Measuring Haptic Experience: Evidence for the HX model through scale development

by

Suji Nivedita Sathiyamurthy

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, 2020

© Suji Nivedita Sathiyamurthy 2020

Author's Declaration

I hereby declare that I am the sole author of this 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.

Abstract

Haptic technology allows one to receive tactile information through the sense of touch. Increasingly, designers and researchers are employing haptic feedback with the aim to improve user experience (UX). While they see the importance and significance of including haptic feedback in everyday applications, there is a lack of standardised tools to assess the quality of these experiences. They currently use qualitative methods or demos for obtaining user feedback; neither approach scales to large studies or remote work. We aim to bridge this gap and complement the existing approaches by developing an instrument that comprehensively measures haptic user experience.

We follow a systematic scale development framework to build, evaluate and establish a first draft of the haptic user experience scale - the Haptic eXperience Index (HXI), that has the potential to measure the effectiveness of haptic experiences. This scale is built upon the recent Haptic Experience (HX) model and it contributes a novel instrument that measures the five foundational constructs for designing haptic experiences: Harmony, Expressivity, Autotelics, Immersion, and Realism. We iteratively developed a set of 20 questions through a series of studies: expert reviews (N=6), face validity (N=8), cognitive interviews (N=9), and exploratory factor analysis (N=261). Our results provide evidence for the HX model's five factors, with enriched description of each factor, and implications for how to measure HX, including a first proposed draft of the HXI.

In this process, we gained an in-depth understanding of the factors we considered for developing HXI; what applications can be chosen for representing a rather diverse set of experiences; understand the limitations and define future work. This HXI is a steppingstone towards a generalized evaluation tool to measure haptic experience.

Acknowledgements

I thank my supervisor Dr. Oliver Schneider for his guidance throughout this research and for providing me with all the right resources to complete it. I am grateful to him for introducing me to the wonderful world of haptics.

I would like to thank Erin Kim for helping me to get started with the research by elucidating how the HX model was built. I thank Melody Lui for her help with setting up user studies and her inputs throughout the process.

I thank all the haptic lab members and the overall games institute community for their continued support in piloting user studies and providing feedback on different stages of this project.

I also thank my family and friends for always believing in me and providing me with all the support I needed.

Dedication

I dedicate my thesis to Sathiyamurthy, Lakshmi, Manoj and Ankesh.

I also dedicate this work to the love of user interaction design, haptic technology, and data analytics.

Table of Contents

List of Figures ix

List of Tables xi

1 Introduction 1

1.1 Research Question 2

1.2 Contribution 4

1.3 Outline 4

2 Related Works 5

2.1 Scale Development 6

2.2 Haptics and UX 6

2.3 Other scales that measure related constructs 7

2.4 Chapter Summary 9

3 Approach 10

3.1 Scale Development Process 10

3.2 Remote Studies & In-person Studies 12

3.3 Chapter Summary 13

4 Scale Inception and User Studies 14
4.1 Scale Inception 15
4.2 User Studies 17
4.3 Study 1: Face Validity 17
4.3.1 Procedure 17
4.3.2 Results 19
4.3.3 Outcome 19
4.4 Study 2: Content Validity 20
4.4.1 Procedure 20
4.4.2 Results 22
4.4.3 Outcome 22
4.5 Study 3: Cognitive Interviews 23
4.5.1 Procedure 23
4.5.2 Results 24
4.5.3 Outcome 24
4.6 Study 4: Survey Pilots 24
4.7 Study 5: Survey Administration 26
4.7.1 Study Design/Setup 26
4.7.2 Significance of Demographic and Exit Survey Questions 27
4.8 Chapter Summary 29

5 Exploratory Factor Analysis 31
5.1 Data Cleaning and Transformations 32
5.2 Correlation Analysis and Scree Plot 32
5.3 Extraction Method, Rotation and Number of Factors 33
5.4 Interpretation of the Factors 35
5.4.1 Goodness of Fit Test 35
5.5 Chapter Summary 40

6 Discussion 42
6.1 Support & Elaboration for the HX Model . . . 42
6.1.1 What each factor tells us . . . 43
6.1.2 Customization & Generalization . . . 44
6.2 Practical Matters for Measuring HX . . . 45
6.2.1 Definition of Haptic Feedback . . . 46
6.3 Applications . . . 46
6.4 Limitations . . . 47
6.5 Future Work . . . 48
6.6 Chapter Summary . . . 48

7 Conclusion 49

References 50

APPENDICES 57

A Recruitment Materials 58

B Content Validity Survey material 60

C Main Study Materials 63
C.1 Qualtrics . . . 63
C.2 Prolific . . . 73

D R code for EFA 75
D.1 R Script . . . 75

E Study Scripts 83
E.1 Face Validity Script . . . 83
E.2 Content Validity Script . . . 85
E.3 Cognitive Interview Script . . . 90

List of Figures

1.1 The five experiential factors of the HX model 3

3.1 Scale Development process - 1) Brainstorming the definitions of the modalities based on literature review chapter 4; 2) Group discussion to evaluate if the items made sense chapter 4; 3) Iterative interview sessions with expert hapticians to assess the content and its rationality chapter 4; 4) Record feedback from target population to analyze if the questionnaire is understood as intended chapter 4; 5) Evaluate tools and techniques to run a large scale online remote study through iterative methods chapter 4; 6) Online study that requires participants to complete a task with haptic experience and answer questions of the HXI based on it chapter 4; 7) A statistical method used to uncover the underlying structure of a relatively large set of variables chapter 5 11

4.1 The scale development process diagram highlighting the steps and studies explained in this chapter. Scale inception - section 4.1; Face Validity - section 4.3; Content Validity - section 4.4; Cognitive Interviews - section 4.5; Survey Pilots - section 4.6; Survey Administration - section 4.7 15

4.2 Definition of the 5 constructs as proposed by the HX model[37] 16

4.3 Evolution of items in the questionnaire through user studies and exploratory factor analysis. We note how the items were introduced as new (N), when they stayed the same (S), if/when they were changed from the initial item (C), and if/when they were rejected outright (R). 18

4.4 Bar graph showing expert's feedback on how relevant each item is to Haptic experience and its underlying factor, and on the clarity of each of these items. 21

4.5 Classification of the questionnaire items into three categories based on the cognitive interviews conducted on a sample of the target population . . . 25

4.6 Left: Prolific study dashboard. Study progress can be monitored and managed here. Participant submissions can be viewed and approved as well; Right: an image uploaded by a prolific participant for completing the study 27
4.7 Bar graphs depicting the distribution of valid user entries based on devices used and genders. 28
4.8 Top: A Bar graph showing the number of participants who answered “yes” or “somewhat” to the question “Did you feel the haptic feedback”. Bottom: A Bar graph showing the distribution of what the participants based their responses on. 30

5.1 Graphical representation of the correlation matrix. We see that there are some “clumps” of items that are positively correlated - evidence of some common factors. 34
5.2 Scree plot showing the number of possible factors in the data. Notice a drop at the 3rd point and a more subtle one at 5. This indicates that there could be 3 - 5 factors underlying the 22 variables in the data. 34
5.3 The “loading” object resulting from factor analysis obtained using the “psych” package from R. The loading highlighted in red are discarded due to low score 37
5.4 The final items and their loadings on each factor of the best factor model. The higher the loading, the more the item is correlated with its factor. We named each factor after the closest theoretical construct from the HX model, as the factors 38
5.5 Path diagram representing the five oblique factors of the final model. The factors were transformed to an oblique solution using promax. Dotted red lines mean a negative loading due to the inverted item. Note that all signs for harmony have been flipped due to factor rotation. 39

A.1 Participant recruitment poster for conducting Cognitive Interviews 58

B.1 Questionnaire provided to the experts in study 2: Expert review for questionnaire content Validity (p1) 61
B.2 Questionnaire provided to the experts in study 2: Expert review for questionnaire content Validity (p2) 62

C.1 Participant recruitment poster created on Prolific for the main study . . . 74

List of Tables

5.1 Kaiser-Meyer-Olkin factor adequacy test 33

5.2 Table showing different goodness of fit metrics to assess the fitness of the model to the data. The basic criteria for model fitness are TLI and CFI >0.90 & RMSEA and RMS values <0.05. The highlighted model has the highest TLI and CFI values and lowest RMSEA and RMS values. 36

5.3 Factor Correlation matrix of the five-factor model 40

Chapter 1

Introduction

Haptic technology is becoming an essential tool for designers seeking to create a great user experience (UX). There is mounting evidence that haptic feedback of different types can contribute to existing UX measures. Mid-air haptic feedback can make experiences measurably more pleasant, unpredictable, and creative [47], while motion seats can evoke better experiences, as measured by EEG and other physiological signals [52]. In virtual reality (VR) environments, well-designed haptic feedback can lead to increased presence [13, 19]. However, these metrics give little insight into how and why haptics contributes to peoples' experiences, and little direction for hapticians[1] to improve their designs.

Currently, designers, researchers, and hapticians use qualitative methods to gain deeper insight into their designs. We understand from Schneider et al.'s work [59] on investigation of haptic experience design that practitioners favour small, in-person acceptance tests to evaluate their designs and iterating them until it just "feels right". When communicating the efficacy of a design, in-person demos are the best way to persuade stakeholders. Neither approach scales to large quantitative studies or translate to remote work, and while some hapticians employ general scales like the AttrakDiff questionnaire [45, 32], there is still a desire for more formal measurement tools with haptic experiences. Developing a measurement tool will not just provide practitioners and researchers with better tools, it will also aid our understanding of haptic experience design. And scale development is considered critical to building knowledge in human and social sciences [20].

In this research, we iteratively develop a set of questions to measure haptic experience (HX), as proposed in the HX model [37]. The author introduced the HX model as a proposed standard for defining HX in terms of its pragmatic and hedonic factors. Based

[1] someone who is skilled at making haptic sensations, technology, or experiences [59]

on this model, we believe that the five experiential factors - “Harmony”, “Autotelics”, “Expressivity”, “Immersion” and “Realism” could be the guiding constructs important to measure HX. We conducted the first iterations of scale development process to assess how the model fits different kinds of devices and applications, understand user’s expectations and emotions around it, and analyze whether the HX model is supported by empirical evidence.

1.1 Research Question

The aim of this research is to explore the feasibility and efficiency of measuring HX using a questionnaire as an evaluation tool. With existing literature around haptic experience and the HX model in particular, we identify the important factors that influence haptic experience. The HX model articulates the key elements that constitute the user experience of haptic interaction through four design parameters, four usability requirements and five experiential factors. The hedonic section of the model defines the experiential factors that are important when it comes to evaluating HX design. Figure 1.1 shows the five hedonic factors that have the potential to influence the design of a haptic system. Through this research, we want to measure these five factors by using scale development’s best practices and further strengthen the evidence and understanding of this model.

Past research has shown enough evidence that haptic feedback has a huge effect on user experience. In order to increase the adaptability of haptic feedback in applications, it is important to refine the literature and framework around designing a haptic enabled system. Kim and Schneider[37] introduced a framework that can be leveraged as a guiding principle to create high quality haptic experiences. Further refining the constructs through developing a scale will help in understanding the components better and get granular details about how to improve HX through structured feedback. This thesis investigates the research questions:

Can haptic experience, rather than UX, be measured? Many evaluative instruments are widely used to measure user experience in various digital applications. One of the most popular and effective ways is to employ validated and reliable questionnaires to conduct user surveys. Designing a questionnaire which helps users easily understand haptic feedback and express their opinion through a likert scale rating, can be an efficient evaluation method. We explore the efficiency and credibility of this approach in this thesis.

How can we measure haptic experience? Using both qualitative and quantitative methods, we create a novel instrument that evaluates the haptic part of a system. This is

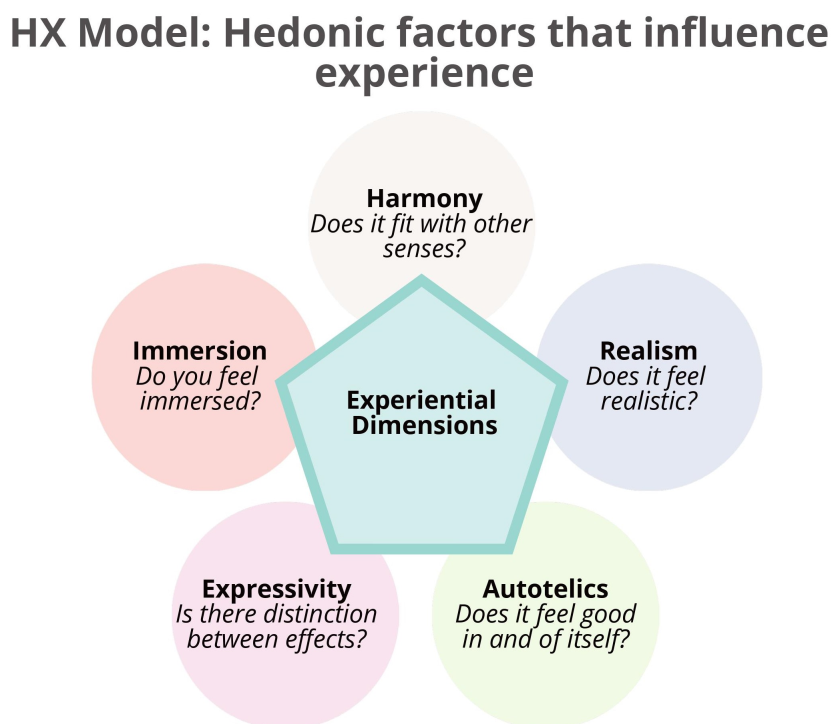


Figure 1.1: The five experiential factors of the HX model

intended to aid hapticians and researchers have a concrete and formal way to assess how their designs affect user experience and also understand the areas to target for improving it. We propose this instrument as a first draft of the HXI by following the item and scale development phases of the scale development best practices model [20]. Ultimately, the goal is to publish the HXI as a universal measurement tool for evaluating the user experience of systems with haptic feedback. This work is a steppingstone in building a valid and reliable scale to measure HX. We also expect that this process will enrich our understanding of haptic design and the components that play a major role in making it appeal to the user.

1.2 Contribution

Using the results of a series of five user studies and Exploratory Factor Analysis (EFA), we propose a 20-item questionnaire, a first draft of what we call the Haptic eXperience Index (HXI) that has the potential to measure HX in multi sensory applications. Although the HXI needs future evaluation for use in practice, our findings contribute:

1. evidence for the HX model [37], and elaboration on its constructs; what variables need to be captured in order to measure each of those construct

2. a first proposed draft of an evaluation instrument for HX which we call the "HXI"; and

3. concrete guidelines for using questionnaires to evaluate haptic experiences. E.g., Using the HXI to evaluate multi-sensory experiences with vibrotactile feedback that lasts for at least 3-5 minutes.

1.3 Outline

The thesis structure is as follows: it starts with exploring current practices in measuring user experiences and various scales that are widely in use (chapter 2). The next section explains the approach followed to develop the scale and the studies we designed (chapter 3). chapter 4 contains a detailed description of the methods and results of the five user studies we ran. chapter 5 elaborates on the exploratory factor analysis done on the collected data and how the factors are interpreted. Finally, it ends with a discussion section that highlights important findings of the empirical results and its applications along with the limitations and avenues for future work.

