct or domain identification which is to specify the theoretical boundaries of the domain [7]. For this, we leverage the work of Kim et al. [39] in which the researchers defined the scale construct by conducting literature review and two user studies. The first user study included a brainstorming session with haptic novices where participants were asked to interact with a haptic device and outline experiential factors that constitute haptic feedback. Factors were labeled and merged based on overlap. In the second study, six haptic experts were invited to critique, modify and rank the developed factors. As a result five experiential factors were defined that constitute Haptic experience: Autotelics, Immersion, Harmony, Realism and Expressivity [39]. These factors and their definitions are mentioned in Figure 3.2.

- Autotelics is a hedonic (non-task) oriented dimension measuring the user's experience of touch in and of itself. An example of this is a user reporting a positive experience simply due to the experience being pleasurable in and of itself.
- Immersion is intended to measure the level of engagement produced by the inclusion of haptic feedback in the system.
- Harmony measures how well the haptic feedback is integrated with the other sensory modalities (visual or auditory) of the system. An unharmonious (which we call disruptive) haptic feedback might lead to distraction and negative user experience to the user.
- Realism measures how believably, the haptic feedback conveys what the device or application was intended to convey. Realism could be achieved independently or in tandem with Harmony.
- Expressivity is the degree by which the user feels that their inputs make impact on the haptic feedback received. This is usually achieved when the haptic feedback is distinguishable with respect to the varying inputs entered by the user.

In addition, the researchers also create the following definition of HX that helps designers and researchers to distinguish it from UX models:

*"a distinct set of quality criteria combining usability requirements and experiential that involves one or more perceived senses of touch, possibly as part of a multi-sensory experience"*

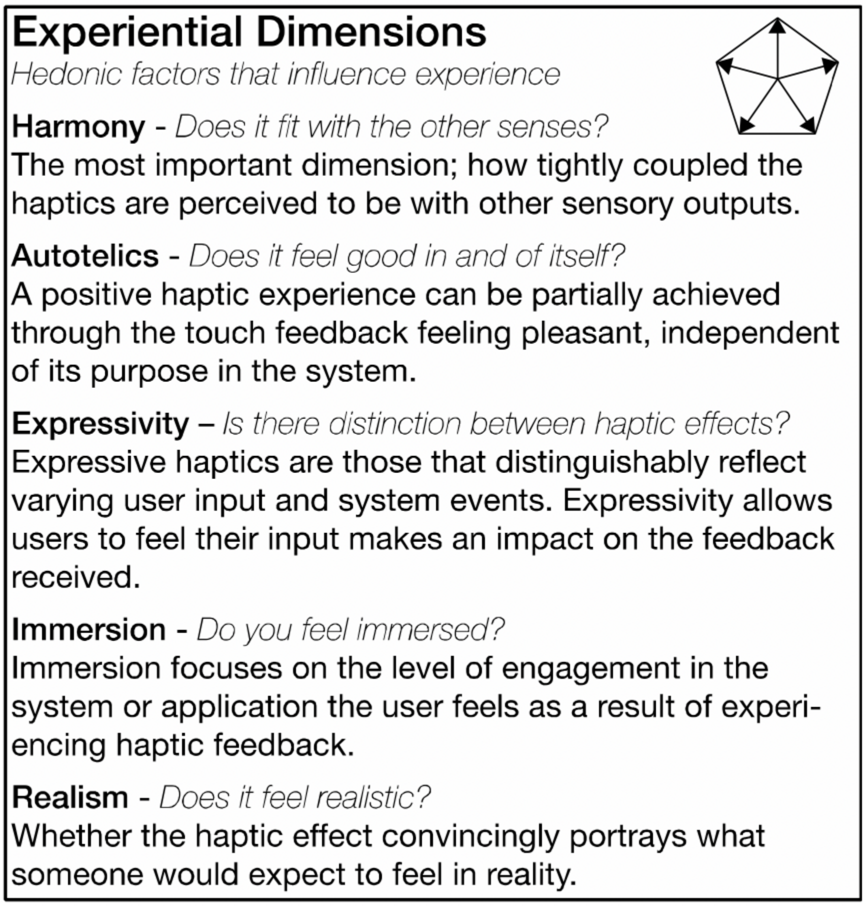


Figure 3.2: Experiential dimensions with definitions of Haptic Experience (HX) model created by Kim and Schneider[39]

This definition alongside the five theoretical constructs of HX, served as the bases for item generation step which is described in the following section.

## 3.2 Item Generation

The next step in scale inception is to generate items based on the previously defined construct. For this we utilized the work of Sathiyamurthy et al. in which the researchers generated 25 items (five per dimension) [58]. These items were then tested and refined to establish face content validity and face validity. Content validity is requried to ascertain if the generated items are adequately measuring the domain of interest. Face validity is required to ensure that generated questions aptly reflect the domain of interest.

## 3.3 Content validity

In this step Sathiyamurthy et al. established the content validity (assessing if genrated items adequately measure the domain of interest) of the initial 25-item questionnaire [58] through an evaluation study by experts. Haptic experts (N=6) were invited to critique the 25-items and asked to rate them based on clarity and their representativeness on the respective experiential dimension. This was followed by a focus-group session to reach a consensus which resulted in item revision or deletion.

The experts pointed out that some of the items would not be apt for some of the haptic devices as some of them are designed to portray only one or more of the five experiential dimensions of the HX model [39]. For example, the dimension of Expressivity would not be fitting for a dental simulator, resulting in a lower score for items pertaining to this dimension. This brought the dilemma of specificity vs generality. If the questionnaire is very broad then scores on some of the devices would be low. However, if the questionnaire is very specific then it would curb its generality. One possible solution to this was to let the end users decide on the dimensions appropriate to their device, but that would also hinder its generality. Since establishing generalizability was one of the objectives of developing this instrument, it was decided on maintaining the questionnaire broad and not restricting its wording to cater to specific haptic modalities. In addition, the inclusion of multiple types of haptic feedback vibrotactile, force-feedback and mid-air would enhance the generalizability of our instrument.

Another issue pointed out was that some of the items were ambiguous which makes it difficult to link it to its indented experiential dimension or to distinguish it from the overall experience of the device. This meant that the score obtained would not only contain the quality of haptic feedback but also other components of the experience. To address this issue, 7 ambiguous items were removed ad 18 items were modified to enhance their clarity and relevance. Four new items were added based on the expert feedback. In addition, questions were reworded to incorporate "haptic feedback" to ensure that it they are intended to measure the haptics aspect. For example, "The feedback was convincing" was reworded to "The haptic feedback was convincing". Furthermore, the questionnaire was modified, so that it includes the definition of haptic feedback [39] to make it clearer to users what we are trying to measure. Following this step, 22 items were left and content validity was established.

## 3.4 Face validity

Following expert feedback, the next step is to establish the face validity of the generated 22-items through evaluation by the target population. Face validity is required to ensure that the generated questions aptly reflect the domain of interest and that answers produce valid measurements. For establishing face validity, one-to-one interviews were conducted with participants (N=9) with varying professional backgrounds and familiarity with haptics. Participants were asked to interact with a haptic device, complete 22-item questionnaire whilst rating each item in terms of difficulty to understand. Participants reported that the definition of Haptics enhanced their understanding while completing the questionnaire. Some of the items were modified, however no items were dropped at this stage. The finalized items are shown in the Table Table 3.1.

## 3.5 Chapter Summary

In this chapter, we discussed the scale development approach outlined by [7]. We described the two initial steps in scale inception construct identification and item generation. In construct identification stage, we summarise the five dimensional (Autotelics, Harmony, Realism, Immersion, Expressivity) HX model developed by Kim et al. [39]. Subsequently, we describe the 22-items generated by Sathiyamurthy et al. [58] based on the 5-factor HX model. The next steps involve testing the theoretical model with exploratory factor analysis (EFA) which we discuss in the next chapter.

| Sub-scale | Item |
|---|---|
| A1 | The haptic feedback felt satisfying. |
| A2 | I like how the haptic feedback itself feels, regardless of its role in the system. |
| A3 | I disliked the haptic feedback |
| A4 | I would prefer the system without the haptic feedback |
| E1 | The haptic feedback all felt the same |
| E2 | I felt adequate variations in the haptic feedback |
| E3 | The haptic feedback helped me distinguish what was going on |
| E4 | The haptic feedback changes depending on how things change in the system |
| E5 | The haptic feedback reflects varying inputs and events |
| I1 | The haptic feedback distracted me from the task |
| I2 | I felt engaged with the system due to the haptic feedback |
| I3 | The haptic feedback helped me focus on the task |
| I4 | The haptic feedback increased my involvement in the task |
| R1 | The haptic feedback was realistic |
| R2 | The haptic feedback was believable |
| R3 | The haptic feedback was convincing |
| R4 | The haptic feedback matched my expectations |
| H1 | The haptic feedback fits well with the other senses |
| H2 | I like having the haptic feedback as part of the experience |
| H3 | The haptic feedback felt disconnected from the rest of the experience |
| H4 | The haptic feedback felt appropriate when and where I felt it |
| H5 | The haptic feedback felt out of place |

Table 3.1: Initial 22-items after content validity for the five dimensional HX model where A = Autotelics, E = Expressivity, I = Immersion, R = Realism, H = Harmony

# Chapter 4

# Survey Administration: User Study

In this section we outline the “Survey Administration” stage of the scale development process as shown in Figure 3.1. This stage is important in order to test the theoretical model developed from the previous steps. We describe the steps taken in designing the study, determining the ideal sample size, summary statistics and data cleaning. This stage is independent of the work done by Sathiyamurthy et al. [58]

## 4.1 In-person vs remote studies

In other scale development studies such as User Motivation Inventory [14], Pictorial System Usability Scale [3] and User Burden Scale [66] researchers have used crowdsourcing platforms such as Amazon Mechanical Turk to recruit participants. Although, research studies have shown that crowdsourcing tools give promising results for conducting large scale haptic based studies [60], we opted to conduct our study in-person because our proposed instrument is intended to be applicable to all types of Haptic devices. However, these devices are not readily accessible to people due to cost constraints. Hence, it was essential to conduct the study in-person to ensure the generalizability of our proposed evaluative instrument.

## 4.2 Determining Sample Size

In previous studies, establishing the sample size for exploratory factor analysis (EFA) has been the subject of debate. Some researchers have recommended basing the sample size on the number of items [10, 17, 27] in the preliminary scale with minimum sample size between 100 and 200 [27, 20, 21]. For instance, Nunnally et al. recommended to have at least 10 respondents per scale item (ratio of 10:1 for respondents to item) [50], and Hair et al. recommended having 5 participants for each scale items [32].

On the other hand, some researchers have suggested that the sample size should be independent of the number of survey items. For instance, Clark et al. [19] recommended using a sample size of 300 whereas Gudagnoli et al. [29] recommended a sample size of 200-300. Comrey et al. weighted the sample size (100 = poor, 200 = fair, 300 = good, 500 = very good, $\geq$ 1000 = excellent) [21].

After extracting factors from EFA, we needed to confirm our model structure using Confirmatory Factor Analysis (CFA) on a independent sample. CFA is important to test whether the data fits the hypothesized model obtained from EFA. Previous research studies have recommended a sample size of around 100 for CFA [8]. Bentvelzen et al. used a sample size of 507 for EFA and 498 for CFA [6] whereas Votipka et al, used a sample size of 157 for EFA and 162 for CFA. Therefore, we also took the sample size requirement for the CFA and aimed to collect around 400 responses. Taking all this together, we aimed for collecting approximately N=300 for EFA and N=100 for CFA which meets the sample size requirements.

## 4.3 Device and demo selection

Our aim is to support the development of a quantitative instrument that enables haptic designers to quickly evaluate and compare their designs in a standardized way at scale. These designs could constitute of one or more types of haptic feedback. It is therefore important that the scale should be generalizable to encompass and effectively measure different haptic feedback; moreover, we believe HX needs to be explored in the context of more diverse devices. Hence, we opted to use five different devices, Haply 2diy [1], Ultraleap Stratos Explore [2], Oculus Quest 2 [3], TanvasTouch [4], 3D Systems Touch [5].

Together, these devices constitute four different types of haptic feedback: force-feedback, vibrotactile, mid-air and surface haptics. For each device, we selected demos that exhibit

[1] https://2diy.haply.co/
[2] https://www.ultraleap.com/product/stratos-explore/
[3] https://store.facebook.com/ca/quest/products/quest-2
[4] https://tanvas.co/products/tanvastouch-dev-kit
[5] https://www.3dsystems.com/haptics-devices/touch

| **Device** | **Feedback** | **Demo Application** |
|---|---|---|
| Haply | force | Maze |
| Tanvas | surface | Cloth Texture |
| Ultraleap | mid-air | Pressing Button |
| Oculus | vibrotactile | Beat Saber |
| 3D Touch | force | Jenga |

Table 4.1: Haptic devices and demos used

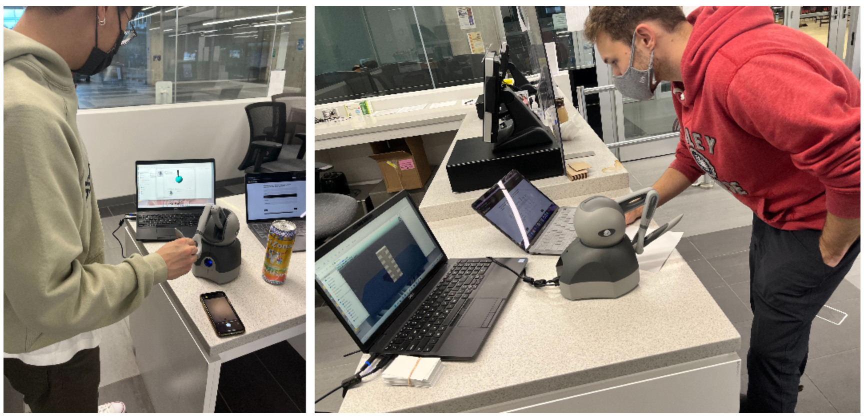

Figure 4.1: Study site situated in Coffee and Donut shop in E7. Participants interacted with the provided haptic device and then filled the questionnaire

the respective haptic feedback of the device to the user. These demos were selected based on the feedback obtained in a pilot study involving 6 participants. For Haply, we opted for a provided maze game in which user has to move from a start position to an end position on the screen using the end-effector while experiencing force feedback. For TanvasTouch, we opted for a clothing texture demo that lets user feel the difference between two different cloth textures using surface haptics. For Ultraleap, we selected the demo where participants had to press a button and move a slider using their hand while experiencing mid-air haptic feedback. For Oculus, we selected beat saber game, where user experiences vibrotactile feedback. For 3D Systems Touch, we opted for the provided Jenga game where user experiences force feedback while lifting blocks. A summary of devices and the respective demos used is provided in Table 4.1.