Chapter 2

Related Works

Haptic technology incorporates tactile experience or feedback as part of its user interface, creating a sense of touch through vibrations, motion, or other forces. Haptic modality consists of two kinds of haptic feedback: tactile and kinesthetic. The tactile feedback addresses the tactile perception of the skin, such as vibrations. The kinesthetic feedback addresses the kinesthetic perception of our own muscular effort. In this research we focus on evaluation of haptic experience with experiments designed around applications with vibrotactile feedback in addition to visual and/or audio components.

In the past decade, the widespread availability of smartphones, gaming consoles and wearable technology has made vibrotactile feedback a common phenomenon. Haptic researchers, both in academic and industrial contexts, have been designing ways to communicate via the sense of touch by means of tactile effects used to provide information such as: navigational cues[66] or textures[41]. The amount of innovation in incorporating tactile feedback in everyday applications is increasing as there is a rise in demand for such applications. This has accelerated research around adding vibrotactile feedback in digital applications. As a result, the availability of these applications in our day-to-day life has increased making it a more common and adapted form of feedback.

However, there is no standard for development and evaluation of user experiences involving haptic components. Although, there are scales that measure specific constructs that may evaluate these kinds of experiences, it is not a complete framework or tool that can be generally adopted. In-lab studies involving direct recording of user emotions and/or using existing UX scales are the most popularly adopted evaluation methods. We explored existing literature and widely adopted user experience questionnaires to identify the significance and limitations that they posit.

2.1 Scale Development

Scales are typically used to capture a behavior, a feeling, or an action that cannot be captured in a single variable or an item. The use of multiple items to measure an underlying latent construct can additionally account for, and isolate, item-specific measurement error, which leads to more accurate research findings[20]. Research shows that the scale development process involves complex and systematic procedures that require theoretical and methodological rigor. Best practices state that an effective instrument development process consists of three phases: item development, scale development, and scale evaluation. Each of theses phases have several steps that are discussed in detail in chapter 3.

2.2 Haptics and UX

There exist several instruments that evaluate UX. The popular evaluating method is to break potential factors into two main dimensions: pragmatic quality and hedonic quality. Pragmatic quality is judged by the practical, goal-oriented aspect of a product, and the efficiency and effectiveness of the users achieving their goals [31]. Hedonic quality is judged by non-tasked quality aspects of the product, such as the aesthetics of the user interface or originality of the design [32]. With the UEQ model [42], the pragmatic dimension is broken down into three aspects - Perspicuity, Efficiency and Dependability, and the hedonic quality is broken down into two aspects - Identity and Stimulation. The scales of the AttrakDiff2 [32] questionnaire offers a similar breakdown of UX evaluation, and there is a relatively high correlation between the UEQ scales and AttrakDiff2 scales. Both scales have a third dimension, attractiveness, which is an overall score for the product where ergonomic quality and hedonic quality are combined.

Another UX evaluation framework, the meCUE questionnaire [49] based on the Component model of User Experience, evaluates UX based on instrumental and non-instrumental product qualities with an emphasis on emotional reactions. This model aligns with the dimensions of the UEQ model [42] and AttrakDiff2 [32] model, where instrumental qualities are similar to pragmatic qualities (usefulness and usability), and non-instrumental qualities correspond to hedonic qualities (aesthetics, status, and commitment). The meCUE model proposes a bi-directional relationship between product qualities and emotions, in which the non-instrumental, instrumental, and emotional reactions influence one another and together determine the overall consequences. The overall rating scores of the meCUE model also have significant correlations with the attraction dimension of the UEQ and AttrakDiff2 model.

However, these UX models have limitations in capturing factors unique to HX. Haptic experiences are highly dependent on the context of the interaction and the feedback from other human senses, such as sound and visuals. These UX models are unable to measure factors such as how well the haptic interaction fits into the experience as a whole, or if the interaction improves the overall experience. Moreover, UX instruments measure the time before, during, and after an interaction, while HX focuses on the experience at the moment of touch.

This HX model [37] defines haptic experience 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.” The model focuses on interaction of the haptic technology with other parts of the design, rather than the technology independently. It proposes “design parameters of Timeliness, Density, Intensity, and Timbre; usability requirements of Utility, Causality, Consistency, and Saliency; experiential factors of Harmony, Expressivity, Autotelics, Immersion, and Realism; and the cross-cutting concern of Personalization as guiding constructs important for HX.” The HX model is used as the focused definition of HX for this study as it includes components unique to haptic experience. The experiential dimensions of the HX model enable us to measure key components such as Realism or Immersion that evaluates haptic feedback in a given environment instead of on its own, or the harmony dimension that evaluates how the haptic feedback interacts with other visual or audio feedback, which is lacking in UX models. The HX model evaluates haptic feedback’s contribution to the whole experience, while existing UX models focus on UX as its own stand-alone experience.

2.3 Other scales that measure related constructs

Other scales related to haptic experience include the Need for Touch (NFT) Scale [53]. NFT Scale measures the user preference of extracting information obtained through haptic experience. The scale focuses on two dimensions similar to the UX model - the autotelic factor and the instrumental factor. The autotelic factor refers to the hedonic-oriented response and the sensory aspect of touch, for enjoyment purposes with no actual practical purpose. The instrumental factor reflects purpose-oriented touch with a goal in mind and evaluating outcomes based on haptic feedback.

The NFT scale is user-centered and limited in measuring only users’ need for touch, not the quality of the haptic experience. The scale measures user preference for any haptic feedback and does not evaluate a specific haptic experience.

Moreover, haptic feedback and physical props can increase presence in virtual environments [12]. The presence questionnaires (Presence Questionnaire (PQ) and Immersive Tendencies Questionnaire (ITQ) [70]) offer scales overlapping with dimensions of the experiential factors of haptic experience. PQ measures the degree to which individuals experience presence in a virtual environment and potential factors that contribute to the intensity of presence, such as control factors, sensory factors, distraction factor and realism factors. ITQ measures the capability or tendency for individuals to experience presence. Similar to the NFT scale, ITQ measures users' preference or tendency to be immersed in experiences, not specific to any haptic technology. PQ touches upon factors applicable to haptics with constructs such as sensory factors and distraction factors, while the majority of the factors refer to the environment as a whole with a focus on visual feedback. PQ is most similar to the immersion or realism dimension of the HX model and does not investigate in depth other aspects of HX - Harmony, Autotelic, Expressivity.

Haptic feedback is often applied to gaming, and there exists several instruments to measure game user experiences. The Game Experience Questionnaire measures video game-playing effects through the degree of engagement elicited in gaming experiences, and focuses on four main aspects - immersion, presence, flow, and psychological absorption [21]. However, the study focuses on users' level of engagement, which is not interchangeable with the quality of the game design. The four aspects explored in the study measures gaming experience as a whole, not applicable for haptic feedback in specific. Unfortunately, this study is also not validated [43]. Another instrument focused on gaming experience is the player traits model which proposes five dimensions - aesthetic orientation, narrative orientation, goal orientation, social orientation, and challenge orientation [63]. The user scores in these five categories reflect how much the user cares about each dimension of the game design. For example, users with a high score in the goal orientation trait prefer to complete games 100%, while those who score low in this trait are more likely to be content with leaving achievements unfinished. The player trait model is user-centered and captures the user preferences for games but does not evaluate the quality of games. Another instrument in games user research is the Player Experience Inventory (PXI)[67], which measures how lower-level game design choices directly impact player's enjoyment. This scale is built based on the Means-End theory [28] that proposes the idea that users choose a product based on certain benefits or desired consequences. Users' perception of a particular design is broken down to two main categories - psychological consequences, including constructs such as immersion, mastery, and autonomy; and functional consequences, which are immediate and tangible consequences. This is a valid and reliable scale in evaluating player experience. Since a haptic feedback is huge component in modern video game development, it is critical for us to understand the PXI and draw inspiration for the haptic experience scale.

We also reviewed literature around tool embodiment and locus of attention in interaction design. Locus of attention is important in the design of computer interfaces. An interface which requires constant shifting of the locus of attention can be very frustrating and rapid changes in locus of attention slows down tasks [4]. This is a crucial aspect to consider while designing any kinds of feedback and especially important in the case of tactile feedback. Work around quantitative measurement of tool embodiment [11] introduces a novel measure, the Locus of Attention Index (LAI), that effectively measures tool embodiment. It is based on the ideology that tools when used proficiently may become an extended part of one's body. The LAI measures the user attention based on detection rate of changes in a system. The way attention and embodiment of tool increases the immersive experience of a system helps us understand how engagement, presence and performance are all directly related to one another.

However, the HXI will differ greatly from all of these scales as the end goal is focused on evaluating haptic experience of a wide variety of haptic systems, predominantly in multi sensory applications.

2.4 Chapter Summary

In this chapter we took a deeper look at the various scales popularly used in the fields of user experience, games, and virtual reality. It is very valuable to understand existing standards and draw inspirations. We also described and elaborated on the HX model which is the framework we will be following for the scale development process.

Chapter 3

Approach

We followed the method of scale development [20, 25] to create, refine, and evaluate a set of items (i.e., questions). Figure 3.1 outlines this process, and Figure 4.3 shows the evolution of items during that process.

3.1 Scale Development Process

We began with *1. Scale Inception* to develop the initial set of items based on our understanding of haptic technology and the HX model. We then iteratively developed these items through three studies: *2. Face Validity* (N=8) for novice evaluation of the items built based on the theoretical model, *3. Content Validity* (N=6) for expert review of the items' Relevance and Clarity, and *4. Cognitive Interviews* (N=9) to evaluate whether the questions are interpretable by the target population using probing questions [29] and think-aloud strategies. We then conducted a *5. Pilot Study* (N=25) to test our procedure and adequacy of data collection for performing factor Analysis. Finally, we ran our *6. Survey Administration* (N=261 after cleaning) and conducted exploratory factor analysis (EFA) to study the correlations between the items and the emergent factors. This process follows best practices of scale development and covers the first two phases of creating an evaluated instrument [20]. The description of each of these steps along with user studies and its respective outcomes have been discussed in detail in chapter 4

The three-phase scale development approach ends with evaluating the scale through confirmatory studies, tests of dimensionality and validity. This final phase (scale evaluation) is left for future work, as it is a substantial effort in its own right.

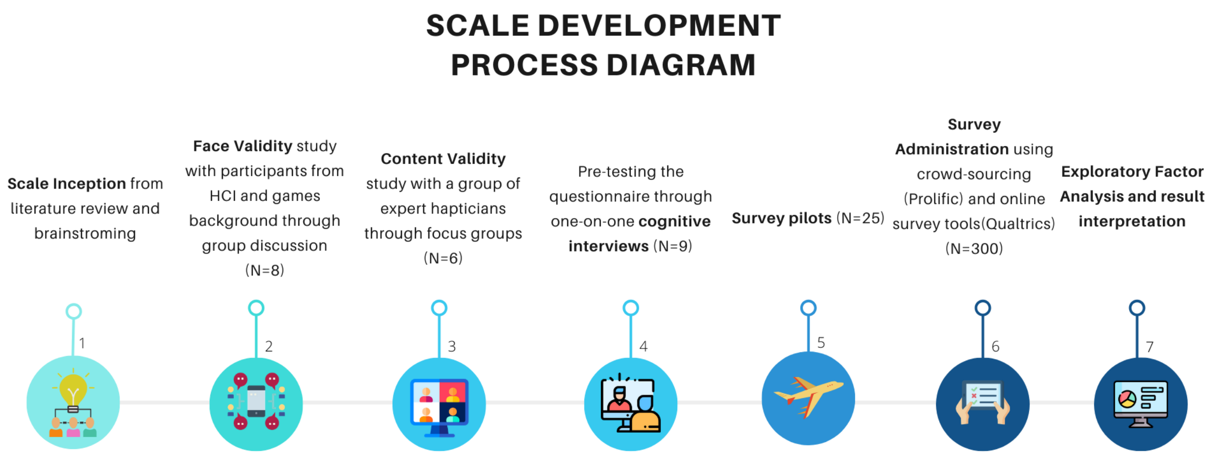


Figure 3.1: Scale Development process - 1) Brainstorming the definitions of the modalities based on literature review chapter 4; 2) Group discussion to evaluate if the items made sense chapter 4; 3) Iterative interview sessions with expert hapticians to assess the content and its rationality chapter 4; 4) Record feedback from target population to analyze if the questionnaire is understood as intended chapter 4; 5) Evaluate tools and techniques to run a large scale online remote study through iterative methods chapter 4; 6) Online study that requires participants to complete a task with haptic experience and answer questions of the HXI based on it chapter 4; 7) A statistical method used to uncover the underlying structure of a relatively large set of variables chapter 5

This process was originally designed to involve a variety of newly commercially available haptic feedback devices, including vibrotactile feedback, variable friction displays, mid-air haptics, and force-feedback; our aim is for a generalized measure for HX across devices. Unfortunately, this research was conducted during the COVID-19 pandemic, and in-person research protocols were suspended. As such, we conducted all steps from *2. Content Validity* onward remotely. We therefore decided to scope our primary data collection in the *5. Pilot Study* and *6. Survey Administration* steps to only include experiences with smartphones and gaming consoles; i.e., vibrotactile feedback in devices with mass adoption suitable for a large online study. While there is evidence that remote haptic studies with vibrotactile content produce similar feedback to in-lab studies [60], this does reduce the generality of the proposed HXI. We unfortunately must leave evaluation of our findings with other haptic modalities to future work.

The HXI is built for hapticians, researchers and designers to assess their work and identify areas for improvement. This means that the audience who will be taking the assessment are electronic media consumers who use products with haptic feedback in it for basic every-day needs, specialized entertainment, and other advanced applications.

3.2 Remote Studies & In-person Studies

The initial study was designed to be conducted with in-person participants. The face validity study was conducted in this manner in the University of Waterloo Games Institute, in a focus group style discussion.

With the situation changes due to COVID-19 we redesigned all our studies to suit remote and online operations. The studies involving one-on-one interviews and group discussions were straight forward and we were able to execute it effortlessly. However, our main data collection study had to be carried out for around 300 participants and thus posed a lot of challenges. We experimented with various user research tools like “Lookback.io” that guides participants through a step-by-step task while recording their screen or actions.

After multiple rounds of testing, we decided not to collect any user screen recording data for privacy concerns. We then decided to use Qualtrics, a survey platform to mimic the guided task process of the previous tool and replace the screen recording with a simple file upload. This approach was compliant to our data and privacy policies and also fulfilled the study requirements.

In order to recruit and manage participants to use our survey based online study, we had to look at crowdsourcing platforms. Using crowdsourcing platforms for large scale user feedback collection is known to be quick and cost effective [34, 38]. However, there are certain drawbacks that come with it. The uncontrolled asynchronous study environment with unknown interruptions and distractions poses threat to the quality of the responses. However, creating a study with simplified tasks and short activities can increase involvement in the study and improve the overall response quality [6]. Since there is some evidence of crowdsourcing tools like Amazon Mechanical Turk yielding promising results for haptic based studies [60] and collecting quality large scale data we decided to consider this approach.

We shortlisted Amazon Mechanical Turk (MTurk) and Prolific and carried out a full investigation on both these platforms to assess their functionality and merits. We found that Prolific was much easier to use as it was specifically created to be used for research purposes. Some of the notable features include - Refining/diversifying participant pool through custom pre-screening, advanced pre-screening options like - "Users with a gaming device", instant anonymous messaging between Researcher and participant and etc.