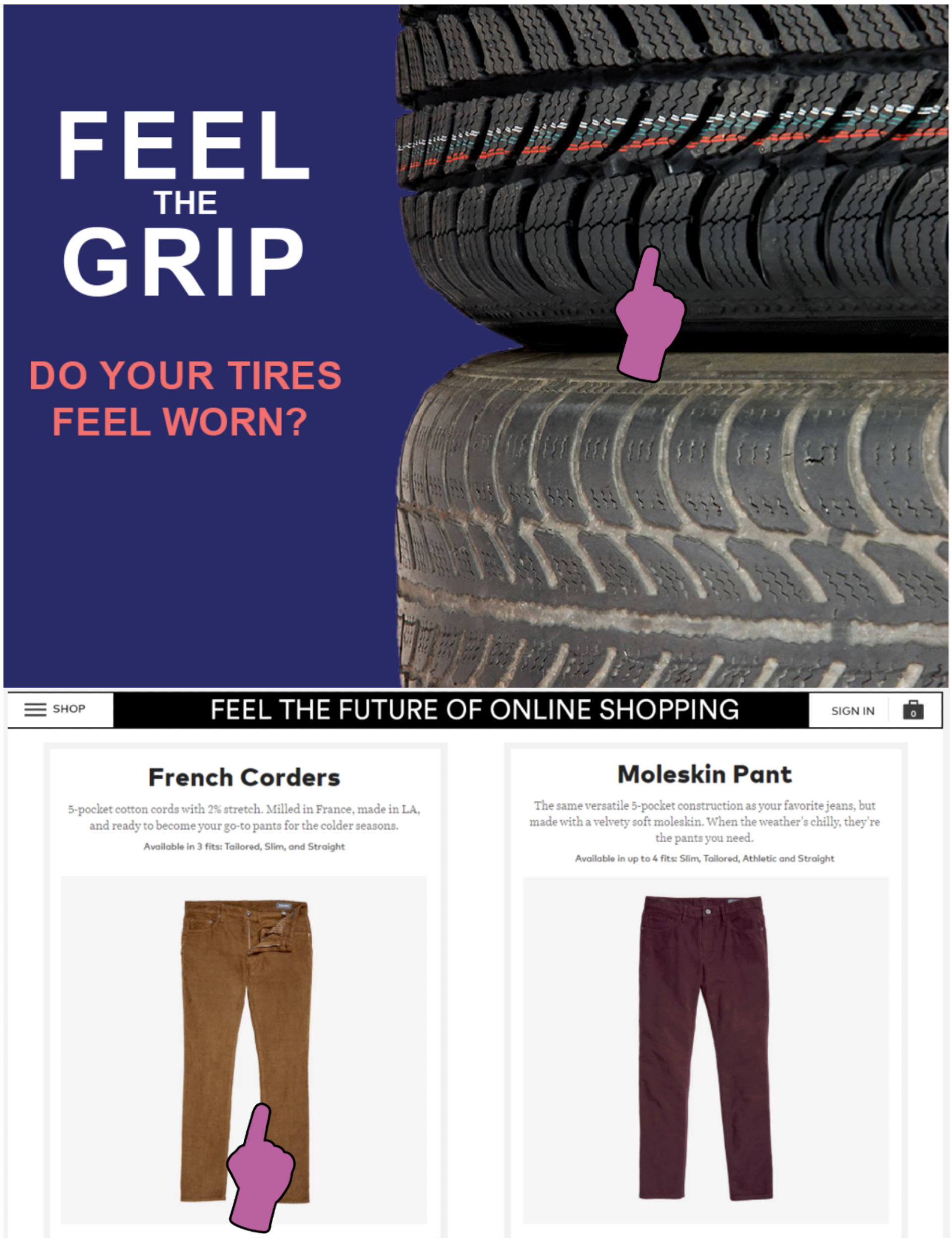


Figure 4.2: Demos shown to users on Tanvas which provides surface haptic feedback using electroadhesion technology. With the demos, it provides different degrees of haptic feedback to represent the texture of a rugged or rough clothing texture to the user.

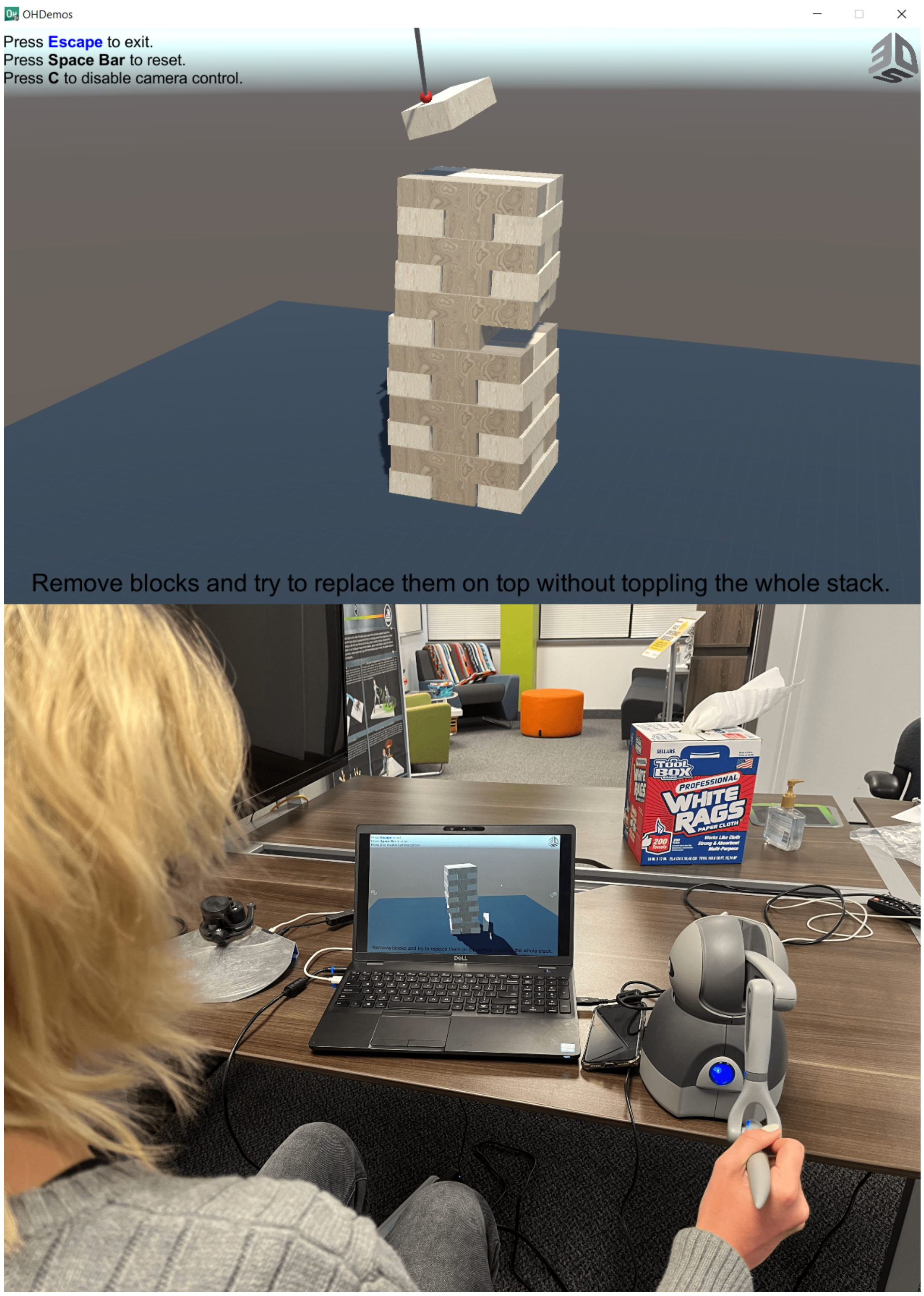


Figure 4.3: Demo used with 3D Touch. The device provides force-feedback to the user when lifting blocks in the jenga game.

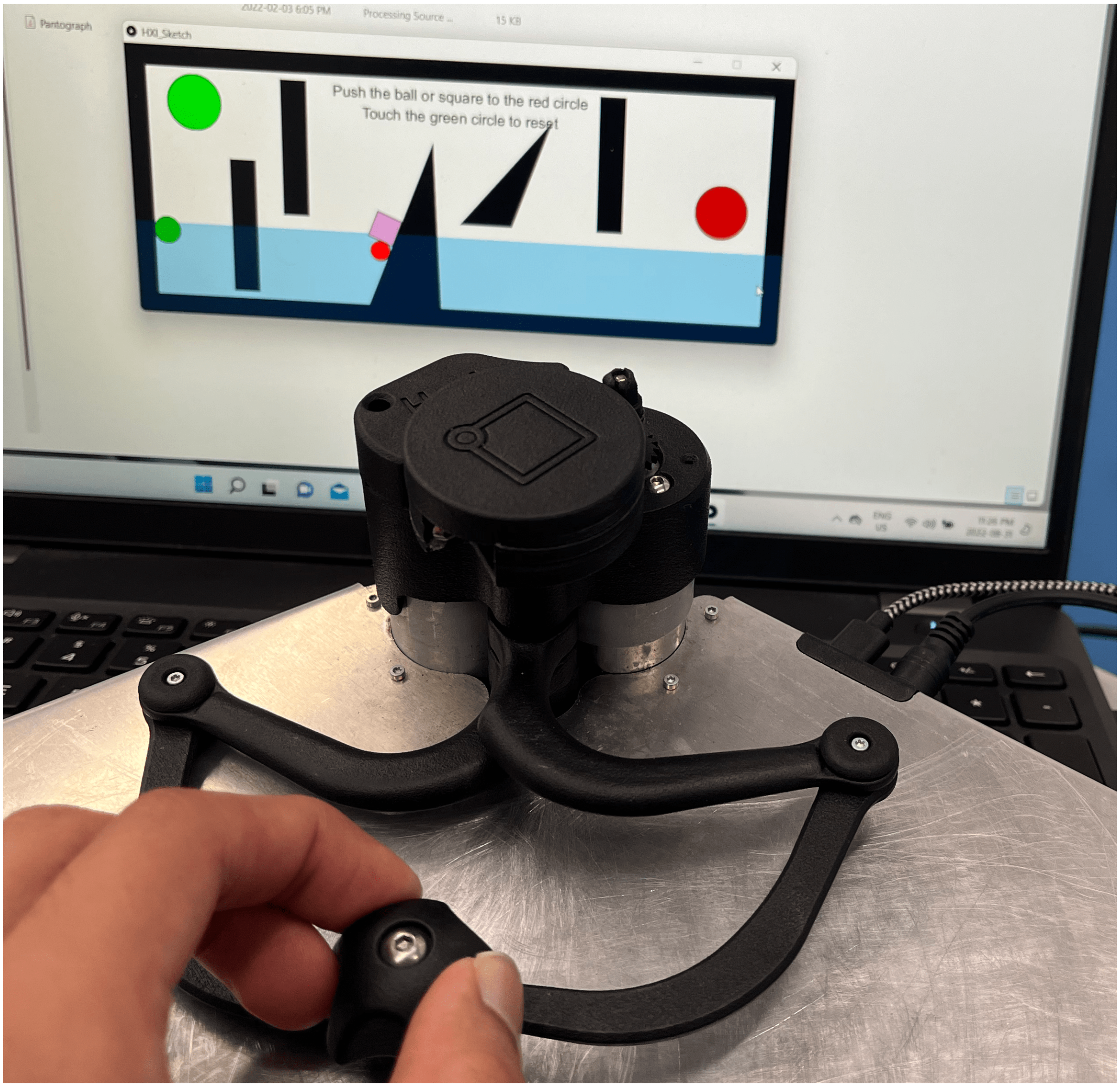


Figure 4.4: Maze demo used with the Haply device which provides force feedback to the user from the end effector. The user feels greater force feedback when in the water or when moving the object uphill.