We ran pilot studies and got in touch with our participants through prolific to assess the quality and the ease of understanding of the study. We made minor modifications to improve the usability of the study and successfully launched it to 300 participants around the world. The study spaces were released in batches of 50 at different times zones in order to collect data from around the world. We were able to complete the entire data collection process in about 5 days.

3.3 Chapter Summary

In this section, we briefly explained the different steps we followed in creating and validating the HXI through different user studies. We also discussed about how the approach shifted from in-person in lab studies to online participants and accommodating remote tasks.

Chapter 4

Scale Inception and User Studies

With an idea of what we want to measure we went through an item generation phase and an elaborate validation phase using several user studies. The term “item” refers to an individual question or statement that respondents are meant to answer. The term “scale” refers to a collection of items intended to measure the same construct, and the items almost always have the same response format. Before getting to the actual item generation phase, there are a few decisions that are to be made regarding the type of question, the type of answer choices and the number of choices.

The first decision after deciding what to measure is deciding exactly how to measure it. We decided to construct statements to which the participants can specify their level of agreement. We decided to go with quantitative responses (using a likert scale) since our goal was to perform factor analysis on the data collected.

Second, we had to decide how many response options to give the respondents. In current practice, most rating scales, including Likert-type scales and other attitude and opinion measures, contain either 5 or 7 response categories [15]. A 5 - point Likert scale was used to increase response rate and response quality along with reducing respondents’ “frustration level” [14]. As we foresee the scale to have 25 items, a 5 - point likert scale was the best option as there is evidence for higher reliability in 5-point scales [35, 46].

Finally, we debated about having a neutral midpoint versus a midpoint that stands for “Not Applicable”. We decided that we were designing our studies in such a way that all the questions are applicable to the context of the task. Including a not applicable (N/A) option will create an imbalance in the scoring as it would mean that the particular item is excluded from the response analysis. Thus, we went with a neutral midpoint (i.e., a midpoint that was scored 3 in a 5 - point scale).

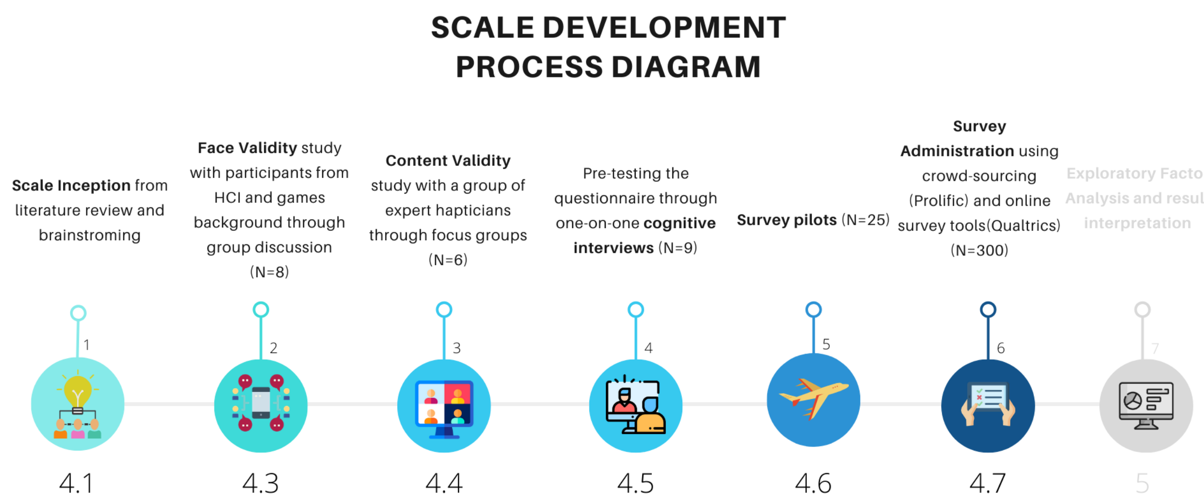


Figure 4.1: The scale development process diagram highlighting the steps and studies explained in this chapter. Scale inception - section 4.1; Face Validity - section 4.3; Content Validity - section 4.4; Cognitive Interviews - section 4.5; Survey Pilots - section 4.6; Survey Administration - section 4.7

With all these decisions made, we carried out the item generation phase by breaking down each construct into measurable attributes and framed them into sentences. The entire process Figure 4.1 of coming up with the final questionnaire is discussed in this chapter.

4.1 Scale Inception

The first step in the scale development process is Scale Inception, wherein we articulate the domain we intend to measure [20] and generate items for the domain. A domain or construct refers to the concept, attribute, or unobserved behavior that is the target of the study [33]; here, we intend to measure the experiential factors of the HX Model [37]. The items are built on the definitions of the dimensions and related constructs from other

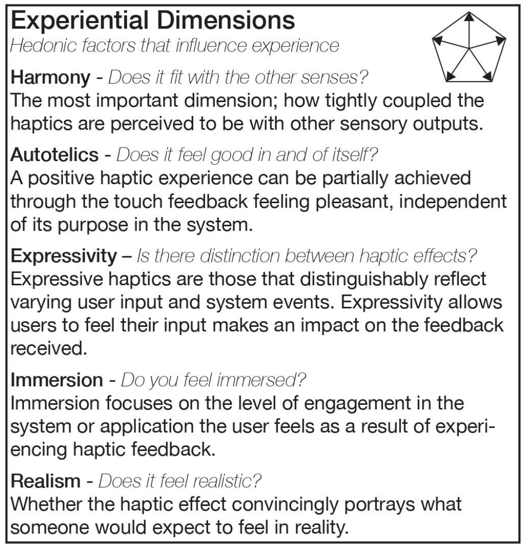


Figure 4.2: Definition of the 5 constructs as proposed by the HX model[37]

scales and aspects that heavily influence it. Figure 4.2 shows the definition of the five constructs from the HX model we are trying to measure. For example, with "Immersion": an immersive experience is said to keep the user "engaged". So, a potential item to measure immersion is measuring the level of engagement, which might be worded into an item such as "I felt engaged with the system". A mix of positive and negative items were generated for cross-validity and robustness; we had at least one negative item (e.g., The feedback felt out of place) as an attention check to help identify respondents disregarding the wording of the items. Through in-lab brainstorming sessions, we refined the pool of items based on the following criteria:

- ease of understanding
- relevance to the domain (i.e., haptic feedback)
- relevance to the dimension (e.g., Immersion)
- ability to extrapolate to varied haptic applications
- ability to extrapolate to all types of haptic feedback (e.g., vibrotactile, force etc.)

We aimed for 5 items for each of the 5 constructs because best practices recommend having 4 to 5 questions per dimension [39, 58]. The result of this step was an initial set of 25 items (5 items for each construct), shown in Figure 4.3. Figure 4.3 represents the evolution of items through the scale development process. The “All Items” column is a repository of all the items that were generated throughout the item generation and refinement steps. The “Initial Items” column denotes items that resulted from the scale inception phase (tagged as New(N)). The subsequent columns represent the other steps explained below and have been tagged “same(S)” if the item did not change at this stage; “changed(C)” if the item was modified in the current stage; “removed” if the item was dropped based on the results of the stage; “new” if the item was generated in the current stage or moved from a different factor in the model.

4.2 User Studies

We conducted three user studies and a formal pilot to refine our initial item set and prepare the instrument for our main data collection. These user studies provided both qualitative and quantitative data to analyze the constructs based on validity, clarity, and coherence. *Study 1: Face validity*, was conducted to understand if the items measure what they intend to measure. *Study 2: Content validity*, was performed to assess the extent to which the items represent all facets of the given construct. *Study 3: Cognitive interviews*, were carried out to comprehend user reaction to the items and understand the thought process it provokes. Finally, *Survey Pilots* were conducted to experiment different crowd-sourcing platforms and study designs.

4.3 Study 1: Face Validity

The purpose of the face validity study is to have new perspectives from participants with a relevant background (in the field of haptics or HCI) to assess the potential of suggested items before we conduct more thorough discussions with experts.

4.3.1 Procedure

We recruited 8 participants with backgrounds in HCI, VR and gaming technologies to screen the questions in an in-lab focus group. We started by explaining what we consider haptic feedback in a system and demonstrated it using the Nintendo switch game “1-2 Switch”. We passed the joy con around the table to let our participants play the game

No.	Dimension	All Items (T=32)	Initial Items (T=25)	Face Validity (T=25)	Content Validity (T=22)	Cognitive Interview (T=22)	EFA (T=20)	Final Questions (T=20)	
1	**Harmony**	The feedback fits well with the other senses	N	C	C	S	R		
2		The experience was more than the sum of its parts	N	C	C	S	R		
3		The feedback felt disconnected from the system	N	S	C	S	C	The haptic feedback felt disconnected from the rest of the experience	H3
4		Feedback fit into its context	N	C	C	S	C	The haptic feedback felt appropriate when and where I felt it	H4
5		The feedback felt out of place	N	C	C	S	C	The haptic feedback felt out of place	H5
6		The haptic feedback distracted me from the task	NA	NA	NA	NA	N	The haptic feedback distracted me from the task	I1
7	**Autotelic**	I look forward to feeling the feedback	N	C	R	R	R		
8		The feedback felt good	N	S	C	S	C	The haptic feedback felt satisfying	A1
9		The feedback would feel good by itself	N	S	C	C	C	I like how the haptic feedback itself feels, regardless of its role in the system	A2
10		I disliked the feedback	N	S	C	S	C	I disliked the haptic feedback	A3
11		I would prefer the game without the feedback	N	C	C	S	C	I would prefer the system without the haptic feedback	A4
12		I like having the haptic feedback as part of the experience	NA	NA	NA	NA	N	I like having the haptic feedback as part of the experience	H2
13	**Expressivity**	The feedback felt expressive	N	C	R	R	R		
14		The feedback all felt the same	N	S	C	S	C	The haptic feedback all felt the same	E1
15		I felt enough variation in the feedback	N	C	C	S	C	I felt adequate variations in the haptic feedback	E2
16		The feedback helps me distinguish my actions	N	S	C	S	R		
17		The feedback changes depending on my actions	N	C	R	R	R		
18		The haptic feedback changes depending on how things change in the system	NA	NA	N	S	S	The haptic feedback changes depending on how things change in the system	E4
19		The haptic feedback reflects varying inputs and events	NA	NA	N	S	S	The haptic feedback reflects varying inputs and events	E5
20	**Immersion**	The feedback was immersive	N	S	R	R	R		
21		I felt engaged with the system	N	S	C	S	C	I felt engaged with the system due to the haptic feedback	I2
22		I felt detached to the real-word because of the feedback	N	C	R	R	R		
23		I was aware of my response to the feedback	N	S	R	R	R		
24		The feedback distracted me away from the system	N	S	C	S	R		
25		The haptic feedback helped me focus on the task	NA	NA	N	S	S	The haptic feedback helped me focus on the task	I3
26		The haptic feedback increased my involvement in the task	NA	NA	N	S	S	The haptic feedback increased my involvement in the task	I4
27		The haptic feedback helped me distinguish what was going on	NA	NA	NA	NA	N	The haptic feedback helps me distinguish what was going on	E3
28	**Realism**	The overall experience was realistic	N	S	C	S	C	The haptic feedback was realistic	R1
29		The feedback matched my expectations	N	S	C	S	R		
30		The experience was close to a real-world experience	N	C	R	R	R		
31		The feedback was convincing	N	C	C	S	C	The haptic feedback was convincing	R3
32		The feedback was believable	N	S	C	S	C	The haptic feedback was believable	R2

N	New	S	Same	C	Changed	R	Removed	NA	Not Applicable

Figure 4.3: Evolution of items in the questionnaire through user studies and exploratory factor analysis. We note how the items were introduced as new (N), when they stayed the same (S), if/when they were changed from the initial item (C), and if/when they were rejected outright (R).

themselves and understand what we were talking about. The HD rumble effects in the 'Ball Count' game served the purpose of explaining how haptic feedback communicates information through the sense of touch. The questionnaire items were displayed on a plasma TV grouped as factors. We went through each factor and the respective items one by one and discussed individual and group opinion on it. The participants were then asked to comment on the items of the questionnaire based on four criteria: discuss individual perception of the item, complexity level, clarity of construct measured, and if the item served its intended purpose. We followed a structured script Appendix E to complete this study.

4.3.2 Results

Participants identified and debated unclear words and complex concepts. For example, the item "I look forward to feeling the feedback" was deleted because the description was vague. The item "The feedback felt appropriate given its context" was changed to "The haptic feedback felt appropriate when and where I felt it" to clarify what "context" meant. Changes to the questionnaire were made on direct participant suggestion provided they were valid and efficient. Other items with unclear or weak suggestions were left unchanged so that we have more qualitative data from the subsequent studies for making better decisions. Perception of the items among the participants were helpful in giving a general indicator of how accurate the items reflect measurements of haptic feedback. For example, participants brought up that immersion seems to be only applicable to games and questioned whether the dimension was fundamentally relevant to haptic experiences.

This step was instrumental in getting qualitative feedback to help refine the questionnaire. Recording user emotions on each item and the scale as whole gave us direction in item refinement. We tried to reduce ambiguity by replacing confusing or complicated wording with clear and concise ones, We also simplified the concepts behind each of the items and further noted down suggestions and criticisms to be revisited in later stages.

4.3.3 Outcome

As a result, 12 of the items were modified. Some of the items which lacked in clarity was kept as is in the pool for further investigation as it needed more evidence for why it needs to change and for gaining insights on how it needs to change as well. For example, participants were expecting more elaborate questions in certain instances like "The feedback was expressive". The word expressive was ambiguous and the participants' comprehension varied. As this lacked in uniformity of how it was understood, it needed refinement in future iterations. Other important questions raised during this session include - whether researchers can choose to evaluate some but not all the dimensions and how a participant will respond to a question that is not applicable.

4.4 Study 2: Content Validity

In order to validate the content of the items, we invited six hapticians as subject matter experts to attend a group discussion evaluating each item. H1 (haptician 1) is a haptic experience industry expert based in North America with 20+ years of experience in the field. H2 is a Europe based researcher whose work is majorly around medical equipment simulations. H3 and H6 are research scientists with backgrounds in physical sensory feedback and haptic intelligence. H4 is an accomplished academic scholars in the field of Human Computer Interaction, affective haptics, haptic perception, and haptics for VR/AR. H5 is a academician researching on Human-Robot Interaction. We selected our experts based on diversity in their fields of study and demographic location.

We asked these experts to evaluate the items based on clarity and representativeness of the corresponding dimension. Expert review was an opportunity for us to trim, reword, and identify overlaps among the items. We focused on modifying items that were commonly ranked as unclear or irrelevant.

4.4.1 Procedure

Experts were interviewed in groups of two. We went through each of the items and asked experts to challenge and critique the relevance of the item to the dimension. Whenever an expert suggested the wording in an item was ambiguous, we further explained our thought process and went back and forth until the initial intentions were clear, and a conclusion was reached. A qualtrics survey was provided to the experts at the end of the interview session to rate each item based on its relevance and clarity. The final survey results were used to assess the clarity and relevance of the items and visualize the feedback from all six experts. Figure 4.4 is a graphical representation of the results of the survey. These survey results along with qualitative feedback was used for fine tuning the questionnaire.

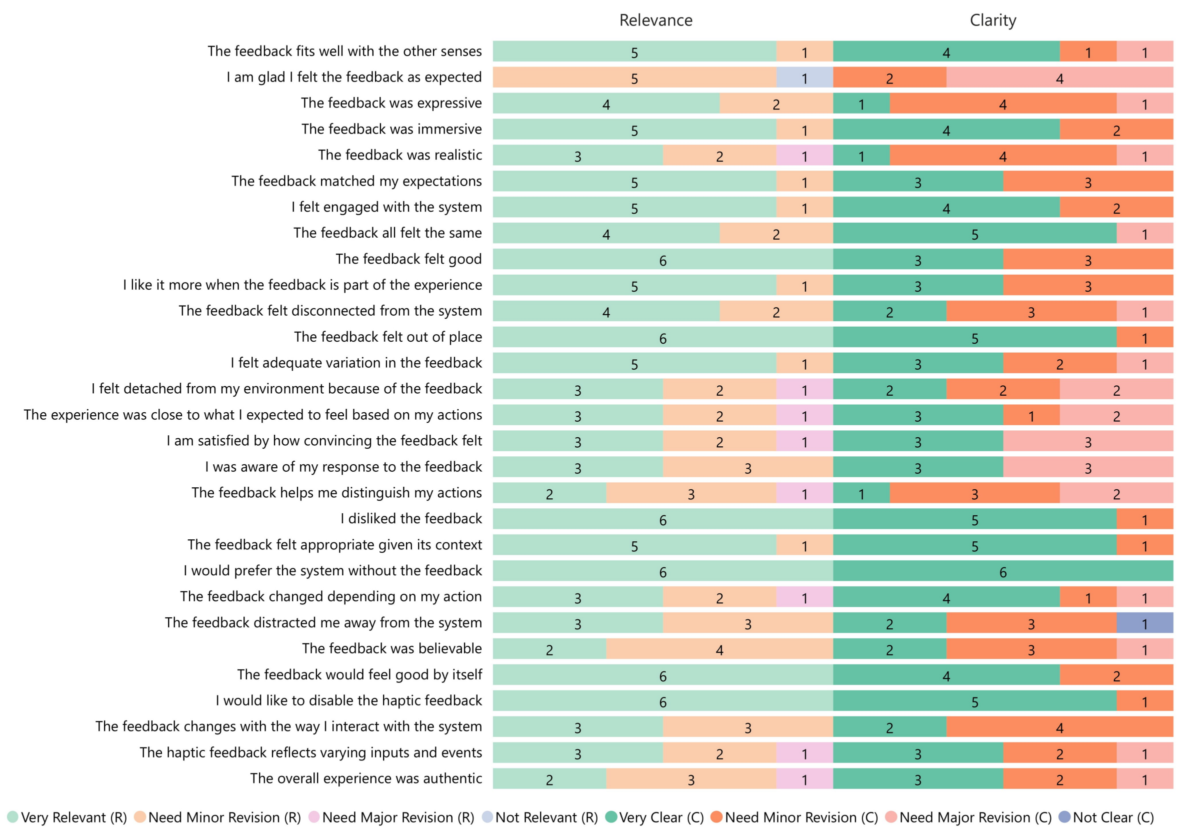


Figure 4.4: Bar graph showing expert's feedback on how relevant each item is to Haptic experience and its underlying factor, and on the clarity of each of these items.

4.4.2 Results

The experts pointed to several items that were confusing or unrelated to the corresponding dimension. First, experts suggested changing "feedback" to "haptic feedback" across all items to make items more explicit and reduce ambiguity. Another point commonly brought up among the three interview groups was that some dimensions or items will only apply to certain haptic applications. For applications that are not designed to reflect one or more experiential factors of the haptics model, users will give it a low rating and the questionnaire will receive a lower overall score. For example, expressivity does not apply to a dental simulator and haptic feedback in this context will score low in the expressivity dimension. With the item "The feedback was expressive", 4 out of 6 experts, indicated that the question was unclear or irrelevant. Similarly, the item "The feedback was realistic" was noted as unclear or irrelevant by 2 of the 6 experts. In an effort to resolve this problem, we raised the question of generality vs. specificity. If the items were too broad, the overall score might be biased towards haptic technologies with unique use cases. One proposed solution for this was to have the questionnaire tailored by the person who is using it to measure haptic feedback based on the device. However, the questionnaire proposed by this study would no longer be a universal standard to measure the quality of haptic feedback as intended.

The experts also brought up that the items which were vague make it hard to pinpoint which dimension of the experience was being evaluated. For example, items such as "the feedback felt good" can be hard for users to distinguish from the experience as a whole. It could be tricky and sometimes impossible to isolate and evaluate solely the haptic component without taking context into account. The rating received might not only reflect the quality of the haptic feedback, but also other parts of the experience.

4.4.3 Outcome

After this study, we had strong evidence that the phrase "Haptic feedback" needed to be included in the items. However, this may make the questionnaire difficult to understand for participants unaware of its definition. We formulated a definition of haptic feedback and included it as part of the HXI to make sure the items are clear while maintaining the clarity of the questionnaire. We included the definition *"Haptic feedback refers to anything that you feel with the sense of touch. It could be vibrations, force, temperature, pressure, or any other physical sensation"*, in the questionnaire to inform our respondents what the term "Haptic feedback" means. As a result of this iteration, a majority of the questionnaire was modified.

4.5 Study 3: Cognitive Interviews

The questionnaire was then pre-tested with a sample of the target population. This study was conducted as a one-on-one interview in two iterations, with 4 to 5 participants in each round. We recruited participants from varying professional backgrounds and levels of exposure to technology. For example, we had a chef, high-school teacher, electrical engineer, students, retired businessman, etc. as participants in this study. Most of our participants were either unaware of haptic technology or had vaguely heard about it. The definition of haptic feedback was incorporated in this version of the HXI as well to capture user feedback.

4.5.1 Procedure

The participants were recruited by publishing a flyer Figure A.1 with study details. The flyer was circulated through email and slack, and the recruitment was carried out using the snowball sampling method. The recruited participants were in the age group of 18-64 (with participants from both extremes of this limit). The following stimuli/experiences were used to represent haptic feedback and experiences: Apple watch, Fitbit, smart phones, Nintendo switch and PS4. To ensure randomness and variety in the haptic experiences, we selected different types of devices and applications and monitored how the questionnaire performed. The participants were guided to complete a task that enabled them to experience haptic feedback in an application. After the task, participants were asked to open the questionnaire on qualtrics, read each item and explain how they understand the item. If participants have questions or doubts, they were not clarified but instead acknowledged and recorded. If participants were stuck on an item, probing questions were asked to help them with their thought process. Refer section E.3 for details on the probing questions used. Based on the results of the first round of interviews, we decided to remove smart watches and fitness trackers from the list of experiences as they did not match the theme of the research i.e., multi-sensory experiences, in most instances. The congruence between the participants' response to the items indicated that smartwatches and fitness trackers might be out-of-scope for this research. We also learnt that participants needed more than 3 minutes to understand and adapt to the haptic sensation to answer our evaluation questions. This raised the question of how participants' responses will change with the length of exposure to the experience. We decided to capture this data in our main study for analysis and future work.

For the subsequent iteration, we increased our task duration to 5 minutes and reduced the set of experiences to games and haptic videos. The combined results of the sessions were used to build the table Figure 4.5 to identify and refine the questions that were highly confusing and the ones that are most likely to be misunderstood.

4.5.2 Results

Including a definition for haptic feedback enhanced the understanding of the items. We noticed that the participants liked having a definition section to refer to and clarity on what the items meant. Decision to modify or drop an item was based on common opinions and feedback. Figure 4.5 shows how we compared the responses of each participant using three categories - Accurately & Easily understood, Misinterpreted/Unsure, Confusing/-Complicated. As mentioned earlier, majority of "misunderstood" or "confused" feedback came from users who used smart watches or fitness trackers to complete the study. This indicated that certain factors like "Immersion" and "Realism" may be confusing to users evaluating a simple, stand-alone, notification type information. No items were dropped in this stage. Slight modifications were made for a few questions.

4.5.3 Outcome

In addition to modifying some of the items, the cognitive interviews helped in finalizing the devices and applications we used for the main study. We shortlisted applications and devices that most participants found easy to access and use. Identification of other variables that can potentially affect haptic experience was also a major take away from this study. For example, two participants with varying duration of exposure to the application, described and rated the experience differently.

4.6 Study 4: Survey Pilots

Through three pilot studies, we were able to design an online study that enabled participants to use commercially available haptic enabled devices (e.g., Smartphone and gaming consoles) to experience haptic feedback and complete the haptics experience index as a survey afterwards.

We first conducted the study using Qualtrics [7], an online survey platform. The purpose was to gauge how smoothly participants could navigate the study. The participants in this study were recruited through university mailing lists and snowballing techniques.

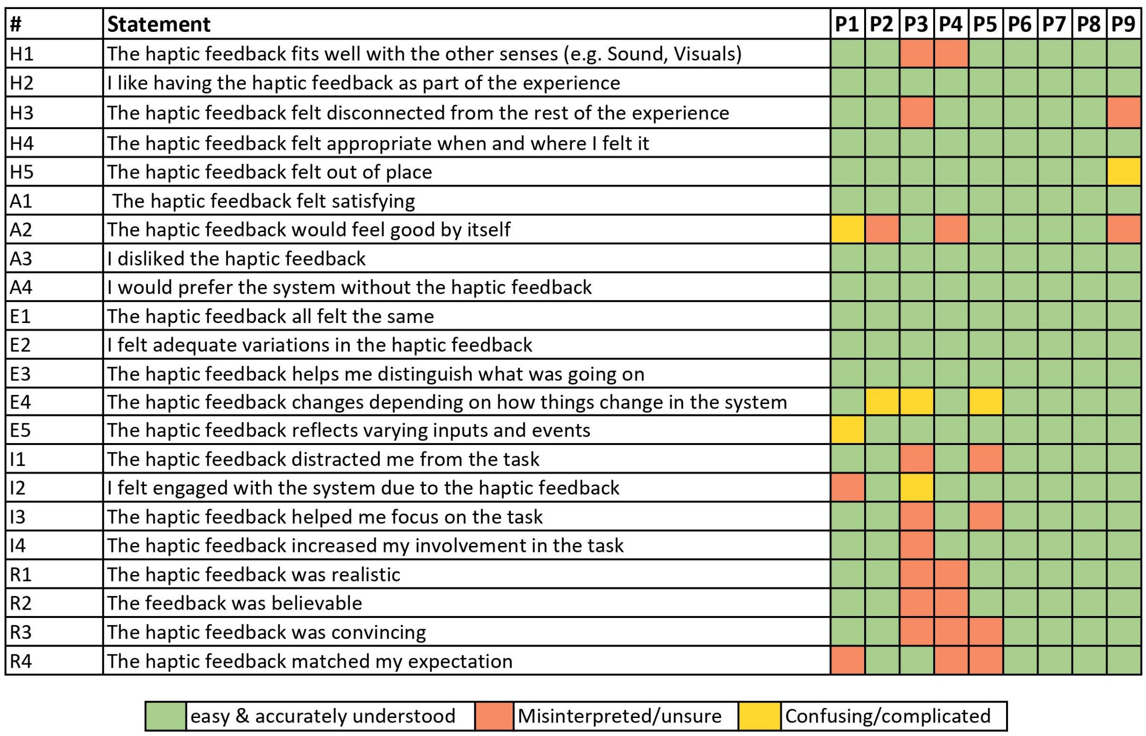


Figure 4.5: Classification of the questionnaire items into three categories based on the cognitive interviews conducted on a sample of the target population

Approximately 6 participants completed this pilot and provided additional feedback and suggestions over an informal chat or email. We were interested in the estimated duration to complete the study, clarity of instructions, and if correct options showed up based on participants' previous selections. Three iterations of the study were conducted on Qualtrics, where improvements on the design of the study based on user feedback were made.

The next step of the pilot study was to conduct the study on Prolific, a crowd-sourcing online platform to recruit participants [5]. The participants in this study were from the participant pool. The participants were between 18 and 34 years old and 36% of the recorded responses were female participants. The first iteration on Prolific involved 10 participants. The results showed that 9 participants used smartphones, and only 1 participant used a gaming console as the preferred haptic device for the study. The 9:1 ratio of smartphones vs. gaming consoles suggested a restriction might need to be placed to limit smartphone participants to ensure study results were valid with both devices, since the questionnaire should be applicable to any haptic device and not just smartphones. We then conducted a second study on Prolific, where we specified that all participants have to complete the survey with a gaming console. The survey was validated and improved based on feedback from the pilot participants.

4.7 Study 5: Survey Administration

With the help of the pilot studies, a remote online study was deployed using Qualtrics and Prolific. The participant pool on Prolific were notified of the study once published and they could complete it at their convenience. The study was released to 50 participants at a time to reach audience from various time-zones. This also helped in keeping the study under control and preventing issues or glitches affecting the entire participant pool. The chat feature was useful to communicate with remote participants when necessary and run the study as effectively as possible.

The data collected were from two categories of devices - smartphones and gaming consoles. All the applications had vibrotactile feedback as part of the experience and a total of 302 responses were recorded. 261 of those responses were identified and verified to have experienced haptic feedback and evaluated it using the HXI.

4.7.1 Study Design/Setup

The following steps were created on Qualtrics [7] to facilitate this study

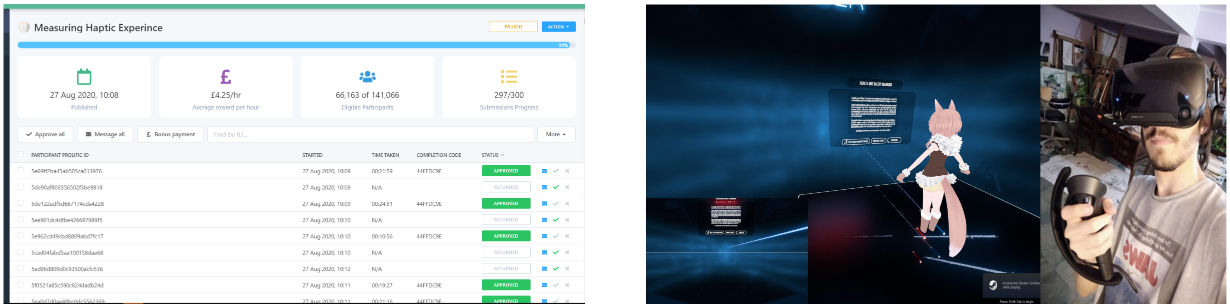


Figure 4.6: Left: Prolific study dashboard. Study progress can be monitored and managed here. Participant submissions can be viewed and approved as well; Right: an image uploaded by a prolific participant for completing the study

- Select Device & Application
- Review & Consent to Study
- Demographic Questionnaire (age, gender, familiarity with haptics, etc.)
- Instructions to complete the task (e.g., Play Animal crossing on Nintendo Switch)
- Upload anonymized images to confirm task completion
- Complete the HXI (22-items, randomized order)
- Exit Survey (rate overall experience and additional comments)

A survey with the above sections was published and a link to the study was added to Prolific. Appendix C shows a sample of this survey. Figure 4.6 shows the Prolific dashboard where the study was monitored and managed and also a sample of the image that was uploaded as part of completing the study. The participant ID, study ID and session information was captured through URL parameters to verify the entries and provide remuneration to successful submissions.