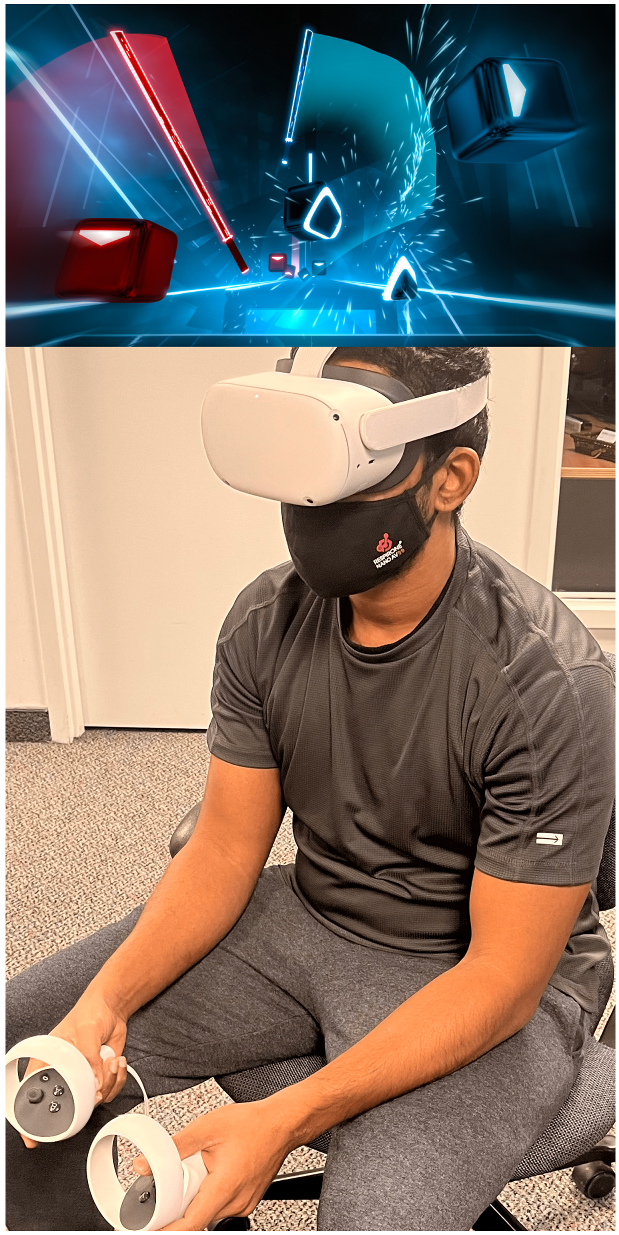

Figure 4.5: Oculus demo which provides vibrotactile haptic feedback from its hand-held controllers when the user strikes a block in a VR game called Beat Saber

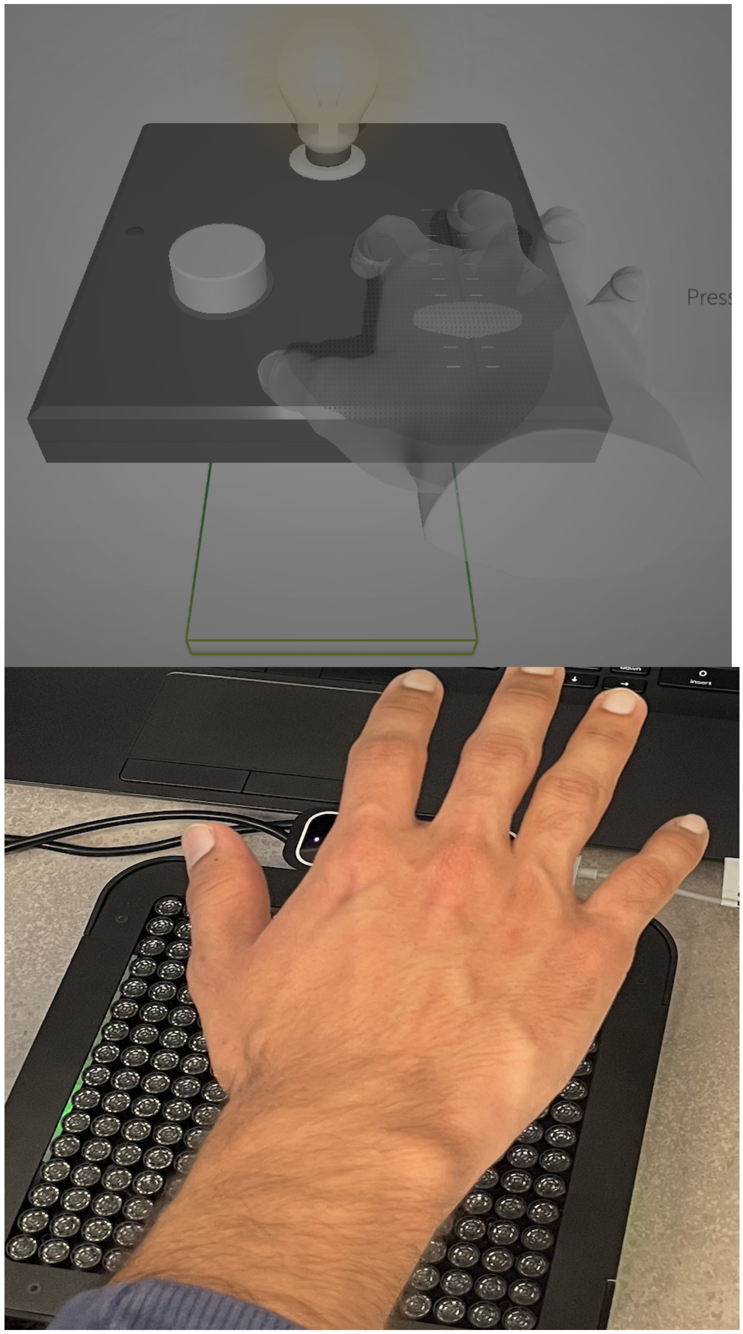

Figure 4.6: Ultraleap demo which provides mid-air haptic feedback using ultrasound sensors when moving the slider and pressing the button

## 4.4 Survey Administration

We conducted our in-person user study at the University of Waterloo and recruited 304 participants from Feb 2022 to April 2022 for EFA and 126 participants from June 2022 to July 2022 for CFA. Participants belonged to various academic fields such as Engineering, Arts and Science. We describe our sample demographics in subsequent section. For remuneration, participants were given $2 gift voucher for an on-campus coffee shop (Coffee and Donut in Engineering 7). To reach out to maximum people, different recruitment methods were used such as in-class recruitment and walk-in participation. Each participant interacted with a single haptic device for approximately 5 minutes and was asked to complete a small predefined task on it. Subsequently they were asked to complete the 22-item questionnaire on Qualtrics. For the ease of participants, option to complete the survey on a smart phone was also included. The following steps were created in Qualtrics to facilitate the study.

- Review and give consent to the study
- Select haptic device from a drop-down menu
- Answer demographic questions
- Complete the task/application using the chosen device
- Complete the questionnaire (22 items in randomized order on a Likert scale)
- Answer exit survey (rate experience)

The exported survey and the survey link have been included in Appendix B.

## 4.5 Data Cleaning and Split

We collected two independent samples and removed responses that were either incomplete or had a completion time of fewer than 2 minutes. In total, we removed 27 data points from both samples combined. After this step, we were left with a sample (N=291) for Exploratory Factor Analysis (EFA) and an independent sample (N=112) for Confirmatory Factor Analysis (CFA). Next, we reverse-coded the negatively phrased items (A3, A4, E1, I1, H3, H5).

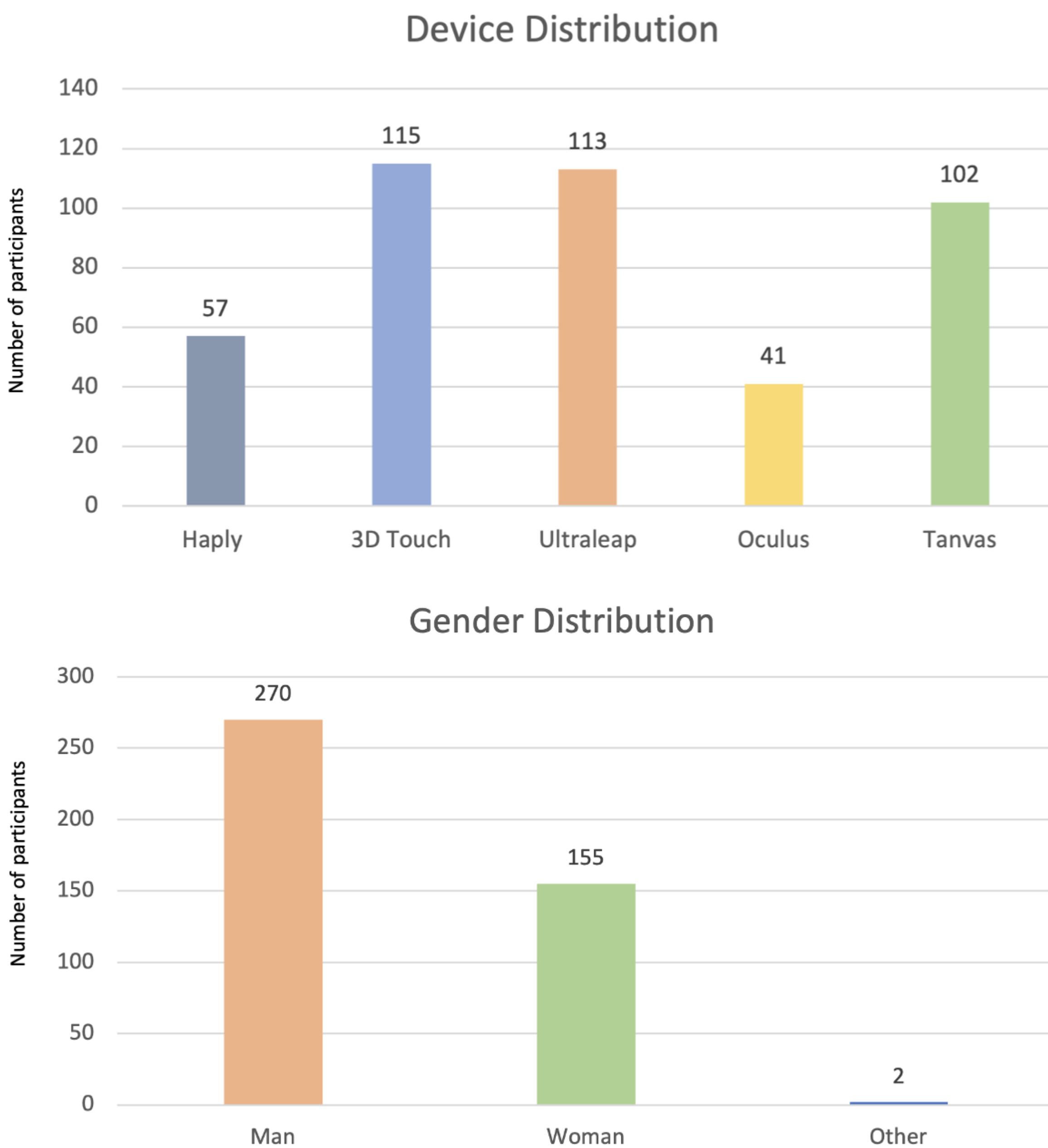


Figure 4.7: Demographic description of the entire sample (N=428)

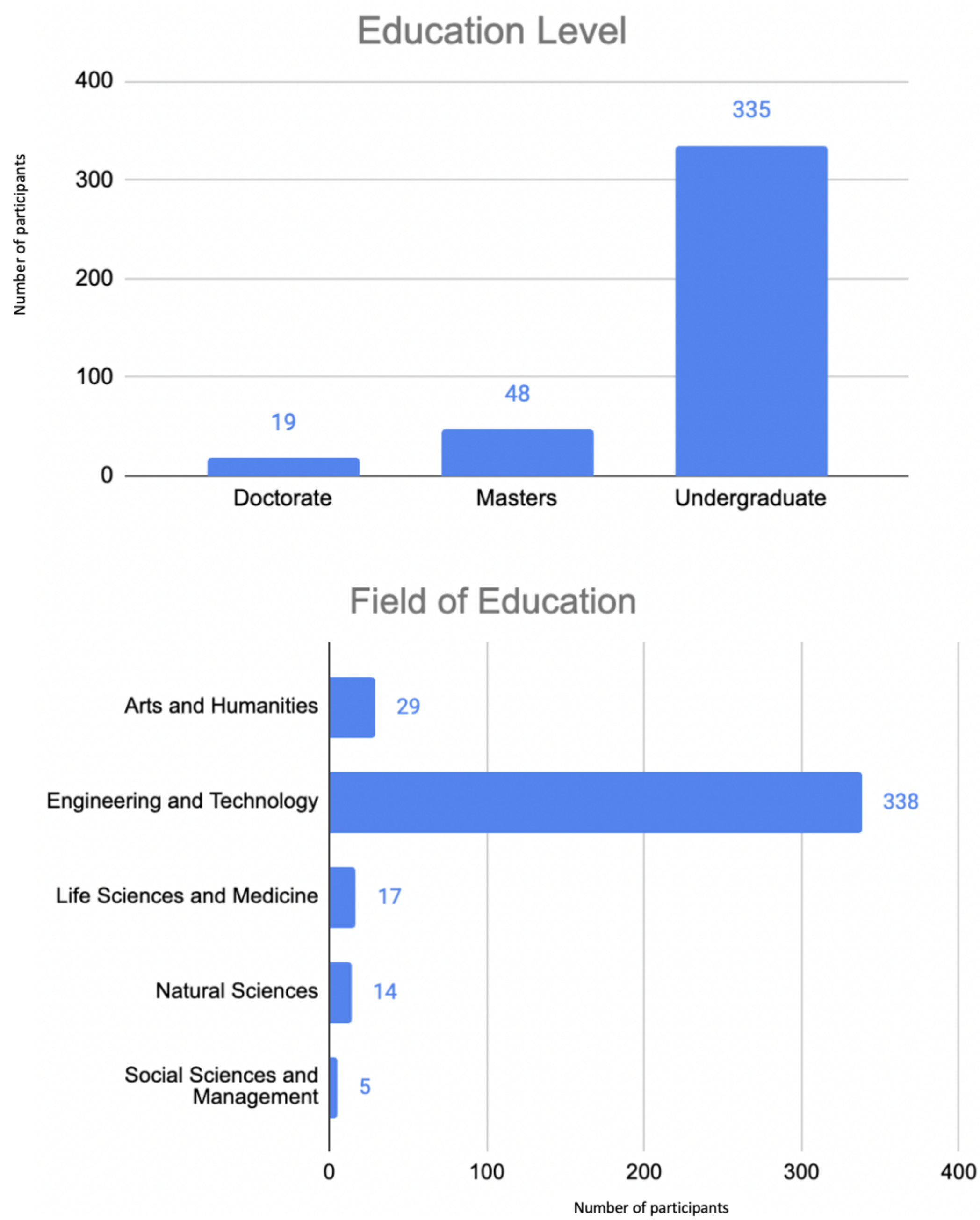


Figure 4.8: Distribution of education level and field of education

## 4.6 Sample Description

Our sample consisted mostly of young students with 50% below the age of 20. Participants were predominantly male (63% Male, 36% Female, $< 1\%$ Non-Binary), from STEM background (85 % STEM, 15% other) and educated (83% current undergraduate students, 12% current masters students). We also asked participants to rate their familiarity with HCI and haptic technology on a 4-point scale. In our collected data, 24% of participants reported being inexperienced with HCI, with 74% reporting moderately experienced and 2% reporting experienced. Similarly, 30% of participants reported being inexperienced with haptics, with 69% reporting moderately experienced and 1% reporting experienced.

## 4.7 Chapter Summary

In this section we discussed the in-person study design process, survey administration and data cleaning and splitting. We described the different demos opted for each haptic device and also elaborated the importance of inclusion of multiple haptic devices encapsulating different haptic feedback. Finally, we discussed the demographics of the collected data.