4.7.2 Significance of Demographic and Exit Survey Questions

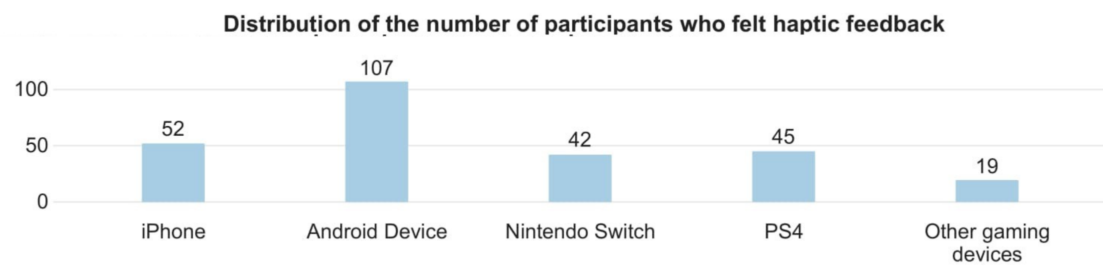
Through the demographic questionnaire, we were able to calculate metrics like distribution of age groups, gender, familiarity with the haptic device, application, and haptic technology. Figure 4.7 shows distribution graphs for some of the demographic questions. The exit

Figure 4.7: Bar graphs depicting the distribution of valid user entries based on devices used and genders.

survey helped in identifying participant's satisfaction with haptic feedback and whether they understood the intention of the study. Questions in the section were used to assess the quality of responses and used in the data cleaning process. Additional questions regarding duration of use were collected for future work. Figure 4.8 shows distribution graphs for some of the exit questions.

4.8 Chapter Summary

This chapter elaborates on the five user studies that we ran in the process of developing a scale. We have highlighted the procedures followed, results and outcomes of each study. The result of completing the steps in this chapter is a well-defined questionnaire with 22 items in which respondents specify their level of agreement to each statement in five points: (1) Strongly disagree; (2) Disagree; (3) Neither agree nor disagree; (4) Agree; (5) Strongly agree. Each of the final 22 items have been carefully crafted based on literature review, feedback from experts and a sample of the target population. The questionnaire was then incorporated as a survey for data collection. The final data collected was used for performing statistical analysis that is explained in detail in chapter 5. Appendix A has recruitment materials for studies 1 to 4. Appendix B has the survey content of study 2 and Appendix C contains the qualtrics and prolific materials used to run study 5.



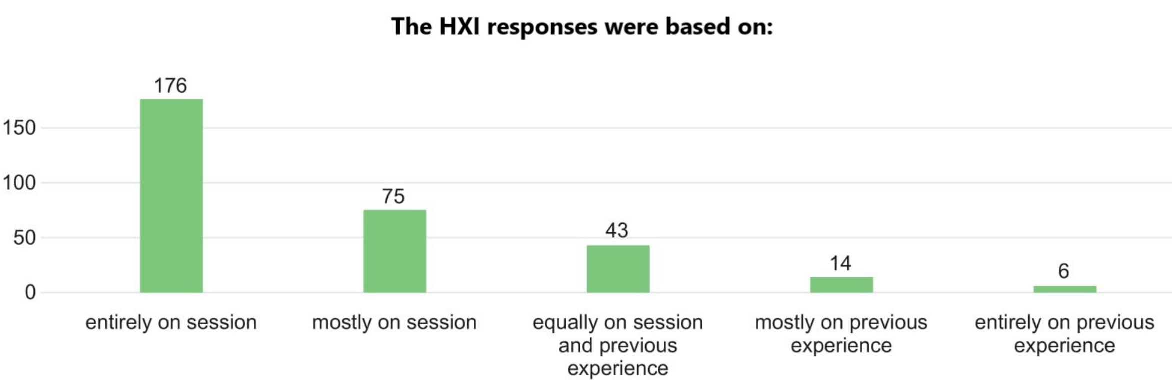


Figure 4.8: Top: A Bar graph showing the number of participants who answered “yes” or “somewhat” to the question “Did you feel the haptic feedback”. Bottom: A Bar graph showing the distribution of what the participants based their responses on.

Chapter 5

Exploratory Factor Analysis

Factor analysis is a statistical method used to study the dimensionality of a set of variables. Factor analysis identifies categories of similar statements. The common factor model reflects the view that the effect of one or more common factors (i.e., shared latent causes) as well as unknown variable-specific variance is expressed by covariation between a set of observed variables. Items are called indicators of the latent variables that underlie their covariation (a form of a reflective latent variable model) in the common factor model. Thus, factor analysis partitions variation in the indicators into common variance and unique variance. Common variance reflects the shared influence of underlying factors on an indicator. Unique variance represents a) reliable variation in the item that reflects unknown latent causes, and b) random error due to unreliability or measurement error.

Each indicator has a communality, which is the total variance in an indicator explained by latent factors. The remaining unexplained variation is called an indicator's uniqueness. Factor analyses are conventionally conducted on standardized data, meaning that the communality and uniqueness should sum to 1.0 for each indicator. Altogether, if a factor model is a good representation of the data, then the correlation among items that load onto a factor should be largely attributable to the factor. More generally, a good factor model should have high communality estimates and low uniqueness for all items. If the uniqueness of an indicator is high, it indicates that variation in the indicator is not explained by the specified factor structure

The factor analyst's first task is to determine how many categories are sufficient to capture the bulk of the information contained in the original set of statements. In factor analysis, latent variables or indicators represent unobserved constructs and are referred to as factors or dimensions. The goal of factor analysis is to model the interrelationships among items. The procedure has four steps:

- Data Cleaning & Transformation
- Correlation Analysis and Scree Plot
- Selecting Extraction Method, Rotation and Number of Factors
- Interpretation of the Factors

5.1 Data Cleaning and Transformations

As part of cleaning the data, we removed incomplete submissions, submissions with invalid image uploads and submissions of users who reported that "haptic feedback was not noticeable at all." A copy of the data was transformed into $p \times m$ matrix in order to perform exploratory factor analysis, where p = number of valid observations (N=261) and m = number of observed variables (N=22). All responses were complete and invalid responses were excluded. The transformed data was loaded into R for statistical analysis and interpretation.

5.2 Correlation Analysis and Scree Plot

The first step in factor analysis is to look at the correlation of the observed variables for patterns [2].

The process begins with a correlation matrix for all the individual items. Using this matrix as a starting point, factor analysis examines the patterns of covariation represented by the correlations among items. This amounts to a provisional assertion that a model having a single latent variable (i.e., a single factor), with a separate path emanating from it to each of the items, is an accurate representation of causal relationships.

An initial look at the correlation table or the graphical representation of the correlation matrix generally gives an idea about variables that cluster together. In Figure 5.1, we can see clusters of blue and red dots which is an indication of factor existence.

Once we collect information around variable correlation, it is a good idea to perform tests to confirm the adequacy of data collected. A Kaiser-Meyer-Olkin (KMO) [3] test is used in research to determine the sampling adequacy of data that are to be used for Factor

Kaiser-Meyer-Olkin factor adequacy										
Overall MSA = 0.93 **MSA for each item**										
H1	H2	H3	H4	H5	E1	E2	E3	E4	E5	R1
0.9720	0.9436	0.9429	0.9543	0.9482	0.8340	0.8697	0.8987	0.9069	0.8960	0.9105
R2	R3	R4	I1	I2	I3	I4	A1	A2	A3	A4
0.9253	0.9720	0.9612	0.9323	0.9278	0.8975	0.9209	0.9652	0.9392	0.9388	0.9225

Table 5.1: Kaiser-Meyer-Olkin factor adequacy test

Analysis. Social scientists often use Factor Analysis to ensure that the variables they have used to measure a particular concept are measuring the concept intended. The KMO test allows us to ensure that the data we have is suitable to run a Factor Analysis and therefore determine whether we have set out to measure what we intended. The statistic computed is a measure of 0 to 1. Interpreting the statistic is relatively straightforward; the closer it is to 1, the better the data is. The overall KMO for our data is 0.93 and all our observed variables show values greater than 0.90 as shown in Table 5.1. This suggests that the collected data is adequate for factor analysis.

Next, we move on to determining the number of factors that the Principal Component Analysis (PCA) suggests, by plotting and interpreting the scree plot (as seen in Figure 5.2). The initial scree plot using parallel analysis suggested that the number of factors was between three and five. .

5.3 Extraction Method, Rotation and Number of Factors

In order to determine the number of factors, we investigated the scree plot [51] to understand the number for underlying factors. The initial graph strongly indicated the presence of 3 highly probable factors. Among the many ways to do latent variable exploratory factor analysis (EFA) [51], we used Maximum likelihood (ML), Ordinary least squares (OLSS) and Principal Axis Factoring (PAF) extraction methods for the initial evaluation. We also tried various rotations including both oblique and orthogonal types to understand the correlations between the variables. Through factor rotation, we can make the output more understandable and it is usually necessary to facilitate the interpretation of factors.

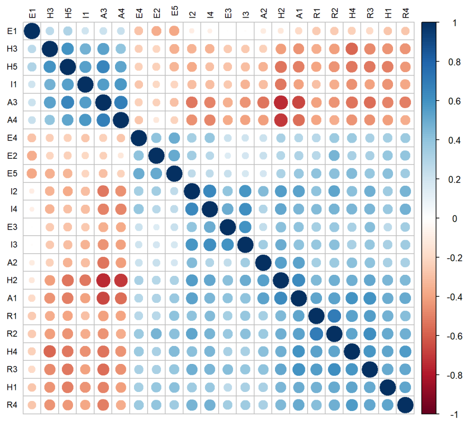


Figure 5.1: Graphical representation of the correlation matrix. We see that there are some "clumps" of items that are positively correlated - evidence of some common factors.

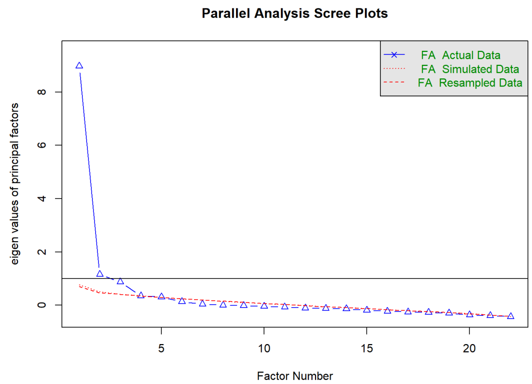


Figure 5.2: Scree plot showing the number of possible factors in the data. Notice a drop at the 3rd point and a more subtle one at 5. This indicates that there could be 3 - 5 factors underlying the 22 variables in the data.

While EFA does identify latent directions in the data, there is no guarantee that these directions are inherently interpretable. Thus, we may wish to rotate the axes of the factor. It is also important to note that rotation does not change the fit or the amount of variance explained. Instead, it redistributes the variance across factors to aid in interpretation. This may change the eigenvalues of each factor, but not their sum (i.e., same total variance explained). Rotations should usually be applied after selecting an appropriate number of factors. Orthogonal rotations such as "varimax" try to rotate the factors to obtain simple structure while keeping the directions at 90 degrees to each other (uncorrelated). This has the advantage of keeping the standardized factor loadings in correlational units. Oblique rotations allow for the factors to correlate but are still motivated to achieve simple structure of the loading matrix. In general, assuming that factors are uncorrelated is probably a bad, or at least a questionable, idea [55]. Therefore, the loadings after applying a rotation no longer represent the correlations between items and factors. Instead, they are standardized partial regression coefficients i.e., the unique effects of a factor on a variable.

The aim is to find a simple solution where each factor has a small number of large loadings and a large number of 0 (or small) loadings. Starting off from 3 factor model, we increased the number of factors to 4 and then 5 in order to obtain the simple factor solution. The goal is to find a model with a strong factor structure, i.e., minimum crossloading, high factor loading, and sum of squared loadings related to each factor. From our initial set of 36 models, we shortlisted eight for further evaluation. Out of these eight shortlisted models, the best model was selected based on highest factor loadings, minimal cross-loadings and other model fit metrics discussed in later sections. Figure 5.4 shows the solution with the best model fit: a five-factor model built with maximum likelihood extraction method and oblique rotation (promax) because of its strong factor structure and metrics (explained in subsection 5.4.1). Table 5.2 shows the eight models we compared, and the selected model is highlighted.

5.4 Interpretation of the Factors

Twenty of the 22 observed variables had significant loadings onto 5 factors (ML1, ML2, ML3, ML4, ML5). The factor analysis object obtained from R using psych package is show in Figure 5.3. The variables and respective factors are described in Figure 5.4. It is evident that the factors “Autotelic”, “Realism” and “Expressivity” have been loaded with variables that we theorized and refined in the previous stages of the study. Interpreting, grouping, and naming these factors were straightforward. However, “Immersion” and “Harmony” have additional properties loaded onto it along with its other three theorized properties. Upon further investigation, we were able to see how these new properties could be associated with the factors that it loaded on and were able to deem the model acceptable. Another interesting point was that the all the signs within “Harmony” were flipped and thus its correlation with other factors were negative. We could either flip all the signs and use it as is or we can re-name the factor to mean the opposite (e.g., Dissonance). As there is no strong evidence to choose one over the other, we decided to keep the factor as Harmony to be more consistent with the HX model [37], with the requirement that item and factor values need to be negated during measurement or interpretation.