# Chapter 5

# Exploratory Factor Analysis

Exploratory factor analysis aims to explore the underlying theoretical structure obtained from the collected response [7]. It does so by clustering similar variables into similar factors to identify underlying latent constructs using the correlation matrix of items. In addition, it also reduces the number of items to a smaller subset to achieve a structure that withstands confirmatory factor and reliability checks. There are three main steps involved in exploratory factor analysis. These are:

- Assessment of the suitability of the data for factor analysis (section 5.1)
- Factor extraction which involves determining optimal number of factors, appropriate rotation type and refining number of items (section 5.2)
- Factor interpretation (section 5.2)

## 5.1 Assessment of the suitability of the data

Testing for sampling adequacy is an important step before diving into the factor analysis. At this stage it is important to determine if factors actually exist in the data or not. If the data is found to be inadequate for factors analysis, then perhaps factor analysis is not the right approach for the problem. We tested our sampling adequacy on four different criteria mentioned below.

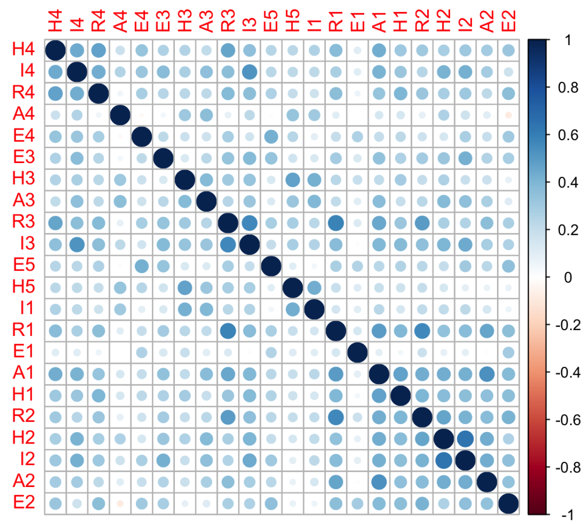


Figure 5.1: Correlation matrix for initial 22-items with N=291. The presence of clusters indicate the presence of factors.

### 5.1.1 Sample size

To determine the suitability of the data for factor analysis, sample size should be adequate [65, 67]. Although, a larger sample size is often recommended for exploratory factor analysis with at least 10 responses for each item, some researchers have found a smaller sample to be sufficient if the item loadings (magnitude of correlation between items and their associated factor) are significantly high. In our case, we had a sample size of 291 responses for 22 items which comes to approximately 13 responses per item. Therefore, our sample meets the sample size criteria as mentioned previously in section 4.2.

### 5.1.2 Evidence of correlation

The second criteria is to determine the strength of relationship among the items. In general there must be presence of correlation coefficients $> .30$ in a correlation matrix of all items [7]. The correlation matrix in Figure 5.1 show significant items with inter-item correlations indicated by the darker blue colors. More specifically, 105 items had inter-item correlation $> 0.30$. This means that there is evidence of correlation in our data which makes it suitable for factor analysis.

| H4 | I4 | R4 | A4 | E4 | E3 | H3 | A3 | R3 | I3 | E5 | H5 | I1 | R1 | E1 | A1 | H1 | R2 | H2 | I2 | A2 | E2 | Overall |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0.91 | 0.89 | 0.9 | 0.8 | 0.8 | 0.93 | 0.86 | 0.91 | 0.85 | 0.87 | 0.9 | 0.83 | 0.87 | 0.91 | 0.65 | 0.92 | 0.92 | 0.91 | 0.89 | 0.87 | 0.9 | 0.86 | 0.88 |

Figure 5.2: Kaiser-Meyer-Olkin factor adequacy test for the initial EFA. All values except E1 are above the recommended threshold of 0.70 [7] which suggests the appropriateness of data for EFA.

### 5.1.3 Kaiser-Meyer-Olkin (KMO)

The third criteria for sampling adequacy is Kaiser-Meyer-Olkin (KMO) test which is a metric to measure the adequacy of the data for factor analysis. It does so by calculating the proportion of variance among variables that might be common variance. In general, lower proportions of inter-item variance makes the data more suitable for Factor analysis. The KMO value ranges from 0 to 1 and is ranked such that a value from 0.8 to 1.0 is considered excellent, 0.7 to 0.79 is considered as adequate, 0.6 to 0.69 is considered as mediocre. KMO values less than 0.6 indicate that the sample is not appropriate for factor analysis [67, 30, 38]. Both item-level and cumulative KMO are required to be calculated for sampling adequacy.

The results of our KMO test are presented in the Figure 5.2. From the results we can observe, that our aggregate KMO score = .88 and all except E1 have KMO score > .70 which indicates that our sample is appropriate for factor analysis.

### 5.1.4 Bartlett's test of Sphericity

The fourth criteria we applied is the Bartlett's test of Sphericity [68]. This test compares a correlation matrix with an identity matrix (i.e. a matrix where each variable has inter-item correlation of 1.0). It is used to test the hypothesis that the variables are not orthogonal i.e. the variables are sufficiently correlated that the correlation matrix is significantly different from an identity matrix). More formally it tests the following research hypothesis:

Null Hypothesis ($H^o$): The variables are orthogonal

Research Hypothesis ($H^a$): The variables are not orthogonal

Based on our results ($\chi^2 = 2335, p < 0.05, df = 231$), we reject the null hypothesis and conclude that the variables are related which makes our sample adequate for factor analysis.

## 5.2 Extraction of Factors

After establishing sampling adequacy, the next step is to determine the optimal number of factors. In this step, we need to determine the optimal number of factors present in the underlying model, the appropriate rotation types followed by item refinement.

### 5.2.1 Ideal number of factors

A scree plot is a graphical tool used to plot the eigenvalues of the factors in decreasing order and is used to determine the optimal number of factors in scale development [7]. The magnitude of the eigenvalues is plotted on the y-axis while the eigenvalue numbers are represented on the x-axis. In general, the optimal number of components is at the point where the plot starts to level off. The scree plot for our data is represented in Figure 5.3. The plot starts to level off at around the 3rd component and again at the 5th component which indicates the presence of 3-5 factors in the 22 items. Though the HX model constitutes of five factors, we also iteratively experimented with 3-factor, 4-factor, and 5-factor models. We compared models based on their conceptual interpretability, number of items per factor and loading score. The 3-factor model had a high number of items per factor but were difficult to interpret due to items from different constructs. The 5-factor model had very low items per factor and low loading on their respective factors. From this iterative process, we arrived at a 4-factor model that had adequate number of items per factor and was easily interpreted.

### 5.2.2 Factor rotation

There are two input parameters for the factor analysis: factoring method and rotation type. If there is no multivariate normality in the data then 'principal axis factoring' is recommended [14]. We tested our data for multivariate normality with Mardia test. The results ($\chi_s^2 = 4861.42, p < 0.01; Z_k = 34.39, p < 0.01$ ) of both tests indicate that the data is non-normal and therefore, 'principal-axis' factoring is appropriate.

The next step is to select the appropriate rotation type. The goal of factor rotation is to clarify the factor structure and make the results of EFA more interpretable [22]. Factor rotation does not changes the fit or the total variance, but redistributes the variances across factors to enhance interpretation.

There are two rotation types in general: oblique and orthogonal. Oblique rotations are appropriate when the factors are expected to be correlated and orthogonal is used

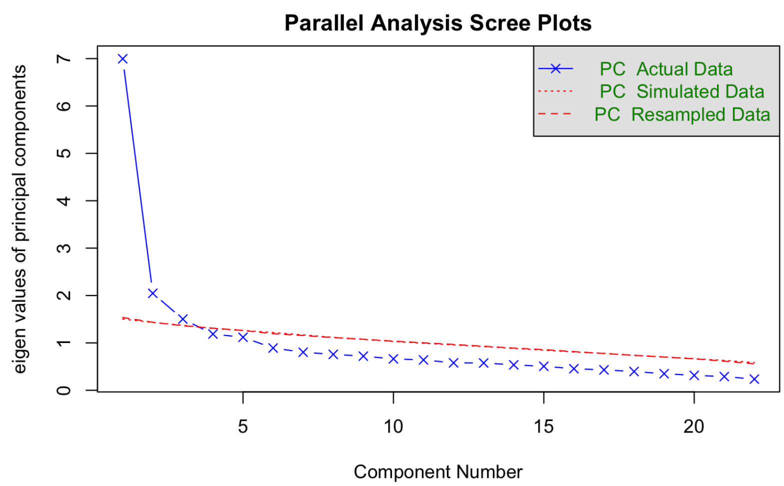


Figure 5.3: Scree plot with decreasing eigenvalues with respect to the number of possible factors. A steep drop at the 3rd component and a more subtle drop at the 5th component indicates that that there could be between 3-5 underlying factors in the 22-items

otherwise [27]. Although the theoretical HX model suggests that the factors are correlated but we still needed to confirm that. If the theoretical model does not provides insight about factor correlations then it is recommended to start with oblique rotation and calculating the factor correlations. If the majority of inter-factor correlations do not exceed 0.30 then orthogonal rotation is more suitable [67]. Therefore, we started with oblique rotation and found the inter-factor correlation score ranged from 0.3 to 0.5 which justifies the use of oblique rotation. Within oblique rotation, we experimented with 'promax', 'oblimin' and 'simplimax' and achieved the best results with 'promax' rotation.

### 5.2.3 Item reduction

After performing factor analysis, we refine our item set based on two inclusion criteria. Firstly, we considered an item if the magnitude of its loading was $> 0.40$. This ensures that the extracted item has a significant association with the underlying factor [49]. This resulted in removing items H4, I4, R4, E3 and E2. The items and their corresponding questions are presented in Table 3.1.

Secondly, we considered an item if it did not have significant cross loading i.e. the item did not load significantly on other factors. This step affirms that variances in items are uniquely associate to their associate factor only and not other factors. As a result items, I3, A1, H1, A2, A3, A4 were removed.

These 11 items were then subjected to a second analysis. The results of Kaiser-Meyer-Olkin (KMO) test (overall KMO = 0.77, none below 0.70) and Batlett's test ($\chi^2 = 884, p < 0.01, df = 55$) indicated adequate factorability of the remaining items. Additionally, the correlation plot as shown in Figure 5.5 indicated the presence of factors.

The inspection of the scree plot shown in Figure 5.6 suggested the presence of four factors and the results of Mardia test ($\chi_s^2 = 1293.27, p < 0.01; Z_k = 24.5, p < 0.01$) indicated that the remaining 11-item data is non-normal. Using principal axis factoring and Oblimin (Promax) rotation, we obtained a refined model as shown in Figure 5.4. All the items loaded significantly on their corresponding factors with no substantial cross-loading. The extracted items and their correlations are shown in Table 5.1.

## 5.3 Naming of factors

After refining our item pool, we obtained our four factor and 11-item model. We named each factor based on the items loaded on it and their interpretability. PA1 and PA4 were named as Realism and Expressivity respectively, as all the items loaded on them corresponded the same experiential dimension. PA2 was named as Harmony as the items as majority of the items corresponded to the dimension of Harmony. However, for PA3 we had to introduce a new dimension of Involvement. We discuss the detail interpretation of these factors in chapter 7.

## 5.4 Summary statistics

Next, we calculated some preliminary statistics for the obtained model using the collected data (N=291) to understand the distribution of scores. Firstly, we calculated the average and standard deviation for each of the 11-items which are shown in Table 5.2. Secondly, we calculated the average score with respect for each of the four extracted experiential dimension with respect to each of the haptic device used in the study shown in Table 5.3.

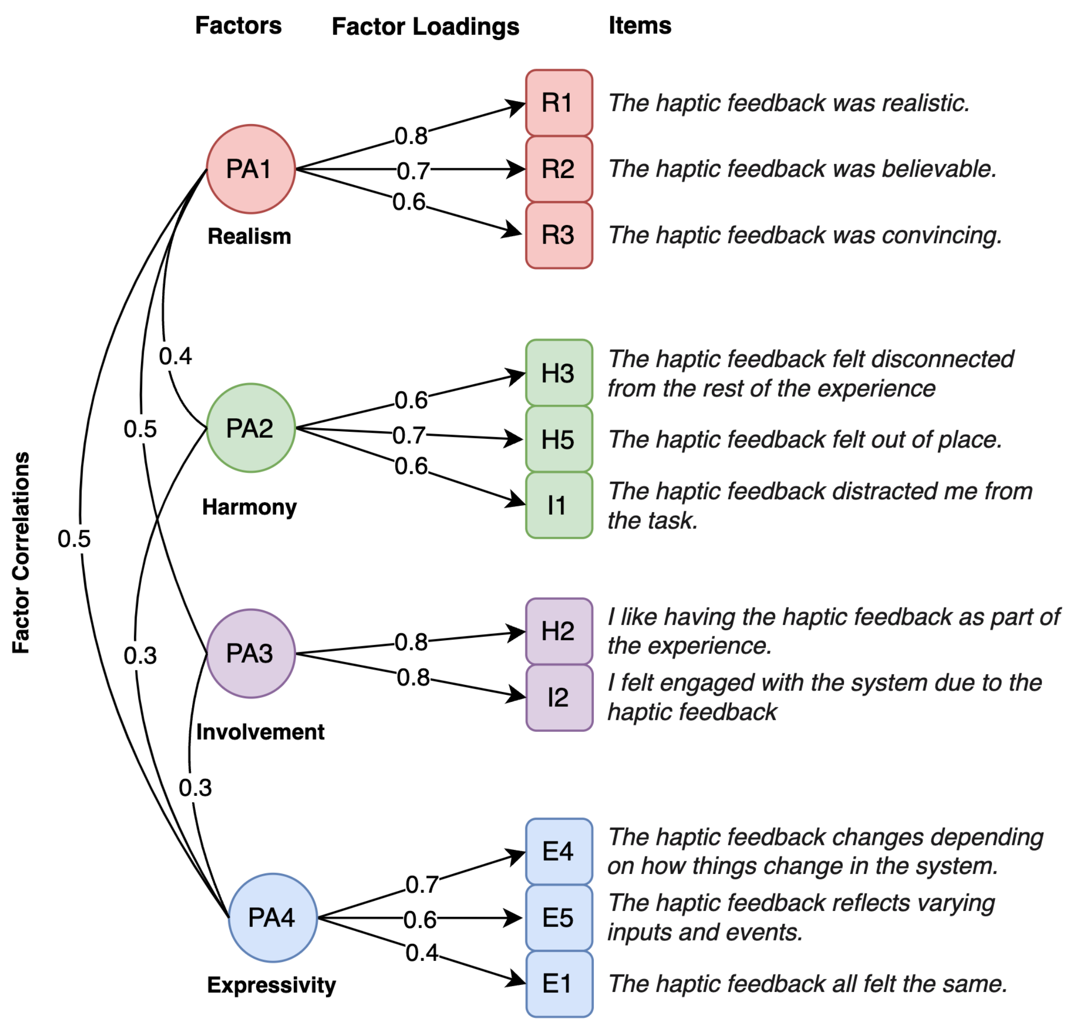


Figure 5.4: Path diagram for the best performing four factor model using promax rotation and principal axis factoring. This is different from the model extracted by Sathiyamurthy et al. as it does not contains the factors of autotelics and immersion. [58]

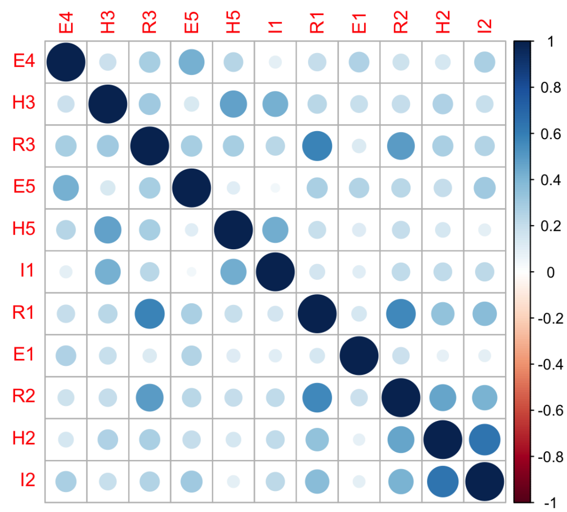


Figure 5.5: Correlation matrix for the refined 11 items with N=291. The presence of clusters indicate the presence of factors.

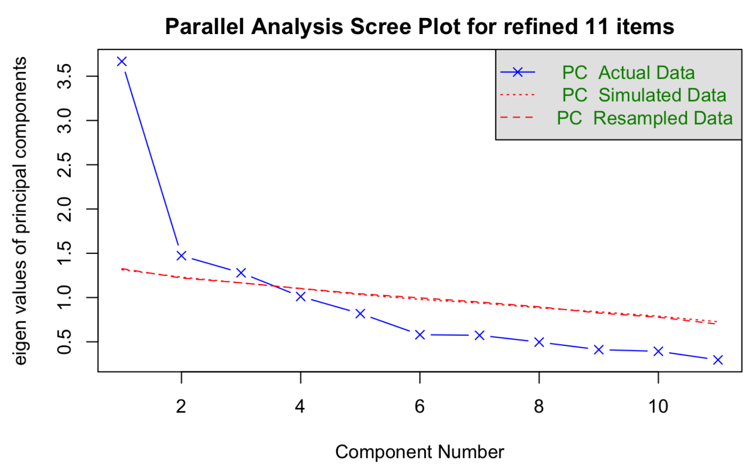


Figure 5.6: Scree plot for the refined 11 items with decreasing eigenvalues with respect to the number of possible factors. A steep drop at the fourth component indicates the presence of four factors.

| Item | Question | PA1 | PA2 | PA3 | PA4 |
|---|---|---|---|---|---|
| R1 | The haptic feedback was realistic | 0.84 | | | |
| R2 | The haptic feedback was believable | 0.63 | | | |
| R3 | The haptic feedback was convincing | 0.72 | | | |
| | | | | | |
| H3 | The haptic feedback felt disconnected from the rest of the experience | | 0.64 | | |
| H5 | The haptic feedback felt out of place | | 0.71 | | |
| I1 | The haptic feedback distracted me from the task | | 0.65 | | |
| | | | | | |
| H2 | I like having the haptic feedback as part of the experience | | | 0.76 | |
| I2 | I felt engaged with the system due to the haptic feedback | | | 0.81 | |
| | | | | | |
| E1 | The haptic feedback all felt the same | | | | 0.36 |
| E4 | The haptic feedback changes depending on how things change in the system | | | | 0.71 |
| E5 | The haptic feedback reflects varying inputs and events | | | | 0.63 |

Table 5.1: 11 items finalized after refinement alongside their loading score on their respective factors. PA1 = Realism, PA2 = Harmony, PA3 = Involvement and PA4 = Expressivity

| Device | Realism | Harmony | Involvement | Expressivity |
|---|---|---|---|---|
| 3D Touch | 12.097 | 11.597 | 9.194 | 12.153 |
| Haply | 12.20 | 11.836 | 8.382 | 12.382 |
| Oculus | 11.694 | 12.528 | 8.361 | 10.028 |
| Tanvas | 11.115 | 10.918 | 8.246 | 11.475 |
| Ultraleap | 11.284 | 11.358 | 8.776 | 11.134 |

Table 5.2: Average score for each haptic device with respect to each of the four experiential dimensions of the obtained model

| | R1 | R2 | R3 | H3 | H5 | I1 | H2 | I2 | E1 | E4 | E5 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Average | 3.801 | 3.979 | 3.893 | 3.838 | 3.818 | 3.904 | 4.337 | 4.306 | 3.261 | 4.144 | 4.151 |
| Std. dev | 0.911 | 0.851 | 0.862 | 0.949 | 0.988 | 0.978 | 0.837 | 0.809 | 1.169 | 0.719 | 0.678 |

Table 5.3: Summary statistics for the obtained 11-item model from the exploratory factor analysis (N=291)

| | REL | INV | EXP | HAR | Cronbach's $\alpha$ |
|---|---|---|---|---|---|
| REL | | | | | 0.79 |
| INV | 0.43 | | | | 0.71 |
| EXP | 0.48 | 0.35 | | | 0.53 |
| HAR | 0.53 | 0.28 | 0.32 | | 0.79 |

Table 5.4: Factor correlations and internal consistence (Cronbach's alpha) for 11 items from Study 1 (N=291) where REL = Realism, INV = Involvement, HAR = Harmony and EXP = Expressivity. Both overall $\alpha$ and majority of the subscale's $\alpha$ exceed the recommended threshold of 0.7 [46]

## 5.5 EFA Reliability

The next step after extraction of factors involve, measuring reliability of the model structure. Reliability is the degree of consistency exhibited when a measure is repeated under identical conditions [54]. There are different statistics designed to measure reliability such as Cronbach's alpha [23], test-retest reliability [55], Revelle's beta [56] and inter-observer reliability [55]. Among all these measures, Cronbach's alpha and test-retest reliability are most commonly used by researchers. Test-retest reliability, also known as the coefficient of stability, measures the consistency of participant's scores over time which requires re-administration of the questionnaire to the same participant at a different time. However, due to a large sample size and time constraints we opted for using Cronbach's alpha for checking our scale's reliability.

Cronbach's alpha measures the internal consistency (i.e, how closely related the items are as a group) of the scale. It is computed by calculating the correlation of each scale item with the total score for each observation then comparing it to the variance for all individual item scores. The value of Cronbach's alpha ranges from 0 to 1.

To confirm that our 11 extracted items maintained their internal reliability, we calculated Cronbach's alpha for the entire scale $\alpha = 0.782$ and for each sub-scale. The results are shown in the Table 5.4. A multi-component scale is considered reliable if overall $\alpha$ exceeds 0.6 and a majority if sub-scale $\alpha$s exceed 0.7 [46]. Based on the results, our scale meets this reliability threshold as $\alpha$ for three (REL, INV, HAR) of the subscales is above the recommended threshold of 0.7 with only EXP having a lower $\alpha$.

## 5.6 Chapter Summary

In this section, we analyzed and established the adequacy of our collected sample based on four criteria (sample size, evidence of correlation, KMO and Barlett's test). We determined the ideal number of underlying factors using scree plot, appropriate factor rotation type and extraction method. Following that, we removed items that cross loaded or did not loaded and obtained a 4-factor model with 11 items using Promax rotation and principal Axis factoring. Finally, we established the reliability of our model using Cronbach's alpha.

# Chapter 6

# Confirmatory Factor Analysis

Next, we verified the structure of the extracted model using Confirmatory Factor Analysis (CFA) with our testing set (N=112). While Exploratory Factor Analysis (EFA) is used to extract the underlying factor structure, it does not measure the model's goodness-of-fit with respect to new data [64]. Hence, to verify the model structure, CFA is used. CFA is a dimensionality test in which the hypothesized model obtained from EFA is tested at a different time point in a longitudinal study or a new sample.

## 6.1 Goodness of fit

Our hypothesized model demonstrated adequate goodness-of-fit with its chi-square ($\chi^2 = 47.07$) below the conservative limit of double the degrees of freedom (DoF = 38) [16]. In addition, using ANOVA, our theoretical model demonstrated better fit than the null model ($\chi^2 = 371.42, p < 0.001$).

We also calculated several other goodness-of-fit metrics. First, we determined the Comparative Fit Index (CFI), which measures the model's fit relative to a more restrictive baseline model [4], and the Tucker-Lewis Index (TLI), a more conservative version of CFI, penalizing overly complex models [5]. Our model performed well in both (CFI = 0.971, TLI = 0.959) with scores over the recommended threshold of 0.90 [49].

Afterwards, we calculated Standardized Root Mean Square Residual (SRMR) which measures the mean absolute difference between observed and predicted correlations [67]. The acceptable threshold for acceptable model fit is $\leq 0.08$ [67]. Our model had SRMR score of 0.070 which is withing the acceptable threshold so it achieves an acceptable fit.

|  | REL | INV | EXP | HAR | Cronbach's $\alpha$ |
|---|---|---|---|---|---|
| REL |  |  |  |  | 0.79 |
| INV | 0.43 |  |  |  | 0.76 |
| EXP | 0.48 | 0.35 |  |  | 0.53 |
| HAR | 0.53 | 0.28 | 0.32 |  | 0.59 |

Table 6.1: Factor correlations and internal consistency (Cronbach's alpha) for 11 items from Study-1 (N=112) where REL=Realism, INV = Involvement, HAR = Harmony and EXP = Expressivity. Note that here, in contrast to the EFA findings, majority of the subscale's Cronbach's $\alpha$ is below the recommended threshold of 0.70 [7]

Next, we calculated the Root Mean Square Error of Approximation (RMSEA), which measures how well the model produces item covariances, instead of a baseline model comparison [71]. To interpret RMSEA, Cudeck et al. recommended RMSEA $\leq 0.05$ as a close fit, $0.05 \leq$ RMSEA $\leq 0.08$ as fair fit, RMSEA $> 0.10$ as indicative of a poor fit between the hypothesized model and the observed data [24]. Our model's RMSEA (0.046) could be interpreted as a 'close fit'.

## 6.2 Reliability

To further confirm the internal consistency of our model, we recalculated the Cronbach's $\alpha$ for the second study. The results are represented in Table 6.1. The overall Cronbach $\alpha$ exceeded the recommended threshold of 0.70 [23, 46]. However, two of the sub-scales (Expressivity and Harmony) were less than the recommended threshold of 0.7. For a scale to be reliable and robust, both the overall Cronbach's $\alpha$ and majority (more than 50%) of the subscale's Cronbach's $\alpha$ should be above the minimum recommended threshold of 0.7. Therefore, this is an indication of weak reliability (internal consistency).

## 6.3 Chapter Summary

In this section we analyzed the reproducibility of our 4-factor model by conducting Confirmatory Factor Analysis (CFA) on an independent sample (N=112). Our model met several goodness-of-fit measures such as RMESEA, SRMR, TLI and CFI. Next, we analyzed the model's reliability using Cronbach's alpha. While the overall model's $\alpha$ was above the minimum acceptable threshold, two subscales (EXP and HAR) could not surpass this threshold. Therefore, we believe another iteration of CFA is needed to confirm the model structure.

# Chapter 7

# Discussion

In this section, we discuss the interpretation of the four extracted experiential factors, including the practical considerations for haptic designers and researchers for building new evaluative instruments. We also distinguish our model from the prior vibrotactile model extracted by Sathiyamurthy el al. [58] Finally, we discuss the limitations and future work.

## 7.1 Interpretation of Factors

Our findings refine existing understanding of hedonic factors underlying HX by showing which items load, and how it compares to prior investigations with vibrotactile feedback in a remote study [58]. Two of our factors (Realism, Expressivity) closely match the intended construct and the measured outcome with vibrotactile feedback. One of our factors (Harmony) seems to align with an existing construct from the HX model. The final factor (Involvement), with only two loading items, is one we had to introduce to explain our results. Our final 4-factor and 11-item questionnaire is presented in Table 7.1.

### 7.1.1 Realism

Realism is the strongest factor in the model with items R1, R2 and R3, consistent with the original proposal. Realism is defined as whether the haptic effects convincingly exhibit what someone expects to feel in reality [39]. However, the variable R4, "The haptic feedback matched my expectation" did not load significantly. This is consistent with the results obtained by Sathiyamurthy et al. [58]. This could be due to the fact that R4 is expecting

| Item | Question | Factor |
|---|---|---|
| R1 | The haptic feedback was realistic | REL |
| R2 | The haptic feedback was believable | REL |
| R3 | The haptic feedback was convincing | REL |
| | | |
| H3 | The haptic feedback felt disconnected from the rest of the experience | HAR |
| H5 | The haptic feedback felt out of place | HAR |
| I1 | The haptic feedback distracted me from the task | HAR |
| | | |
| H2 | I like having the haptic feedback as part of the experience | INV |
| I2 | I felt engaged with the system due to the haptic feedback | INV |
| | | |
| E1 | The haptic feedback all felt the same | EXP |
| E4 | The haptic feedback changes depending on how things change in the system | EXP |
| E5 | The haptic feedback reflects varying inputs and events | EXP |

Table 7.1: Final 4-factor and 11-item questionnaire, where REL = Realism, HAR = Harmony, INV = Involvement and EXP = Expressivity

the user to have some prerequisite expectations about the haptic feedback which might be perplexing the user.

### 7.1.2 Involvement

Involvement is the only newly-defined factor that emerged from our investigation. Although the items (H2 and I2) in this factor correspond to different intended constructs (Harmony and Immersion) from the hypothesized model, taken together they could be interpreted as measuring the user's "Involvement." Involvement is 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 [72]. Involvement depends on the degree of significance or meaning that the individual attaches to the stimuli, activities, or events. The fact that both item statements start with "I" suggests that it is focused on the subjective experience. In this case, we suppose that Involvement, as measured by these two items (H2 and I2), could represent engagement with the system due to meaningful haptics.

Involvement as a construct has not yet been included in the discourse surrounding HX, and we note that this is the only resulting factor with only two items. As such, this factor will need to bear the most scrutiny in future work, as it may be underspecified.