5.4.1 Goodness of Fit Test

“How well does the hypothesised model fit?” This is a critical question in almost every application of Structural Equation Modeling (SEM) [71]. There are multiple metrics that can be used to assess and prove the goodness of fit of an EFA model. The object returned by the fa function from psych gives us a lot of information that we can use to diagnose and interpret the factor model. We can look at metrics like SS loadings which is the sum of squared loadings related to each factor. It is the overall variance explained in all the 20 variables by each factor. As a rule of thumb, the SS loading value of the factors should be

#	Model	RMSR	RMSEA	$\chi 2$/df	TLI	CFI
1	Extraction: Maximum Likelihood (ML): Factors: 4 Rotation: oblimin	0.0330	0.0606	1.9639	0.9240	0.9515
2	Extraction: Principal Axis Factoring (PAF): Factors: 4 Rotation: oblimin	0.0328	0.0611	1.9791	0.9227	0.9507
3	Extraction: Ordinary Least Squares (OLS) Factors: 4 Rotation: oblimin	0.0328	0.0611	1.9791	0.9227	0.9507
4	Extraction: Maximum Likelihood (ML): Factors: 5 Rotation: oblimin	0.0266	0.0499	1.6547	0.9482	0.9710
5	Extraction: Maximum Likelihood (ML): Factors: 5 Rotation: promax	0.0265	0.0490	1.6550	0.9480	0.9710
6	Extraction: Principal Axis Factoring (PAF): Factors: 5 Rotation: oblimin	0.0260	0.0508	1.6783	0.9463	0.9700
7	Extraction: Principal Axis Factoring (PAF): Factors: 5 Rotation: promax	0.0260	0.0508	1.6783	0.9463	0.9700
8	Extraction: Ordinary Least Squares (OLS): Factors: 5 Rotation: oblimin	0.0260	0.0508	1.6783	0.9463	0.9700

Table 5.2: Table showing different goodness of fit metrics to assess the fitness of the model to the data. The basic criteria for model fitness are TLI and CFI >0.90 & RMSEA and RMS values <0.05. The highlighted model has the highest TLI and CFI values and lowest RMSEA and RMS values.

```
##
## Loadings:
##                                  ML1    ML5    ML3    ML2    ML4
## Fits well(H1)                           -0.331
## Like as part of exp(H2)           0.805
## disconnected(H3)                          0.857
## appropriate(H4)                          -0.438
## out of place(H5)                          0.712
## All same(E1)                                          -0.493
## Adequate Variations(E2)                                0.799
## distinguishable(E3)                              0.622
## consistency(E4)                                        0.675
## Reflects variation(E5)                                 0.647
## Realistic(R1)                                                 0.826
## Believable(R2)                                                0.743
## Convincing(R3)                                                0.495
## Matched expectation(R4)
## Distraction(I1)                           0.561
## engagement(I2)                                   0.564
## focus(I3)                                        0.788
## Involvement(I4)                                  0.822
## Satisfying(A1)                    0.495
## Like by itself(A2)                0.704
## dislike(A3)                      -0.676  0.376
## prefer no haptics(A4)            -0.642
##
##                   ML1   ML5   ML3   ML2   ML4
## SS loadings     2.483 2.384 2.244 1.947 1.907
## Proportion Var  0.113 0.108 0.102 0.089 0.087
## Cumulative Var  0.113 0.221 0.323 0.412 0.498
```

Figure 5.3: The “loading” object resulting from factor analysis obtained using the “psych” package from R. The loading highlighted in red are discarded due to low score

Factor 1 & Loading size		Factor 2 & Loading size		Factor 3 & Loading size		Factor 4 & Loading size		Factor 5 & Loading size	
2.483		2.244		1.907		1.947		2.384	
The haptic feedback felt satisfying	.495	I felt engaged with the system due to the haptic feedback	0.564	The haptic feedback was realistic	.826	The haptic feedback all felt the same	-.493	The haptic feedback felt disconnected from the rest of the experience	.857
I like how the haptic feedback itself feels, regardless of its role in the system	.704	The haptic feedback helped me focus on the task	0.788	The haptic feedback was believable	.743	I felt adequate variations in the haptic feedback	.799	The haptic feedback felt appropriate when and where I felt it	-.438
I disliked the haptic feedback	-.676	The haptic feedback increased my involvement in the task	0.822	The haptic feedback was convincing	.495	The haptic feedback changes depending on how things change in the system	.675	The haptic feedback felt out of place	.712
I would prefer the system without the haptic feedback	-.642	The haptic feedback helps me distinguish what was going on	0.622			The haptic feedback reflects varying inputs and events	.647	The haptic feedback distracted me from the task	.561
I like having the haptic feedback as part of the experience	.805								
Autotelic		**Immersion**		**Realism**		**Expressivity**		**Harmony**	

Figure 5.4: The final items and their loadings on each factor of the best factor model. The higher the loading, the more the item is correlated with its factor. We named each factor after the closest theoretical construct from the HX model, as the factors

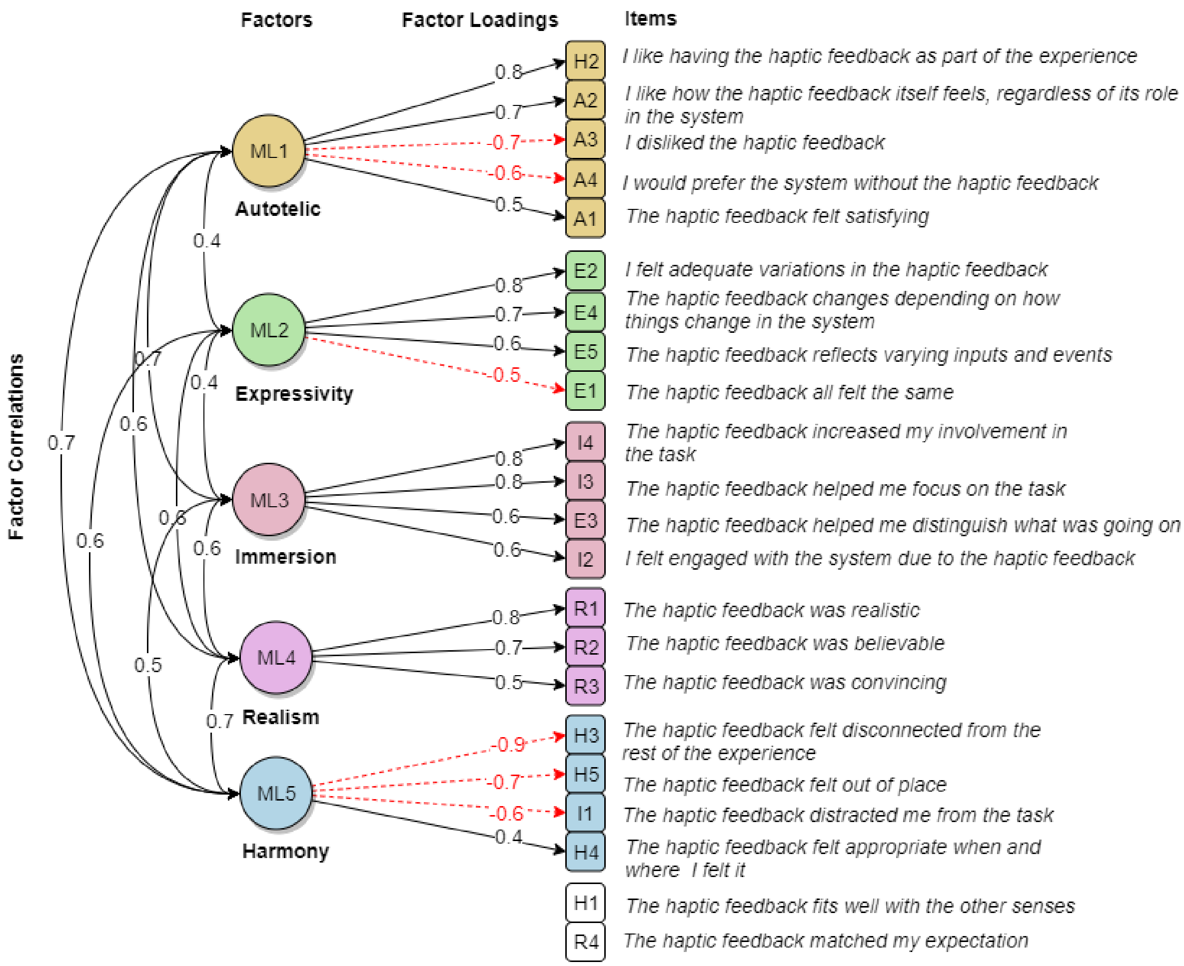


Figure 5.5: Path diagram representing the five oblique factors of the final model. The factors were transformed to an oblique solution using promax. Dotted red lines mean a negative loading due to the inverted item. Note that all signs for harmony have been flipped due to factor rotation.

Factor Correlations					
	ML1	**ML2**	**ML3**	**ML4**	**ML5**
ML1	1				
ML2	0.45	1			
ML3	0.66	0.41	1		
ML4	0.62	0.65	0.61	1	
ML5	0.69	0.59	0.5	0.66	1

Table 5.3: Factor Correlation matrix of the five-factor model

greater than 1. From Figure 5.4, we can see that our model has values ranging from 2 to 2.5 - indicating that these factors explain a large variance in its variables.

The magnitude of item-factor correlation ranges from 0.5 to 0.9 which is arguably an indication of a strong structure. The factors are evidently correlated with each other as expected. The correlations range from 16% (0.4*0.4*100) to 49% (0.7*0.7*100). High correlation between factors is not ideal but best practices dictate that the value should not exceed 0.7 [8].

Three model fit indices that are widely applied are considered in this study, all of which are based on a fit function given a specific estimation method. These are Root Mean Square Error of Approximation (RMSEA) [62, 61], Comparative Fit Index (CFI) [17] and Tucker–Lewis index (TLI) [64]. RMSEA is an absolute fit index, in that it assesses how far a hypothesized model is from a perfect model. On the contrary, CFI and TLI are incremental fit indices that compare the fit of a hypothesized model with that of a baseline model. The key advantage of TLI is the fact that it is not significantly affected from sample size. The application of RMSEA, CFI, and TLI is heavily contingent on a set of cut-off criteria. Earlier research [36] suggests that that an RMSEA value of <0.05 indicates a "close fit". It is recommended that TLI >0.90 indicates an acceptable fit [18]. And a CFI larger than 0.95 indicates relatively good model-data fit in general. From Table 5.2, we can see that the final model has an excellent model fit (RMSEA = 0.049, TLI = 0.948, CFI = 0.971, $\chi 2 = 1.655$).

5.5 Chapter Summary

The goal of EFA is to evaluate the dimensionality of a set of indicators in order to identify the smallest number of latent factors that explain the pattern of correlations. EFA tries to identify a simple model and is a data-driven exploratory technique with little intervention from the scientist in the estimation of the model. The primary decisions made are deciding how many factors to extract and how/whether to rotate the factors for interpretability. This chapter further elaborates the statistical process and methods used to extract the factors important in measuring haptic experience and explains how we arrive at the final structure. It also shows statistical evidence as to why the final model is an adequate estimation of the data. Refer Appendix D for the complete R code used for the analysis.

Chapter 6

Discussion

Our findings provide evidence and increased detail for the HX model, give concrete guidelines for measuring HX, and form a first draft of the HXI which can help researchers, designers, and hapticians measure the experiential factors of their haptic designs.

6.1 Support & Elaboration for the HX Model

The five-factor model obtained from exploratory factor analysis provides evidence to the previously proposed HX model [37]. Because the model with the best fit has five factors, each of which involves mostly questions from one HX model factor, it shows that participants were responding to our proposed items in a way that varies according to the five factors. The evidence of an internal structure provided by exploratory factor analysis confirms that the underlying factors of the HX model in fact measures haptic experience. This also supports our idea of measuring the haptic component of a system with a five-dimensional scale, rather than a simplified model with fewer dimensions

From a factor analysis perspective, factor loading is the correlation between the observed score and the latent score. Generally, it is better if this score is high since the square of factor loading can be directly translated as item reliability. We can see from Figure 5.5 that most of our factor loadings are between 0.5 to 0.8 with one factor loading at 0.44. This indicates strong item reliability. It is important to note that the negative sign on the factor loadings are that of "negatively-worded" items in the questionnaire and thus they are negatively correlated with the factor. In total, the extracted factors explain 54.7% of the variance.

We suspected correlation between the factors and so we used the oblique rotation for the factor model. As a rule of thumb, correlations between factors should not exceed 0.7 [8]. A correlation greater than 0.7 indicates a majority of shared variance (0.7 * 0.7 = % shared variance). As we can see from the factor correlation matrix Table 5.3, all the factors have a correlation of less than 0.7. This confirms discriminant validity. Figure 5.5 shows that the five factors are correlated with each other: the lowest correlation was 0.4 between Immersion - Expressivity and Expressivity - Autotelics, and the highest correlation was 0.7 between Harmony - Autotelics, Autotelics - Immersion and Harmony - Realism. These results lead to interpretations like "Feedback that feel good by itself help increase the immersion and realism of an application" given how highly they correlate. These findings help define what designers and hapticians must focus on in order to elevate their application design.

6.1.1 What each factor tells us

The scale development process has further enriched our understanding of each construct.

Autotelic: ML1 turned out to be the strongest factor with maximum influence on its variables. All variables of *Autotelic* (A1, A2, A3, A4) loaded significantly onto ML1 along with another variable from harmony (H2). The high factor loading of all the items is an indicator of how important it is to design a feedback that is likable and satisfying to the user. The item (H2) questions the satisfaction or desirability of the haptic feedback in the context of the system and has a very high factor loading of 0.805. While Autotelic by definition means that the feedback feels good in and by itself, the factor loadings indicate that the context of the system might significantly influence the likability of the feedback. Thus, we could understand that for an application to have a good autotelic experience, the design also needs to consider the context in which it is applied.

Expressivity: ML2 consisted of four variables designed for *Expressivity (E1, E2, E4, E5)*, but E3 ("The haptic feedback helped me distinguish what was going on") loaded on Immersion, suggesting that Expressivity has more to do with variance of feedback than causality, and that feedback of someone's actions is connected to Immersion. From the original definition that inspired the formulation of this construct "..Expressivity allows users to feel their input make an impact on the feedback received", we could verify that this item can potentially be measuring the expressivity of a system. This is not surprising, given that Immersion involves perceiving oneself to be interacting with an environment that has continuous stream of stimuli and experiences [70], but it is an important distinction for HX. This distinction helps us categorize the properties of immersion and expressivity with more clarity than before.

Immersion: All the positive items of *Immersion* (12, I3, I4) loaded significantly on ML3 thus confirming our existing understanding of the dimension. We can thus conclude that increased involvement, focus, and engagement is in fact an indicator of an immersive experience. Additionally, we can see that an item "The haptic feedback helped me distinguish what was going on" (E3) has been loaded onto this construct. What was originally thought of as an indicator of an expressive experience has now been shifted to be an attribute of Immersion. This factor overall lends support to the popular idea that when users are able to both affect and be affected by the system, they become more immersed in it [70].

Realism: Factor ML4 consisted only of variables for *Realism (R1, R2, R3)*, with the variable R4, "The haptic feedback matched my expectation" not loading significantly. This could simply be because the item is not capturing the construct. Our initial screening had some evidence that participants had some reservations about this question as it requires them to have expectations about the experience as a prerequisite. The non-significant loading of the factor gives additional evidence that this item may not be suitable to measure realism in an experience. Another interesting question raised during content validity discussion was about reliability vs believability. Some of the experts argued that believability might be a construct of its own and some agreed that it is mostly an aspect of realism. Since "The haptic feedback was believable" has a high factor loading of 0.743, it is safe to say that the factor "Realism" extracts sufficient variance from the variable "believability". Thus, believability might in fact be an underlying variable of the latent variable realism.

Harmony: ML5 was originally loaded with three negative items (H3, H5, I1) and one positive item (H4). All the negative items were positively correlated with the factor and the positive one was negatively correlated with it. Interestingly, the correlation between ML5 and all other factors was also a negative correlation. We were able to deduce that this factor was in fact the opposite of the construct *Harmony*. On further investigation and brainstorming sessions, we concluded that if we flip the signs, then we can group this factor as harmony. This suggests that harmony can indeed be the absence of a disruptive feedback. For example, if the haptic feedback does not feel "disconnected" from the system or "out of place" or "distracting", then we could accept that the feedback is harmonious with the system.

6.1.2 Customization & Generalization

While the study was done using commercially available haptic devices, which mostly fall under smartphone or gaming console categories, we have created the questionnaire with inputs from experts in different fields like robotics and medical sciences. This boosts the applicability of the scale to a wider range of applications. We even conducted the initial screening with smartwatches and fitness tracking devices and discovered that all our constructs may not be applicable to simpler applications such as feedback from a timer or a "step goal reached" alert. For example, the aforementioned feedback doesn't have a real-world equivalent thus questions around "believability", "realistic" and "convincing" seem irrelevant to the experience. This proposes that the HXI may need to be devised as sub scales that can be customised according to what the application desires its experience to contain.

The question of whether haptic experience can be measured is also partially answered by this study. By setting up a step-by-step guided task and providing a definition of what haptic feedback is, helped participants to complete the task and recognise the haptic feedback in them. They were aware of what they "felt" and easily connected it to "haptic sensation" and answered the questionnaire accordingly. Approximately 12% of the participants reported that they did not feel any haptic feedback in the guided task and their responses consisted only of neutral options, indicating that the HXI is in fact only applicable for applications with a prominent haptic feedback component.

What started as a five-dimensional scale with 5 items each has been refined to a five-dimensional scale with 3 to 5 items each by the studies conducted above. The first draft of the questionnaire contains 20 items that represent these five underlying factors and is ready to be put to further testing using other haptic modalities such as force feedback, mid-air haptic, electrotactile and conducting confirmatory factor analysis on the results.

6.2 Practical Matters for Measuring HX

Measuring experience is difficult. Focusing on a particular type of feedback to evaluate while keeping in mind the overall experience is even more challenging. In order to build our studies around measuring haptics, we had to provide participants with applications that are experienced as a system, but also have recognizable haptic feedback. For example, watching a haptic ad is very similar to watching any video with sounds and visuals, but has additional vibrations that notably catches user attention and elevates the experience when done right. We selected multi-sensory applications with visuals, sounds and vibrations and made sure the participants interacted with it for approximately 5 minutes. The approximate time was calculated based on the qualitative feedback obtained from cognitive interviews. This gave enough time for users to understand the application and the feedback, and fully experience the system.

One of the main debates we constantly had was whether a questionnaire is the right choice for evaluating HX. Through brainstorming and reviewing work around measurement instruments, we were convinced that a well-designed questionnaire can collect systematic data of user emotions and expectations about a system or design. We found participants typically interacted with haptic experience for five minutes, with the entire task and questionnaire taking 10 minutes. This suggests that participants can complete a 20-question instrument in less than five minutes, suitable for evaluating systems without being too onerous on participants.

Terminologies used in the questionnaire is another important area of concern. We initially referred to the unit measured as “feedback”, in order to keep it generic and simple. But this was ambiguous as participants were confused about what feedback we were referring to. We then decided to change it to “haptic feedback” based on the input from experts. However, this change could potentially make the questionnaire difficult to comprehend for users who did not know what “haptic” was. Statistics of our study shows that 57.7% of our participants have not heard of the word or unsure of what it means. Thus, we included a simple definition of haptic feedback to the HXI, which allowed non-expert participants to evaluate the haptics in their experience.

6.2.1 Definition of Haptic Feedback

The following sentence is a part of the HXI that explains what we mean by haptic feedback in our items:

> *Haptic feedback refers to anything that you feel with the sense of touch. It could be vibrations, force, temperature, pressure, or any other physical sensation.*

We found that defining haptic feedback was necessary to help users understand the items of the scale as reported by 82.9% of respondents.

6.3 Applications

Although the scale is yet to be validated using confirmatory studies, the HXI can still be used along with other measurement scales to understand user feedback. We have experimented this within other lab projects and the results seem beneficial to the researcher. One of the best applications for this version of the HXI could be to receive early feedback about mobile applications and game prototypes built with novel haptic effects. We also consider the HXI to be an effective tool to be used in the design process. It gives the developers a framework to focus their design on. For example, if the idea is to build an immersive experience, the developer needs to make sure that the haptic feedback is designed to help users "focus" on the task, increase user "involvement", enhance user "engagement" and help the user comprehend "what is going on". We found the draft HXI to be very efficient for mobile applications and games since majority of our data was collected based of these experiences and participants seem very comfortable with it As we are able, we hope to explore other, more varied haptic experiences with the HXI to see if it can evaluate longer experiences, or work for in-person studies with novel hardware.

6.4 Limitations

While there are promising results from the study, there are several limitations. Since game-like applications were used in the study to measure haptic feedback, we cannot be certain that other types of haptic applications will perform the same and produce similar significant factor structures. It is possible that users associate games with pleasant experiences and therefore lean towards giving higher scores. This could lead to a social desirability bias. The social desirability bias is considered to be a systematic error in self-reporting measures resulting from the desire of respondents to avoid embarrassment and project a favorable image to others [26]. Users tend to look for the right answer creating biased data which threatens the validity of the research. However, we could identify, test, and prevent it while validating the final results in future.

Moreover, it is not necessary that all applications consider these five dimensions as important factors in their design. We have yet to explore and evaluate the usage of HXI as sub-scale derivatives that can be customized according to the needs of designers and researchers.

The quantitative studies we conducted are cross-sectional studies as the survey was a snapshot of the population on which the data was collected. About 98% of the participants answered the HXI based mostly or entirely on the study session. Since we have made inferences about a population of interest at one point in time, we fail to take into account the good or bad effect time has on these experiences.

Due to the remote nature of the main study, we were not able to interact with our participants as much as we wanted to. While the messaging functionality offered by the crowd-sourcing tool helped resolve some of the issue and questions the participants had, a total control of what participants were doing during the task was lacking. We were only able to use commercially available devices with vibrotactile feedback, since finding remote participants with advanced setups (and controlling them) was very tedious. However, we believe our findings will translate to similar in-person studies, as prior work has found that crowd-sourced vibrotactile studies on commodity hardware yield similar findings to remote studies [60].

6.5 Future Work

Now that there is evidence that provides support for the HX model, confirmatory studies need to be conducted to support our results and to generalize it. Furthermore, the confirmatory studies can benefit greatly with the use of other haptic modalities like force, mid-air, and pressure. We may also need to lookout for other unknown dimension of haptic experience that these novel devices and applications bring to the table.

Other variables that can likely influence HX have been identified as part of this study. This data needs to be investigated and analysed as it has the potential to uncover new aspects of haptic user experiences to consider. For example, some users reported that the vibrations on their smartwatches "startled them" when they first started using it. But they got used to the sensation and did not notice it as much with regular usage. Another notable example is how some users prefer to "turn-off" haptic feedback on their smartphones as it is "irritating" or "annoying" with prolonged usage. Just like how measurement of UX over time has been gaining traction in fully understanding user experience [40], haptic experience over time can also be a major influence on the overall effects of haptic feedback in a system. A well-designed longitudinal study could help us record user responses to the HXI over different periods in time and can help gain the insights that we are looking for.

6.6 Chapter Summary

In this chapter, we discussed how the research has provided evidence for a reliable haptic evaluation instrument. We also touched upon how this work enhances our existing understanding of haptic experience and design. We further understood the current applications and limitations to the first draft of the HXI and concluded with the opportunities of future work.

Chapter 7

Conclusion

In this paper, we introduced the HXI as the first draft of the scale developed using the experiential dimensions of HX model. The HXI provides evidence for the underlying model and indicates that a five-dimensional scale is a viable tool to measure haptic experience in a multisensory vibrotactile application. The construction, validation and analysis of the model were carried out using five user studies with 348 total participants. The constructs in the HXI have been refined and validated and thus, this tool can readily be adopted by hapticians, researchers and developers of multisensory applications to evaluate their design through systematic user feedback.

Through this project, we explore why it is important to measure haptic experience, what the currently adopted tools and methods of measurements are and the need for a new standardised tool. We gained many valuable insights from literature review and user studies that has helped us understand haptic feedback and its role in multisensory applications. By interacting with haptic researchers and users, we were able to see how the field of haptics is expanding and influencing design choices for many digital applications. Introducing haptic feedback to communicate information is said to have highly positive reaction in users when done right. Through tools like the HXI, developers and researchers will have a way to understand the components of a good experience and be able to produce innovative applications that users will love.

We hope that the next time you create a prototype or an application which employs haptic feedback, you will be able to use the HXI and enjoy structured user feedback!

References

[1] A Practical Introduction to Factor Analysis: Exploratory Factor Analysis.

[2] Intro - Basic Exploratory Factor Analysis — Quantdev.

[3] Kaiser-Meyer-Olkin (KMO) Test for Sampling Adequacy - Statistics How To.

[4] LOCUS OF ATTENTION.

[5] Prolific — Online participant recruitment for surveys and market research.

[6] Prolific's Best Practice Guide – Prolific.

[7] Qualtrics XM // The Leading Experience Management Software.

[8] StatWiki.

[9] Handbook of psychology: Research methods in psychology, Vol. 2, 2nd ed., 2013.

[10] Enas Almanasreh, Rebekah Moles, and Timothy F. Chen. Evaluation of methods used for estimating content validity. *Research in Social and Administrative Pharmacy*, 15(2):214–221, 2 2019.

[11] Ayman Alzayat, Mark Hancock, and Miguel A. Nacenta. Quantitative measurement of tool embodiment for virtual reality input alternatives. In *Conference on Human Factors in Computing Systems - Proceedings*. Association for Computing Machinery, 5 2019.

[12] Mahdi Azmandian, Mark Hancock, Hrvoje Benko, Eyal Ofek, and Andrew D. Wilson. Haptic Retargeting. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, pages 1968–1979, 5 2016.

[13] 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 *Conference on Human Factors in Computing Systems - Proceedings*, 2016.

[14] E Babakus and W G Mangold. Adapting the SERVQUAL scale to hospital services: an empirical investigation. *Health services research*, 26(6), 1992.

[15] William O. Bearden, Richard G. Netemeyer, and Kelly L. Haws. *Handbook of marketing scales: Multi-item measures for marketing and consumer behavior research.* 2011.

[16] André Beauducel. The factor paradox: Common factors can be correlated with the variance not accounted for by the common factors! Technical report.

[17] P. M. Bentler. Comparative fit indexes in structural models. *Psychological Bulletin*, 1990.

[18] P. M. Bentler and Douglas G. Bonett. Significance tests and goodness of fit in the analysis of covariance structures. *Psychological Bulletin*, 1980.

[19] Christopher C Berger, Mar Gonzalez-Franco, Eyal Ofek, and Ken Hinckley. The uncanny valley of haptics. *Science Robotics*, 3(17):Art—-No, 2018.

[20] 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(June):1–18, 2018.

[21] 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, 7 2009.

[22] James Dean Brown. Choosing the Right Type of Rotation in PCA and EFA. Technical report.

[23] Lydia B. Chilton, Greg Little, Darren Edge, Daniel S. Weld, and James A. Landay. Cascade. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems - CHI '13*, page 1999, 2013.

[24] Jacob Cohen. A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20:37–46, 1960.

[25] Robert F DeVellis. Scale Development Theory and Applications (Fourth Edition Robert). *SAGE Publication*, 4:256, 2016.

[26] Robert J. Fisher. Social Desirability Bias and the Validity of Indirect Questioning. *Journal of Consumer Research*, 20(2):303, 9 1993.

[27] Marcello Giordano, John Sullivan, and Marcelo M. Wanderley. Design of Vibrotactile Feedback and Stimulation for Music Performance. pages 193–214. Springer, Cham, 2018.

[28] Jonathan Gutman. A Means-End Chain Model Based on Consumer Categorization Processes. *Journal of Marketing*, 46(2):60–72, 4 1982.

[29] Heather Haeger, Amber D. Lambert, Jillian Kinzie, and James Gieser. Cognitive Interviews to Improve Survey Instruments. *Annual Forum of the Association for Institutional Research*, (June):15, 2012.

[30] Abdelwahab Hamam, Mohamad Eid, Abdulmotaleb El Saddik, and Nicolas D. Georganas. A quality of experience model for haptic user interfaces. *Proceedings of the 2008 Ambi-Sys Workshop on Haptic User Interfaces in Ambient Media Systems, HAS 2008*, pages 1–6, 2008.

[31] Marc Hassenzahl. The Hedonic/Pragmatic Model of User Experience. In *Towards a UX Manifesto. COST294-MAUSE affiliated workshop*, 2007.

[32] Marc Hassenzahl, Michael Burmester, and Franz Koller. AttrakDiff: Ein Fragebogen zur Messung wahrgenommener hedonischer und pragmatischer Qualität. pages 187–196. 2003.

[33] Stephen N. Haynes, David C. S. Richard, and Edward S. Kubany. Content validity in psychological assessment: A functional approach to concepts and methods. *Psychological Assessment*, 7(3):238–247, 9 1995.

[34] Jeffrey Heer and Michael Bostock. Crowdsourcing graphical perception. In *Proceedings of the 28th international conference on Human factors in computing systems - CHI '10*, page 203, 2010.

[35] G. Douglas Jenkins and Thomas D. Taber. A Monte Carlo study of factors affecting three indices of composite scale reliability. *Journal of Applied Psychology*, 62(4), 1976.

[36] K. G. Jöreskog and D. Sörbom. *LISREL 8: Structural equation modeling with the SIMPLIS command language. Scientific Software International.* 1993.

[37] Erin Kim and Oliver Schneider. Defining Haptic Experience : Foundations for Understanding , Communicating , and Evaluating HX. pages 1–13, 2020.

[38] Aniket Kittur, Ed H. Chi, and Bongwon Suh. Crowdsourcing user studies with Mechanical Turk. In *Proceeding of the twenty-sixth annual CHI conference on Human factors in computing systems - CHI '08*, page 453, 2008.

[39] Paul Kline. *Handbook of Psychological Testing*. Routledge, 11 2013.

[40] Sari Kujala, Virpi Roto, Kaisa Väänänen-Vainio-Mattila, Evangelos Karapanos, and Arto Sinnelä. UX Curve: A method for evaluating long-term user experience. *Interacting with Computers*, 23(5):473–483, 2011.

[41] Ki Uk Kyung and Jun Young Lee. Ubi-Pen: A Haptic Interface with Texture and Vibrotactile Display. *IEEE Computer Graphics and Applications*, 29(1):56–64, 2009.

[42] Bettina Laugwitz, Theo Held, and Martin Schrepp. Construction and Evaluation of a User Experience Questionnaire. pages 63–76. 2008.

[43] 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 *The Annual Symposium on Computer-Human Interaction in Play Extended Abstracts - CHI PLAY '18*, pages 257–270, 2018.

[44] Jane Lessiter, Jonathan Freeman, Edmund Keogh, and Jules Davidoff. A cross-media presence questionnaire: The ITC-sense of presence inventory. *Presence: Teleoperators and Virtual Environments*, 10(3):282–297, 2001.

[45] Jane Lessiter, Jonathan Freeman, Edmund Keogh, Jules Davidoff, Vahid Zamanzadeh, Akram Ghahramanian, Maryam Rassouli, Abbas Abbaszadeh, Hamid Alavi-Majd, Ali-Reza Nikanfar, User Interface Design GmbH, Sari Kujala, Virpi Roto, Kaisa Väänänen-Vainio-Mattila, Evangelos Karapanos, and Arto Sinnelä. AttrakDiff. *Presence: Teleoperators and Virtual Environments*, 10(3):282–297, 6 2001.

[46] Robert W. Lissitz and Samuel B. Green. Effect of the number of scale points on reliability: A Monte Carlo approach. *Journal of Applied Psychology*, 60(1), 1975.

[47] Emanuela Maggioni, Erika Agostinelli, and Marianna Obrist. Measuring the added value of haptic feedback. *2017 9th International Conference on Quality of Multimedia Experience, QoMEX 2017*, pages 1–6, 2017.

[48] Kristen Miller. Cognitive Interviewing. *Question Evaluation Methods: Contributing to the Science of Data Quality*, pages 49–75, 2011.

[49] Michael Minge, Manfred Thüring, and Ingmar Wagner. Developing and Validating an English Version of the meCUE Questionnaire for Measuring User Experience. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 60(1):2063–2067, 9 2016.