### 7.1.3 Harmony

This factor has two items related to Harmony (H3, H5) and one item corresponding to Immersion (I1). The theoretical HX model suggests that Immersion could also be produced by attaining Harmony if desired for the system [39]. Therefore, I1 could also be interpreted as explaining Harmony as well. The items in this factor are all negatively-phrased and could be interpreted as "disruption" which is opposite to the Harmony present in our theoretical model. However, these items are positively correlated with the underlying factor because all the negatively-phrased items were reverse-coded at the data cleaning stage. This means that after reverse-coding, low scores on negative items were changed to higher ones leading to a positive correlation. Therefore, this factor could be interpreted as measuring the absence of disruption or Harmony in the system. Items H1 "The haptic feedback fits well with the other senses" and H4 "The haptic feedback felt appropriate when and where I felt it" did not load onto this factor. One reason could be that these items are inclined towards measuring the Hygiene factors which when not present causes user dissatisfaction.

### 7.1.4 Expressivity

Factor 2 consists of items E1, E4 and E5 where E1 is a negatively-phrased item (reversed-coded in data cleaning stage). Expressivity has been defined as such that it allows users to feel the haptics distinguishably reflect varying user input and system events [39]. This construct is captured by the three items. E4 and E4 measures the extent to which a user's different interactions with the system result in different forms of haptic feedback. E1 being a negatively-phrased item measures the opposite of that and after reverse-coding it could be interpreted as the haptic feedback was not constant. The fact that E2 and E3 didn't load significantly may indicate the expressivity here demands correspondence with inputs and other parts of the system, rather than just variations.

## 7.2 Differences with the vibrotactile model

Our extracted model differs from the model extracted by Sathiyamurthy et al. [58] in a similar study. Our model has only four factors with the absence of Immersion as an independent factor and has lower items per factor. This could be due to the fact that for developing the previous version of HX, recognizable commercially available haptic devices (smartphone and gaming console) were used. Moreover, the study was restricted to vibrotactile feedback only. However, for developing the current version of HX, non-commercial devices (Tanvas, Ultraleap, 3D Touch and Haply) with different forms of haptic feedback (mid-air, vibrotactile, force) were used. This raises the question of generalizability versus specificity as the previous version was limited to vibrotactile whereas the current version encompassed various haptic feedbacks. Therefore, it could be inferred that a generalizable scale could have different constructs as compared to one developed for a specific type of haptic feedback.

Another difference is that no item related to Autotelics was included in the final extracted model. Autotelics is a purely hedonic factor measuring the experience of touch in and of itself [51]. For instance, the haptic feedback might contribute to the experience but without any significant impact on the pragmatic goals i.e. task completion. The absence of Autotelics related items, suggests that this version of the HX model focuses more on the pragmatic contributors of overall haptic experience.

## 7.3 Expertise in training

The HX model was developed with input from novices and expert hapticians [39]. Novices typically had experience from vibrotactile devices such as smartphones or consoles. Although novices were told about other types of devices and haptic feedback, they did not have the option to try non-vibrotactile devices during the model construction. Kim and Schneider [39] had to rely on expert input to provide generality to the model. As such, it could be that the HX model includes constructs that are meaningful to expert hapticians generally, but non-hapticians only for vibrotactile feedback.

The result could be that the 5-factor HX model can be used to elicit feedback on typical vibrotactile devices such as smartphones and game controllers, and possibly be used by trained experts to evaluate other types of devices. However, the 5-factor model might be inappropriate for end-user evaluation of more varied systems, precisely the type of evaluation used in our study: people without haptics training evaluating varied haptic feedback.

Unfortunately, we did not have sample size to conduct factor analysis of each modality separately to see if we could replicate the results of Sathiyamurthy et al. [58], as we only had one vibrotactile device (Oculus Quest 2) in our sample with 41 responses. In contrast, Sathiyamurthy et al. used a sample of 300 to obtain the five factor structure using vibrotactile devices only. Thus, we would be interested in conducting a study involving vibrotactile devices only and compare the results with those obtained by Sathiyamurthy et al. to check the model consistency.

## 7.4 Hygienes and Motivators

In our resulting model, we found that items (H2, H3, H5, I1, I2) intended for Harmony and Immersion were intertwined into two other factors: our resulting Harmony (H3, H5, I1) and Involvement (H2, I2). In the theoretical HX model, Harmony and Immersion are highly related: poor Harmony would almost certainly break Immersion, while Immersion could also be produced by attaining Harmony if desired for the system [39].

We wonder if we have found an alternative structure, one that is potentially stronger or more relevant than directly measuring Harmony and Involvement. Perhaps the 4-factor model separately measures negative factors (H3, H5, I1) that disrupt the experience and lead to poor Harmony and Immersion, and positive factors (H2, I2) that contribute to a meaningful, engaging experience.

This could possibly relate to the notions of *hygiene* and *motivators*, as adapted to UX by Tuch et al. [70]. Hygiene factors only affect a user's experience if they are below an acceptable threshold. For example, users might not appreciate the perfect functioning of their smartphones, but as soon as the smartphone does not work, they would experience the interaction as negative. Therefore, hygiene factors are important to avoid negative experiences but do not necessarily create positive experiences. On the other hand, motivators contribute to positive user experiences but their absence does not leads to negative experiences [70]. In other words, our resulting structure might measure disruptive features (hygiene) that negatively impact experience, and then positive features (motivators) that make it is a good experience.

## 7.5 Practical considerations for a cross-device scale

Measuring a user's haptic experience, distinguishing it from the overall user experience and generalizing it across all haptic devices using a scale is challenging. However, we believe using a scale compared to qualitative feedback is more accessible, affordable and efficient.

While the theoretical HX scale was intended to be used across different types of haptic devices, the refined model obtained from this study incorporating different types of haptic feedbacks (vibrotactile, mid-air, force feedback) is different from a similar study done by Sathiyamurthy et al. [58] using only vibrotactile feedback. This raises the question of generalizability versus specificity in terms of evaluating the haptic experience. In practical terms, this means that the five item scale designed by Suji et al. could be useful for a designer interested in evaluating the vibrotactile feedback of their device. However, for evaluating the general haptic feedback, our four-item (Realism, Harmony, Involvement, Expressivity) scale is more suitable. In broader terms, we believe that different evaluation scales might be appropriate for measuring the haptic experience of specific haptic modalities (vibrotactile, force, mid-air). Therefore, this scale can serve as the basis for future generalizable scales. However, for more nuanced and modality specific evaluation of haptic experience, we recommend the development of different scales.

## 7.6 Limitations

While this study gives evidence of a structure for a HX model, it is not without limitations. These limitations provide us with insights that could be utilized in the construction of future evaluative instruments.

### 7.6.1 Reliability not achieved

The most significant limitation is that reliability was not achieved during the Confirmatory Factor Analysis stage. Reliability in the context of scale development is important to assess the reproducibility of the extracted model and is measured through Cronbach's $\alpha$. While the overall $\alpha$ was adequate, two sub-scales had lower $\alpha$s which is why the extracted model could not be considered reliable with great confidence. Therefore, we do not advocate the use of this version of the scale in user studies.

### 7.6.2 Novelty Effect

The devices used in the study were non-commercial and not easily accessible to respondents. In addition, the haptic feedback provided by these devices was unique and quite different from the more readily accessible vibrotactile haptic feedback. Therefore, there might be a novelty effect in our data which might have led participants to respond more positively. Koch et al. has defined novelty effect as "an increased motivation to use something, or an increase in the perceived usability of something, on account of its newness. When novelty eventually fades, usage patterns and/or perceived usability changes" [41]. In addition, the work of Rutten et al. involving mid-air haptic feedback has shown the existence of novelty effect in UX research studies which results in increased attractiveness towards the device for participants [57]. Consequently this results in participants responding more positively in subsequent questionnaires.

### 7.6.3 Difficulty in recruitment

We found in-person recruitment with such a large sample size to be quite challenging. For the ease of participants, we had opted to use Qualtrics for the entire survey and allowed people to use the researcher provided laptop to answer the survey or their own phones for their convenience. No identification information on Qualtrics or paper was included in the study to preserve anonymity of the participant and minimize the survey completion time. While we prevented multiple response submissions from the same participant using Qualtrics, it is possible that a participant could have participated twice using a different device to answer the survey.

### 7.6.4 Social Desirability bias

The presence of researchers on the study site along with the respondent might have led to social desirability bias. Social desirability bias occurs when a respondent opts for responses that they consider more socially acceptable rather than choosing responses that are a reflection of their true responses [28]. This leads to over-reporting of socially desirable responses and under-reporting of less socially desirable responses. Since, our survey included negatively-phrased questions, we believe that it might be prone to social desirability bias. To mitigate this, we recommend that future studies must be conducted in a controlled laboratory setting without the presence of the researcher.

### 7.6.5 Less diversity in sample

Although we had sought to recruit people from diverse backgrounds, it is quite possible that our sample does not fully represent the general population. This is because the study was conducted entirely at the University of Waterloo involving enrolled students which might have limited the generalizability of the scale.

### 7.6.6 Small number of devices

Finally, we believe that owing to the rapidly evolving technology landscape, new and intricate haptic devices are being developed and released commercially. While we designed our scale to encompass different haptic feedbacks (vibrotactile, mid-air, force), our study was restricted by the number of devices (Oculus, Tanvas, 3D Touch, Haply, Ultraleap) used which could have led to lower generalizability.

## 7.7 Future Work

The HX model obtained from this study was confirmed using Confirmatory Factor Analysis (CFA). Although the model structure remained intact in CFA, it was not able to meet the reliability threshold. In future, we would like to explore the validity of the obtained model dimensions (Harmony, Expressivity, Involvement, Realism) using convergent and divergent validity studies. Subsequently, these dimensions alongside new items could abet the development of future versions of HX.

## 7.8 Chapter Summary

In this section, we discussed the interpretation of each of the extracted factors i.e. Realism, Expressivity, Harmony and Involvement. Next, we distinguished the obtained model from the previously obtained vibrotactile model obtained by Sathiyamurthy et al. [58]. Afterwards, we provided practical considerations for haptic designers to use the obtained model in their designs and insights for future studies. Finally, we discussed the limitations of the structure we found.

# Chapter 8

# Conclusion

In this paper, we present the development of a multidimensional scale using the experiential dimensions of HX developed by Kim and Schneider [39] and items generated by Sathiyamurthy et al [58]. Through an extensive literature review we distinguish haptic experience from traditional user experience and emphasize the importance of measuring haptic experience on a standalone basis. Next from analysis of current evaluation methods of haptic experience, we stress the need of a quantitative and scalable instrument to measure haptic experience. Through two in-person user studies incorporating diverse haptic modalities and involving 430 participants we extracted and evaluated a 4-factor model that is usable with different haptic modalities. These four factors consist of Realism, Harmony, Involvement and Expressivity.

Using our 4-factor model, hapticians can obtain insights into their designs regardless of the type of haptic feedback. These insights can help them in distinguishing and comparing the parameters of a good haptic experience for the end user. The obtained model also has theoretical and practical implications for haptics research. Researchers can use our obtained model for improved understanding of haptic experience and for the development of future evaluative instruments for measuring haptic experience.

# References

[1] Vero Vanden Abeele, Katta Spiel, Lennart Nacke, Daniel Johnson, and Kathrin Gerling. Development and validation of the player experience inventory: A scale to measure player experiences at the level of functional and psychosocial consequences. *International Journal of Human-Computer Studies*, 135:102370, 2020.

[2] Mahdi Azmandian, Mark Hancock, Hrvoje Benko, Eyal Ofek, and Andrew D Wilson. Haptic retargeting: Dynamic repurposing of passive haptics for enhanced virtual reality experiences. In *Proceedings of the 2016 chi conference on human factors in computing systems*, pages 1968–1979, 2016.

[3] Juergen Baumgartner, Naomi Frei, Mascha Kleinke, Juergen Sauer, and Andreas Sonderegger. Pictorial system usability scale (p-sus) developing an instrument for measuring perceived usability. In *Proceedings of the 2019 chi conference on human factors in computing systems*, pages 1–11, 2019.

[4] Peter M Bentler. Comparative fit indexes in structural models. *Psychological bulletin*, 107(2):238, 1990.

[5] Peter M Bentler and Douglas G Bonett. Significance tests and goodness of fit in the analysis of covariance structures. *Psychological bulletin*, 88(3):588, 1980.

[6] Marit Bentvelzen, Jasmin Niess, Mikołaj P Woźniak, and Paweł W Woźniak. The development and validation of the technology-supported reflection inventory. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, pages 1–8, 2021.

[7] Godfred O Boateng, Torsten B Neilands, Edward A Frongillo, Hugo R Melgar-Quiñonez, and Sera L Young. Best practices for developing and validating scales for health, social, and behavioral research: a primer. *Frontiers in public health*, 6:149, 2018.

[8] Kenneth A Bollen. *Structural equations with latent variables*, volume 210. John Wiley & Sons, 1989.

[9] Barbara Boos and Henning Brau. Erweiterung des ueq um die dimensionen akustik und haptik. In Steffen Hess and Holger Fischer, editors, *Mensch und Computer 2017 - Usability Professionals*, Regensburg, 2017. Gesellschaft für Informatik e.V.

[10] Richard W Brislin. Cross-cultural research methods. In *Environment and culture*, pages 47–82. Springer, 1980.

[11] Jeanne H Brockmyer, Christine M Fox, Kathleen A Curtiss, Evan McBroom, Kimberly M Burkhart, and Jacquelyn N Pidruzny. The development of the game engagement questionnaire: A measure of engagement in video game-playing. *Journal of experimental social psychology*, 45(4):624–634, 2009.

[12] Timothy A Brown. *Confirmatory factor analysis for applied research.* Guilford publications, 2015.

[13] Timothy A Brown and Michael T Moore. Confirmatory factor analysis. *Handbook of structural equation modeling*, 361:379, 2012.

[14] Florian Brühlmann, Beat Vollenwyder, Klaus Opwis, and Elisa D Mekler. Measuring the "why" of interaction: Development and validation of the user motivation inventory (umi). In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, pages 1–13, 2018.

[15] M Burmester, M Hassenzahl, and F Koller. Attrakdiff: Ein fragebogen zur messung wahrgenommener hedonischer und pragmatischer qualität. In *Mensch & Computer*, volume 57, page 2003, 2003.

[16] Edward G Carmines. Analyzing models with unobserved variables. *Social measurement: Current issues*, 80, 1981.

[17] Raymond B Cattell. *Factor analysis: an introduction and manual for the psychologist and social scientist.* Harper, 1952.

[18] Barbara A Cerny and Henry F Kaiser. A study of a measure of sampling adequacy for factor-analytic correlation matrices. *Multivariate behavioral research*, 12(1):43–47, 1977.

[19] Lee Anna Clark and David Watson. *Constructing validity: Basic issues in objective scale development.* American Psychological Association, 2016.

[20] Andrew L Comrey. Common methodological problems in factor analytic studies. *Journal of consulting and clinical psychology*, 46(4):648, 1978.

[21] Andrew L Comrey and Howard B Lee. *A first course in factor analysis*. Psychology press, 2013.

[22] Anna B Costello and Jason Osborne. Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis. *Practical assessment, research, and evaluation*, 10(1):7, 2005.

[23] Lee J Cronbach. Coefficient alpha and the internal structure of tests. *psychometrika*, 16(3):297–334, 1951.

[24] Robert Cudeck. of assessing model fit. *Testing structural equation models*, 154:136, 1993.

[25] Jurgen A Doornik and Henrik Hansen. An omnibus test for univariate and multivariate normality. *Oxford bulletin of economics and statistics*, 70:927–939, 2008.

[26] Michel Goossens, Frank Mittelbach, and Alexander Samarin. *The LaTeX Companion*. Addison-Wesley, Reading, Massachusetts, 1994.

[27] Richard L Gorsuch. *Factor analysis: Classic edition*. Routledge, 2014.

[28] Pamela Grimm. Social desirability bias. *Wiley international encyclopedia of marketing*, 2010.

[29] Edward Guadagnoli and Wayne F Velicer. Relation of sample size to the stability of component patterns. *Psychological bulletin*, 103(2):265, 1988.

[30] Louis Guttman. Some necessary conditions for common-factor analysis. *Psychometrika*, 19(2):149–161, 1954.

[31] Joseph F Hair, William C Black, Barry J Babin, Rolph E Anderson, and Ronald L Tatham. Multivariate data analysis . uppersaddle river. *Multivariate Data Analysis (5th ed) Upper Saddle River*, 5(3):207–219, 1998.

[32] Joseph F Hair, William C Black, Barry J Babin, Rolph E Anderson, and Ronald L Tatham. others.(2006). *Multivariate data analysis*, 6, 2006.

[33] Yoel Haitovsky. Multicollinearity in regression analysis: Comment. *The Review of economics and statistics*, pages 486–489, 1969.

[34] Marc Hassenzahl. The hedonic/pragmatic model of user experience. *Towards a UX manifesto*, 10:2007, 2007.

[35] Marc Hassenzahl, Michael Burmester, and Franz Koller. Attrakdiff: A questionnaire to measure perceived hedonic and pragmatic qualityät. In *human & computer 2003*, pages 187–196. knight, 2003.

[36] Jeffrey Heer and Michael Bostock. Crowdsourcing graphical perception: using mechanical turk to assess visualization design. In *Proceedings of the SIGCHI conference on human factors in computing systems*, pages 203–212, 2010.

[37] Frederick Herzberg. The motivation-hygiene theory. *Organizational behavior one: Essential theories of motivation and leadership, eds JB Miner, ME Sharpe Inc, New York*, 2(4):61–74, 2005.

[38] Henry F Kaiser. *A second generation little jiffy.* Springer, 1970.

[39] Erin Kim and Oliver Schneider. Defining haptic experience: Foundations for understanding, communicating, and evaluating hx. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, pages 1–13, 2020.

[40] Donald Knuth. *The TEXbook.* Addison-Wesley, Reading, Massachusetts, 1986.

[41] Michael Koch, Kai von Luck, Jan Schwarzer, and Susanne Draheim. The novelty effect in large display deployments–experiences and lessons-learned for evaluating prototypes. In *Proceedings of 16th European conference on computer-supported cooperative work-exploratory papers*. European Society for Socially Embedded Technologies (EUSSET), 2018.

[42] Leslie Lamport. *LATEX — A Document Preparation System.* Addison-Wesley, Reading, Massachusetts, second edition, 1994.

[43] Bettina Laugwitz, Theo Held, and Martin Schrepp. Construction and evaluation of a user experience questionnaire. In *Symposium of the Austrian HCI and usability engineering group*, pages 63–76. Springer, 2008.

[44] Effie L-C Law, Florian Brühlmann, and Elisa D Mekler. Systematic review and validation of the game experience questionnaire (geq)-implications for citation and reporting practice. In *Proceedings of the 2018 Annual Symposium on Computer-Human Interaction in Play*, pages 257–270, 2018.

[45] Emanuela Maggioni, Erika Agostinelli, and Marianna Obrist. Measuring the added value of haptic feedback. In *2017 ninth international conference on quality of multimedia experience (QoMEX)*, pages 1–6. IEEE, 2017.

[46] Robert K McKinley, Terjinder Manku-Scott, Adrian M Hastings, David P French, and Richard Baker. Reliability and validity of a new measure of patient satisfaction with out of hours primary medical care in the united kingdom: development of a patient questionnaire. *Bmj*, 314(7075):193, 1997.

[47] Michael Minge, Manfred Thüring, and Ingmar Wagner. Developing and validating an english version of the mecue questionnaire for measuring user experience. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, volume 60, pages 2063–2067. SAGE Publications Sage CA: Los Angeles, CA, 2016.

[48] Thomas Muender, Michael Bonfert, Anke Verena Reinschluessel, Rainer Malaka, and Tanja Döring. Haptic fidelity framework: Defining the factors of realistic haptic feedback for virtual reality. In *CHI Conference on Human Factors in Computing Systems*, pages 1–17, 2022.

[49] Richard G Netemeyer, William O Bearden, and Subhash Sharma. *Scaling procedures: Issues and applications.* sage publications, 2003.

[50] Jum C Nunnally. *Psychometric Theory, McGraw-Hill Book Company, New York, NY.* McGraw-Hill, 1978.

[51] Marianna Obrist, Sriram Subramanian, Elia Gatti, Benjamin Long, and Thomas Carter. Emotions mediated through mid-air haptics. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, pages 2053–2062, 2015.

[52] Horea Pauna, Pierre-majorique Léger, Sylvain Sénécal, Marc Fredette, François Courtemanche, Shang-Lin Chen, Élise Labonté-Lemoyne, and Jean-François Ménard. The psychophysiological effect of a vibro-kinetic movie experience: The case of the d-box movie seat. In *Information systems and neuroscience*, pages 1–7. Springer, 2018.

[53] Joann Peck and Terry L Childers. Individual differences in haptic information processing: The “need for touch” scale. *Journal of Consumer Research*, 30(3):430–442, 2003.

[54] Miquel Porta. *A dictionary of epidemiology.* Oxford university press, 2008.

[55] Tenko Raykov and George A Marcoulides. *Introduction to psychometric theory*. Routledge, 2011.

[56] William Revelle. Hierarchical cluster analysis and the internal structure of tests. *Multivariate Behavioral Research*, 14(1):57–74, 1979.

[57] Isa Rutten, Lawrence Van den Bogaert, and David Geerts. From initial encounter with mid-air haptic feedback to repeated use: the role of the novelty effect in user experience. *IEEE Transactions on Haptics*, 14(3):591–602, 2020.

[58] Suji Sathiyamurthy, Melody Lui, Erin Kim, and Oliver Schneider. Measuring haptic experience: Elaborating the hx model with scale development. In *2021 IEEE World Haptics Conference (WHC)*, pages 979–984. IEEE, 2021.

[59] Oliver Schneider, Karon MacLean, Colin Swindells, and Kellogg Booth. Haptic experience design: What hapticians do and where they need help. *International Journal of Human-Computer Studies*, 107:5–21, 2017.

[60] Oliver S Schneider, Hasti Seifi, Salma Kashani, Matthew Chun, and Karon E MacLean. Hapturk: crowdsourcing affective ratings of vibrotactile icons. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, pages 3248–3260, 2016.

[61] Martin Schrepp, Andreas Hinderks, and Jörg Thomaschewski. Applying the user experience questionnaire (ueq) in different evaluation scenarios. In *International Conference of Design, User Experience, and Usability*, pages 383–392. Springer, 2014.

[62] Noora Shrestha. Detecting multicollinearity in regression analysis. *American Journal of Applied Mathematics and Statistics*, 8(2):39–42, 2020.

[63] Noora Shrestha. Factor analysis as a tool for survey analysis. *American Journal of Applied Mathematics and Statistics*, 9(1):4–11, 2021.

[64] Si-Mui Sim and Raja Isaiah Rasiah. Relationship between item difficulty and discrimination indices in true/false-type multiple choice questions of a para-clinical multidisciplinary paper. *Annals-Academy of Medicine Singapore*, 35(2):67, 2006.

[65] J Stevens. Categorical data: The log linear model. *Applied multivariate statistics for the social sciences (3rd ed., pp. 518-557). Mahwah, NJ: Lawrence Erlbaum Associates*, 1996.

[66] Hyewon Suh, Nina Shahriaree, Eric B Hekler, and Julie A Kientz. Developing and validating the user burden scale: A tool for assessing user burden in computing systems. In *Proceedings of the 2016 CHI conference on human factors in computing systems*, pages 3988–3999, 2016.

[67] Barbara G Tabachnick, Linda S Fidell, and Jodie B Ullman. *Using multivariate statistics*, volume 5. pearson Boston, MA, 2007.

[68] Sigmund Tobias and James E Carlson. Brief report: Bartlett's test of sphericity and chance findings in factor analysis. *Multivariate behavioral research*, 4(3):375–377, 1969.

[69] Gustavo F Tondello, Karina Arrambide, Giovanni Ribeiro, Andrew Jian-lan Cen, and Lennart E Nacke. "i don't fit into a single type": A trait model and scale of game playing preferences. In *IFIP Conference on Human-Computer Interaction*, pages 375–395. Springer, 2019.

[70] Alexandre N Tuch and Kasper Hornbæk. Does herzberg's notion of hygienes and motivators apply to user experience? *ACM Transactions on Computer-Human Interaction (TOCHI)*, 22(4):1–24, 2015.

[71] Daniel Votipka, Desiree Abrokwa, and Michelle L Mazurek. Building and validating a scale for secure software development self-efficacy. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, pages 1–20, 2020.

[72] Bob G Witmer and Michael J Singer. Measuring presence in virtual environments: A presence questionnaire. *Presence*, 7(3):225–240, 1998.

[73] Svetlana Yarosh, Panos Markopoulos, and Gregory D Abowd. Towards a questionnaire for measuring affective benefits and costs of communication technologies. In *Proceedings of the 17th ACM conference on Computer supported cooperative work & social computing*, pages 84–96, 2014.

[74] Nick Yee, Nicolas Ducheneaut, and Les Nelson. Online gaming motivations scale: development and validation. In *Proceedings of the SIGCHI conference on human factors in computing systems*, pages 2803–2806, 2012.

# APPENDICES

# Appendix A

# Recruitment Material

## A.1 Recruitment Poster

## A.2 Recruitment Social Media

**For Twitter, slack channels (and other social media)**

The Haptic Computing Lab is inviting individuals to participate in user experience study around haptic devices. As a participant, you would be asked to come to campus, test a haptic-feedback device and provide your feedback through a user experience questionnaire. The entire study will take 10 minutes and you will receive a **$2 Coffee and Donut (C&D) voucher** which you can redeem at E7 and CPH.

For safety reasons, you may not participate if you have the following: visual, audio or balance disorders, musculoskeletal disorders, medical implants (pacemaker), neurological disorders (e.g., stroke, seizures, migraine), or pregnancy.

Please e-mail ahmed.anwar@uwaterloo.ca for further inquiries.

## A.3 Recruitment Email

Hello,

My name is Ahmed Anwar and I am a Masters student working under the supervision of Dr. Oliver Schneider at the Haptic Computing Lab in the Management Sciences Department at the University of Waterloo. Our team is creating a user experience questionnaire for commercial devices incorporating haptic feedback. These devices include Oculus Quest VR headset, Haply, Tanvas and Ultraleap.

We are currently seeking individuals to be a part of our study. Participation will be required to come to campus and complete a small task with a haptic device and subsequently complete of a questionnaire. Participation in this study is entirely voluntary and will take approximately 10 minutes of your time.

If you are interested in participating, please contact me at ahmed.anwar@uwaterloo.ca. I will then inform you about the study location and the available time. As remuneration for participating, you will receive a **$2 Coffee and Donut (C&D)** voucher, which you can redeem at any of the C&D locations on University of Waterloo campus.

For safety reasons, you may not participate if you have the following: visual, audio or balance disorders, musculoskeletal disorders, medical implants (pacemaker), neurological disorders (e.g., stroke, seizures, migraine), or pregnancy

This study has been reviewed and received ethics clearance through a University of Waterloo Research Ethics Committee (#43516). If you have questions for the committee, contact the Research Ethics Board, at 1-519-888-4567 ext. 36005 or reb@uwaterloo.ca.

Sincerely,

Ahmed Anwar

## A.4 Recruitment in-class

Hello,

My name is Ahmed Anwar and I am a Masters student working under the supervision of Dr. Oliver Schneider at the Haptic Computing Lab in the Management Sciences Department at the University of Waterloo. Our team is creating a novel scale for measuring the user experience pertaining to Haptic feedback for commercial devices. These devices include Oculus Quest VR headset, Haply, Tanvas and Ultraleap.

We are intending to recruit participants from classrooms to get maximum response for the development of our scale. Therefore, we are requesting few minutes at the end of your class <Class Name> that takes place in <Room number> to verbally recruit students to take part in our study. The study will take place at the end of the class so no class time will be used. The equipment (Laptop and Haptic device) will be brought by the researcher. In addition, we have also checked the schedule of the room and made sure that there is no incoming class right away.

During the study, participants will perform a small task with a Haptic device and then complete our survey. As remuneration, they will be provided with **$2 Coffee and Donut (C&D) voucher**. Participation is entirely voluntary.

This study has been reviewed and received ethics clearance through a University of Waterloo Research Ethics Committee (#43516). If you have questions for the committee, contact the Research Ethics Board, at 1-519-888-4567 ext. 36005 or reb@uwaterloo.ca.

Sincerely,

Ahmed Anwar

# Appendix B

# Qualtrics Survey

UNIVERSITY OF WATERLOO

Welcome

### Welcome to the research study!

**Project Title:** Evaluating User Experience of Haptic Interactions
**Principal Investigator:** Oliver Schneider, Assistant Professor, Department of Management Science
**Student Investigator:** Ahmed Anwar, MASc Student, Department of Management Science

**Study Overview**
This purpose of this study is to develop a novel scale to measure haptic experiences from various devices. This scale is intended to assist haptic designers to effectively gauge the user experience of their haptic devices and make improvements accordingly. The data from this study will be used to validate the scale through multiple statistical techniques. If you have any questions, feedback, or concerns about the study or related research, please contact any member of the research team listed above.

Instructions

### Instructions

The present study aims to better understand users' experience with devices rendering haptic feedback. Haptic Experience is defined as:

> *"a distinct set of quality criteria combining usability requirements and experiential dimensions that are the most important considerations for people interacting with technology that involves one or more perceived senses of touch, possibly as part of a multisensory experience"*

During this study you will be asked to complete the following tasks:

1. Provide your consent to participate in this study.
2. Answer a demographic questionnaire
3. Complete a specific task using a Haptic device
4. Complete a survey based on your interaction with the device.

Consent

### Acknowledgement

By signing this consent form, you are not waiving your legal rights or releasing the investigator(s) or involved institution(s) from their legal and professional responsibilities

By clicking the button below, you acknowledge:

1. You have read the information provided regarding this study
2. You are 18 years old or older.
3. Your participation in the study is voluntary.
4. You are aware that you may choose to terminate your participation at any time.
5. You are aware of the risks associated outlined in the information letter.

- ○ I consent, begin the study
- ○ I do not consent, I do no wish to participate

**Device Selection**

## Please select a Haptic Device to complete the study.

- ○ Haply
- ○ 3D Touch
- ○ Ultraleap
- ○ Oculus
- ○ Tanvas
- ○ iPhone

**Demographic Questionnaire**

## Let's start with a little bit about yourself!

*Please note that none of the questions will be used to identify you. In addition, all of you answers be kept strictly confidential and will be only used in aggregate.*

## What is your age ?

- ○ 18-20
- ○ 21-34
- ○ 35-44
- ○ 45-54
- ○ 55-64

## What is your gender ?

- ☐ Man
- ☐ Woman
- ☐ Non-binary / third gender
- ☐ Prefer not to say
- ☐ __________ Prefer to self-describe

## What is your current education level?

Please specify the program you are **currently enrolled in or the highest achieved**. For example, if you are a 3rd year undergraduate student, you would select "Undergraduate", and if you are not a student but have a master's degree, you would select "Masters".

- Undergraduate
- Masters
- Doctorate
- Other

## What is your field of study ?

- Arts and Humanities
- Engineering and Technology
- Life Sciences and Medicine
- Natural Sciences
- Social Sciences and Management

## Please select your familiarity with the following

| | Never Heard of it | Heard of it | Some Experience | Expert |
|---|---|---|---|---|
| Human Computer Interaction (HCI) | O | O | O | O |
| Haptic Technology | O | O | O | O |

**Task Completion Step**

## Task Completion Step

Please complete a brief task using your selected Haptic device.
An experimenter will be present throughout if you need any assistance.

## Thank you for completing the task

Now please answer a few questions about your interaction with the haptic device.

**HXI Questionnaire**

**In these questions, haptic feedback refers to anything that you feel with the sense of touch. It could be vibrations, force, temperature, pressure, or any other physical sensations**

| | Strongly disagree | Somewhat disagree | Neither agree nor disagree | Somewhat agree | Strongly agree |
|---|---|---|---|---|---|
| The haptic feedback felt appropriate when and where I felt it | O | O | O | O | O |
| The haptic feedback increased my involvement in the task | O | O | O | O | O |
| The haptic feedback matched my expectations | O | O | O | O | O |
| I would prefer the system without the haptic feedback | O | O | O | O | O |
| The haptic feedback changes depending on how things change in the system | O | O | O | O | O |

**In these questions, haptic feedback refers to anything that you feel with the sense of touch. It could be vibrations, force, temperature, pressure, or any other physical**

**sensations**

| | Strongly disagree | Somewhat disagree | Neither agree nor disagree | Somewhat agree | Strongly agree |
|---|---|---|---|---|---|
| The haptic feedback helped me distinguish what was going on | O | O | O | O | O |
| The haptic feedback felt disconnected from the rest of the experience | O | O | O | O | O |
| I disliked the haptic feedback | O | O | O | O | O |
| The haptic feedback was convincing | O | O | O | O | O |
| The haptic feedback helped me focus on the task | O | O | O | O | O |

**In these questions, haptic feedback refers to anything that you feel with the sense of touch. It could be vibrations, force, temperature, pressure, or any other physical sensations**

| | Strongly disagree | Somewhat disagree | Neither agree nor disagree | Somewhat agree | Strongly agree |
|---|---|---|---|---|---|
| The haptic feedback reflects varying inputs and events | O | O | O | O | O |
| The haptic feedback felt out of place | O | O | O | O | O |

**In these questions, haptic feedback refers to anything that you feel with the sense of touch. It could be vibrations, force, temperature, pressure, or any other physical sensations**

| | Strongly disagree | Somewhat disagree | Neither agree nor disagree | Somewhat agree | Strongly agree |
|---|---|---|---|---|---|
| The haptic feedback distracted me from the task | O | O | O | O | O |
| The haptic feedback was realistic | O | O | O | O | O |
| The haptic feedback all felt the same | O | O | O | O | O |
| The haptic feedback felt satisfying | O | O | O | O | O |
| The haptic feedback fits well with other senses (e.g. sound, visuals) | O | O | O | O | O |

**In these questions, haptic feedback refers to anything that you feel with the sense of touch. It could be vibrations, force, temperature, pressure, or any other physical sensations**

| | Strongly disagree | Somewhat disagree | Neither agree nor disagree | Somewhat agree | Strongly agree |
|---|---|---|---|---|---|
| The haptic feedback was believable | O | O | O | O | O |
| I liked having the haptic feedback as part of the experience | O | O | O | O | O |
| I felt engaged with the system due to the haptic feedback | O | O | O | O | O |
| I liked how the haptic feedback itself feels, regardless of its role in the system | O | O | O | O | O |
| I felt adequate variations in the haptic feedback | O | O | O | O | O |

**Post Study Questionnaire**

## We are almost done!

**Please answer some questions about your experience with the study**

**In general, how would you rate your experience with the following ?**

| | |
|---|---|
| The Haptic Device | |
| The application | |
| The haptic feedback | |

## How noticeable was the haptic feedback ?

- ○ Far too little
- ○ Slightly too little
- ○ Neither too much nor too little
- ○ Slightly too much
- ○ Far too much

## Was the definition of Haptic feedback useful in understanding the survey questions ?

- ○ Not at all useful
- ○ Slightly useful
- ○ Moderately useful
- ○ Very useful
- ○ Extremely useful

## Has your understanding of haptic technology increased after this study ?

- ○ Definitely not
- ○ Probably not
- ○ Might or might not
- ○ Probably yes
- ○ Definitely yes

## Do you think your responses would change if you performed the task for longer period of time ?

- ○ Definitely not
- ○ Probably not
- ○ Might or might not
- ○ Probably yes
- ○ Definitely yes

# Appendix C

# R code scripts

## C.1 Data Cleaning

```
df <- read.csv('~/Library/CloudStorage/OneDrive-UniversityofWaterloo/
               HXI/HXI Code/Data/HXI-July-2-2022.csv')

df2 <- rename(df,
    'age' = 'Q2',
    'gender'='Q3',
    'edu_lvl'='Q5',
    'field'= 'Q6',
    'fam_hci'='Q7_1',
    'fam_haptics' = 'Q7_2',

    'H4'='Q1_1', 'I4'='Q1_2',
    'R4'='Q1_3', 'A4'='Q1_4',
    'E4'='Q1_5', 'E3'='Q2_1',
    'H3'='Q2_2', 'A3'='Q2_3',
    'R3'='Q2_4', 'I3'='Q2_5',
    'E5'='Q3_1', 'H5'='Q3_2',
    'I1'='Q4_1', 'R1'='Q4_2',
    'E1'='Q4_3', 'A1'='Q4_4',
    'H1'='Q4_5', 'R2'='Q5_1',
    'H2'='Q5_2', 'I2'='Q5_3',
    'A2'='Q5_4', 'E2'='Q5_5',
```

```
    'exp_device'='Q1_1.1',
    'exp_app'='Q1_2.1',
    'exp_haptic'='Q1_3.1',
    'noticeable'='Q2.1',
    'h_def'='Q3.1',
    'increase_understanding'='Q4',
    'LongerTime'='Q6.1')

df2 <- df2[-c(1,2),]
rownames(df2) <- seq(nrow(df2))

## Remove incomplete respnses
df2 <- df2[df2$Finished=="1",]

## Remove fast responses
df2$Duration..in.seconds. <- as.integer(df2$Duration..in.seconds.)
df2 <- df2[df2$Duration..in.seconds. >140,]
rownames(df2) <- seq(nrow(df2))


HXI <- df2[,c("H4","I4", "R4", "A4", "E4",
        "E3", "H3", "A3", "R3","I3",
        "E5","H5",
        "I1","R1","E1","A1","H1",
        "R2","H2","I2","A2","E2")]


## Reverse code negative items
HXI$A3 <- recode(HXI$A3, `1`=5,`2`=4,`3`=3,`4`=2,`5`=1)
HXI$A4 <- recode(HXI$A4, `1`=5,`2`=4,`3`=3,`4`=2,`5`=1)
HXI$I1 <- recode(HXI$I1, `1`=5,`2`=4,`3`=3,`4`=2,`5`=1)
HXI$H5 <- recode(HXI$H5, `1`=5,`2`=4,`3`=3,`4`=2,`5`=1)
HXI$H3 <- recode(HXI$H3, `1`=5,`2`=4,`3`=3,`4`=2,`5`=1)
HXI$E1 <- recode(HXI$E1, `1`=5,`2`=4,`3`=3,`4`=2,`5`=1)

HXI <- data.frame(sapply(HXI, as.numeric))
rownames(HXI) <- seq(nrow(HXI))
```

```
## Generate correlation matrix for all columns

rmat <- cor(HXI, use = "complete.obs")
corrplot(rmat)



### Split data into training and testing

len <- dim(HXI)[1]
training <- HXI[1:291,] # Tian 291
testing <- HXI[292:len,] # Tian 112


### Save data into seperate csv files

write.csv(training, "~/Library/CloudStorage/OneDrive-UniversityofWaterloo/
    HXI/HXI Code/Study_1/study_1.csv",
    row.names = FALSE)

write.csv(testing, "~/Library/CloudStorage/OneDrive-UniversityofWaterloo/
    HXI/HXI Code//Study_2/study_2.csv",
    row.names = FALSE)
```

## C.2 Exploratory Factor Analysis

### C.2.1 Iteration 1

```
## LOAD DEPENDENCIES

library(dplyr) # for data wrangling
library(psych) # for factor analysis
library(MVN) # for multivariate normality test
library(corrplot)
library(ltm)

## Read the data for EFA
```

```
HXI <- read.csv('~/Library/CloudStorage/OneDrive-UniversityofWaterloo/
                HXI/HXI/Code/Study_1/study_1.csv')

## Calculate correlation matrix
rmat <- cor(HXI, use = "complete.obs")

## Barlett's test for sphericity
cortest.bartlett(rmat, nrow(HXI))

## Kaiser Meyer Olkin (KMO) test for sampling adequacy
KMO(rmat)
min(KMO(rmat)$MSAi)

## Mardia Test for checking multi-variate normality
mvn(HXI, mvnTest="mardia")

## Parallel analysis for determining ideal number of factors
fa.parallel(HXI, fm = "pa")

## Factor extraction
fit <- fa(HXI, nfactors=4, rotate = "promax", fm="pa")

## Plot factor path diagram
diagram(fit)

## ITEM REFINEMENT
# removed H4, I4, R4, E3, E2,
# CL I3, A1, H1, A2, A3, A4
```

### C.2.2 Iteration 2

```
HXI1 <- HXI[,c("E4","H3","R3","E5","H5","I1",
                "R1","E1","R2","H2","I2")]
```

```
## Calculate correlation matrix
rmat <- cor(HXI1, use = "complete.obs")

## Barlett's test for sphericity
cortest.bartlett(rmat, nrow(HXI1))

## Kaiser Meyer Olkin (KMO) test for sampling adequacy
KMO(rmat)
min(KMO(rmat)$MSAi)

## Mardia Test for checking multi-variate normality
mvn(HXI1, mvnTest="mardia")

## Parallel analysis for determining ideal number of factors
fa.parallel(HXI1, fm = "pa")

## Factor extraction
fit <- fa(HXI1, nfactors=4, rotate = "promax", fm="pa")

## Plot factor path diagram
diagram(fit)

### TESTING SCALE RELIABILITY

### Reliability for overall scale
cronbach.alpha(HXI1)

### Reliability for all subscales

cronbach.alpha(HXI1[,c("R1","R2","R3")])
cronbach.alpha(HXI1[,c("I1","H3","H5")])
cronbach.alpha(HXI1[,c("H2","I2")])
cronbach.alpha(HXI1[,c("E4","E5","E1")])
```

## C.3 Confirmatory Factor Analysis

```
## LOAD DEPENDENCIES
```

```
library(dplyr) # for data wrangling
library(psych) # for factor analysis
library(MVN) # for multivariate normality test
library(corrplot)

library(lavaan)
library(semTools); # CFA, SEM


## LOAD THE VALIDATION DATA
validation <- read.csv('~/Library/CloudStorage/OneDrive-UniversityofWaterloo/
                        HXI/HXI Code/Study_2/study_2.csv')

## STRUCTURE EQUATION MODELING
model <- "MR1 =~ R1+R2+R3
          MR2 =~ H3+H5+I1
          MR3 =~ E1+E4+E5
          MR4 =~ I2+H2
         "
fit.mod <- cfa(model, data=validation)
fitMeasures(fit.mod, c("cfi","tli","srmr","rmsea"))
summary(fit.mod,fit.measures=TRUE,standardized=TRUE,rsquare=TRUE)

## RELIABILITY IN CFA

cronbach.alpha(validation)
cronbach.alpha(validation[,c("R1","R2","R3")])
cronbach.alpha(validation[,c("I1","H3","H5")])
cronbach.alpha(validation[,c("H2","I2","I4")])
cronbach.alpha(validation[,c("E1","E4","E5")])
```