[50] Fabiane F. R. Morgado, Juliana F. F. Meireles, Clara M. Neves, Ana C. S. Amaral, and Maria E. C. Ferreira. Scale development: ten main limitations and recommendations to improve future research practices. *Psicologia: Reflexão e Crítica*, 30(1):3, 1 2018.

[51] Jason Osborne, Jason W Osborne, Anna B Costello, and J Thomas Kellow. *Best Practices in Exploratory Factor Analysis*. Number January 2005. 2011.

[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, 12 2003.

[54] Mikki H. Phan, Joseph R. Keebler, and Barbara S. Chaparro. The Development and Validation of the Game User Experience Satisfaction Scale (GUESS). *Human Factors*, 58(8):1217–1247, 2016.

[55] Kristopher J. Preacher and Robert C. MacCallum. Repairing Tom Swift's Electric Factor Analysis Machine. *Understanding Statistics*, 2(1):13–43, 2 2003.

[56] Kristopher J. Preacher and Robert C. MacCallum. Repairing Tom Swift's Electric Factor Analysis Machine. *Understanding Statistics*, 2003.

[57] Karin Schermelleh-Engel, Helfried Moosbrugger, and Hans Müller. Evaluating the fit of structural equation models: Tests of significance and descriptive goodness-of-fit measures. *MPR-online*, 8(2):23–74, 2003.

[58] John A Schinka, Wayne F Velicer, and Irving B Weiner. *Handbook of psychology: Research methods in psychology, Vol. 2, 2nd ed.* 2012.

[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(August 2016):5–21, 2017.

[60] Oliver S Schneider, Hasti Seifi, Salma Kashani, Matthew Chun, and Karon E MacLean. HapTurk: Crowdsourcing Affective Ratings for Vibrotactile Icons. In *CHI '16*, pages 3248–3260, New York, New York, USA, 5 2016. ACM Press.

[61] J. H. Steiger and J. C. Lind. Statistically based tests for the number of common factors. *Annual Meeting of the Psychometric Society*, 1980.

[62] James H. Steiger. Structural Model Evaluation and Modification: An Interval Estimation Approach. *Multivariate Behavioral Research*, 1990.

[63] 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. pages 375–395. 2019.

[64] Ledyard R. Tucker and Charles Lewis. A reliability coefficient for maximum likelihood factor analysis. *Psychometrika*, 1973.

[65] User Interface Design GmbH. *AttrakDiff*. PhD thesis.

[66] Jan B.F. Van Erp, Hendrik A.H.C. Van Veen, Chris Jansen, and Trevor Dobbins. Waypoint Navigation with a Vibrotactile Waist Belt. *ACM Transactions on Applied Perception*, 2(2):106–117, 4 2005.

[67] Vero Vanden, Katta Spiel, Lennart Nacke, Daniel Johnson, and Kathrin Gerling. International Journal of Human-Computer Studies Development and validation of the player experience inventory : A scale to measure player experiences at the level of functional and psychosocial consequences. *Journal of Human Computer Studies*, 135(October 2019):102370, 2020.

[68] Robert J Vandenberg. Statistical and methodological myths and urban legends: Where, pray tell, did they get this idea? *Organizational Research Methods*, 9(2):194–201, 2006.

[69] Yon Visell. Tactile sensory substitution: Models for enaction in HCI. *Interacting with Computers*, 21(1-2):38–53, 1 2009.

[70] Bob G. Witmer and Michael J. Singer. Measuring Presence in Virtual Environments: A Presence Questionnaire. *Presence: Teleoperators and Virtual Environments*, 7(3):225–240, 6 1998.

[71] Yan Xia and Yanyun Yang. RMSEA, CFI, and TLI in structural equation modeling with ordered categorical data: The story they tell depends on the estimation methods. *Behavior Research Methods*, 51(1):409–428, 2 2019.

[72] Yuchun Zhou. A Mixed Methods Model of Scale Development and Validation Analysis. *Measurement*, 2019.

APPENDICES

Appendix A

Recruitment materials

Figure A.1: Participant recruitment poster for conducting Cognitive Interviews pages

University of Waterloo
Research Participant's Acknowledgement of
Receipt of Remuneration
and
Self-Declared Income

Section A: To be completed by Principal Investigator or designate

Principal/Faculty Investigator's Name: **Oliver Schneider**

Student Investigator(s)'s Name: **Suji Nivedita Sathiyamurthy**

Department: Management Sciences

Study Title: Evaluation of Haptics Experience

Section B: To be completed by research participant

In appreciation of my involvement as a research participant in the above study,
I acknowledge that I have received $ _10.00_ from the University of Waterloo.

I further acknowledge that:

- this amount received from the University of Waterloo is taxable;
- that it is my responsibility to report the amount received for income tax purposes; and
- the University of Waterloo will not issue a tax receipt for the amount received.

Participant's Name: ___________________________________

Participant's Signature: ___________________________________

Date: ___________________________________

Witness' Name __Suji Nivedita Sathiyamurthy________

Witness' Signature: ___________________________________

Date: ___________________________________

Form issued: December 20, 2011

Appendix B

Study 2: Content Validity

The content validity study was conducted in two parts. First, the expert hapticians were invited to a group discussion on each of the items and then asked to take a survey to rate the items on Relevancy and Clarity. Figure B.1 and Figure B.2 is a screenshot of the survey conducted on Qualtrics.

Please rate each question based on its relevance to Haptics Experience and its clarity

	Relevancy				Clarity			
	Not relevant	Need major revision	Need minor revision	Very relevant	Not clear	Need major revision	Need minor revision	Very clear
The feedback fits well with the other senses	○	○	○	○	○	○	○	○
I am glad I felt the feedback as expected	○	○	○	○	○	○	○	○
The feedback was expressive	○	○	○	○	○	○	○	○
The feedback was immersive	○	○	○	○	○	○	○	○
The feedback was realistic	○	○	○	○	○	○	○	○
The feedback matched my expectations	○	○	○	○	○	○	○	○
I felt engaged with the system	○	○	○	○	○	○	○	○
The feedback all felt the same	○	○	○	○	○	○	○	○
The feedback felt good	○	○	○	○	○	○	○	○
I like it more when the feedback is part of the experience	○	○	○	○	○	○	○	○
The feedback felt disconnected from the system	○	○	○	○	○	○	○	○
The feedback felt out of place	○	○	○	○	○	○	○	○
I felt adequate variation in the feedback	○	○	○	○	○	○	○	○
I felt detached from my environment because of the feedback	○	○	○	○	○	○	○	○
The experience was close to what I expected to feel based on my actions	○	○	○	○	○	○	○	○
I am satisfied by how convincing the feedback felt	○	○	○	○	○	○	○	○

Figure B.1: Questionnaire provided to the experts in study 2: Expert review for questionnaire content Validity (p1)

	Relevancy				Clarity			
	Not relevant	Need major revision	Need minor revision	Very relevant	Not clear	Need major revision	Need minor revision	Very clear
I was aware of my response to the feedback	○	○	○	○	○	○	○	○
The feedback helps me distinguish my actions	○	○	○	○	○	○	○	○
I disliked the feedback	○	○	○	○	○	○	○	○
The feedback felt appropriate given its context	○	○	○	○	○	○	○	○
I would prefer the system without the feedback	○	○	○	○	○	○	○	○
The feedback changed depending on my action	○	○	○	○	○	○	○	○
The feedback distracted me away from the system	○	○	○	○	○	○	○	○
The feedback was believable	○	○	○	○	○	○	○	○
The feedback would feel good by itself	○	○	○	○	○	○	○	○
I would like to disable the haptic feedback	○	○	○	○	○	○	○	○
The feedback changes with the way I interact with the system	○	○	○	○	○	○	○	○
The haptic feedback reflects varying inputs and events	○	○	○	○	○	○	○	○
The overall experience was authentic	○	○	○	○	○	○	○	○

Suggestions and other comments:

Powered by Qualtrics

Figure B.2: Questionnaire provided to the experts in study 2: Expert review for questionnaire content Validity (p2)

Appendix C

Main Study Materials

C.1 Qualtrics

The main study was designed and published on Qualtrics. A sample view of the study:

Welcome page

Welcome to the study: **Evaluation of Haptic Experiences**

The purpose of this study is to assess your experience with an application providing haptic (touch) feedback. You will be asked to complete a task (e.g. play a game on your game console or smartphone) using one of the devices listed below for approximately 5 minutes. An uploaded image of the device's screen will be requested to verify the completion of the task. You will then be directed to fill out a survey based on your experience with your device and assigned application. To continue with this study, you need to have access to a haptic enabled device(eg: Smartphone, Nintendo Switch, PS4). Haptic enabled devices **refer to anything that provide a feedback that you feel with the sense of touch. It could be vibrations, force, temperature, pressure, or any other physical sensations**

The entire process will take approximately 10 minutes.

Note: Your session will end in 44 minutes.

Please select the device with which you would like to complete the task:

Smartphones : 0 place(s) remaining
Gaming & other devices: 27 place(s) remaining

- ○ iPhone
- ○ Android Device
- ○ Nintendo Switch
- ○ PS4
- ○ My haptic enabled device is not listed here

Enter the name of your device
(eg: XBOX)

Please select the application you prefer to use for this study
All applications listed below are certified App Store applications. If you do not see a suitable application on the list, please select a different device

Ghost POP game

Please select the application you prefer to use for this study
If you do not see a suitable application on the list, please select a different device

Haptic Advertisements

Please select the application you prefer to use for this study
If you do not see a suitable application on the list, please select 'other'

Animal Crossing

Please enter the details of the application you prefer to use for this study
Eg: Playing Half-life Alyx on Vive

Please select the application you prefer to use for this study

If you do not see a suitable application on the list, please select 'other'

GTA 5

Enter the name of your game

Return submission

The maximum quota for using smartphones (both iPhone & Android) has been reached. Thank you for your interest

You have not completed this study. Please return to Prolific to return this submission. You can try to retake the study with a different device or contact the researcher if you are facing any technical issues. Thank you for your time.

Informed Consent

Consent for Participation in the Study

Title of the study: Evaluating User Experience of Haptic Interactions

Principal Investigator: Oliver Schneider, Assistant Professor, Department of Management Sciences, University of Waterloo, Canada, oliver.schneider@uwaterloo.ca

Student Investigator: Suji Nivedita Sathiyamurthy, Graduate Research Student, Department of Management Sciences, University of Waterloo, Canada, snsathiy@uwaterloo.ca

The purpose of the research study is to receive responses to the Haptic Experience Questionnaire, a scale currently in development intended to evaluate users' experience with interactive touch feedback systems. In order to be eligible, you must be between the ages of 18-64 years, have typical sensitivity to touch, and no medical hearth conditions. You will not be eligible to participate if you fall outside of the age range or, for safety purposes, have an atypical sensitivity to touch or known heart condition

In this study, you will be presented with a task in which you interact with a device rendering haptic (touch) feedback. You may be asked to upload screenshots of your selected device to verify your completion of the task. Finally, you will be asked to complete a survey based on your experience with the device. Your identity will be confidential.

You will be completing the study by an online survey operated by Qualtrics. When information is transmitted or stored on internet privacy cannot be guaranteed. There is always a risk your responses may be intercepted by a third party (e.g., government agencies, hackers). Qualtrics temporarily collects your contributor ID and computer IP address to avoid duplicate responses in the data-set but will not collect information that could identify you personally.

Any data collected will be stored for a minimum of three years on ECResearch, a secure network file server for the University of Waterloo Engineering researchers. The data will be anonymous and no names or personal information will be shared. Anonymous data may eventually be posted online for other researchers to use.

The entire process will take approximately 10 minutes. In appreciation of your time, you will receive a 1.25 (GBP) through Prolific. You may refuse or skip any task or question without affecting your remuneration. Your participation in this study is entirely voluntary and you may refuse to participate or withdraw from the study at any time up until the point at which your data is anonymized at the end of the study. You may refuse to participate or withdraw before this point without jeopardy. By indicating your consent, you are not waiving your legal rights or releasing the investigator(s) or involved institution(s) from their legal and professional responsibilities

You can contact the Student Investigator at Suji Sathiyamurthy <snsathiy@uwaterloo.ca> for any additional information or concerns.

This study has been reviewed and received ethics clearance through a University of Waterloo Research Ethics Committee (ORE #40872).

If you have questions for the Committee contact the Office of Research Ethics, at 1-519-888-4567 ext. 36005 or ore-ceo@uwaterloo.ca.

If you have questions about the study or related research, please contact Dr. Oliver Schneider at oliver.schneider@uwaterloo.ca.

I hereby CONSENT to participate in this study

- ○ Yes
- ○ No

Does not consent

As you do not wish to participate in this study, please return your submission on Prolific by selecting the 'Stop without completing' button

introduction block

Introduction

The present study aims to better understand users' experience with devices rendering haptic feedback. You will be asked to complete a task (i.e. game application) using a device you currently own and provide insight to your experience with a survey.
You will be asked to complete the following tasks:

1. Answer a demographic questionnaire.
2. Complete a guided task using your selected device.
3. Upload screen images to confirm task completion.
4. Fill out a survey based on your experience with the device.

The study should take no more than 10 minutes to complete. If you have arrived from Prolific, you will have 45 minutes until the session times out.

If you close this website you will not be able to continue from where you left off and will have to start from the beginning of the study. If you experience failure in attempting to complete the study, you will be sent an email for assistance. Please respond to the email to avoid conflicts.

Press Next to proceed.

Demographic Questionnaire

Let's start with a little bit about yourself!

None of the following questions can be used to identify you. Also, note that all of your answers will be kept strictly confidential and will only be used in aggregate

Which of the following age groups do you fall under?

○ 18 - 24 ○ 55 - 64

- 25 - 34
- 35 - 44
- 45 - 54
- 65 or older
- I prefer not to answer

What is your gender?

- Male
- Female
- Non-binary/third gender
- I prefer not to say
- ________ I prefer to self describe

What is your occupation?

____

How long have you been using your ${e://Field/choice}?

- Less than one month
- Two months to less than a year
- One year or more

How familiar are you with ${e://Field/app}?

- I have never tried it before
- I have briefly tried it a couple times
- I am very familiar

What is your experience with haptic technology?

- I am an expert haptician
- I have some experience
- I have heard of it but yet to learn more
- What is haptic technology?

Task block

These page timer metrics will not be displayed to the recipient.
First Click: *0 seconds*
Last Click: *0 seconds*
Page Submit: *0 seconds*
Click Count: *0 clicks*

Instructions page

Instructions to complete this task

Step 1: Download the app ${e://Field/app} from the ${e://Field/store}
Link: ${e://Field/app_link}

Step 2: Open ${e://Field/app} on your ${q://QID10/ChoiceGroup/SelectedChoices}

Step 3: Before beginning to play, please take a screenshot of your initial screen

Instructions page

Instructions to complete this task

Step 1: Open google chrome on your ${q://QID10/ChoiceGroup/SelectedChoices}

Step 2: Go to http://gleemlabs.haptter.com/website/haptic-video-marketing/haptic-videos-examples/
(you must see a few videos in this page)

Step 3: Please take a screenshot of this initial screen

Instructions page

Instructions to complete this task

Step 1: Open ${e://Field/app} on your ${q://QID10/ChoiceGroup/SelectedChoices}

Step 2: Let the game load and wait for start screen

Step 3: Please take a picture of the initial screen and your device

Instructions page

Instructions to complete this task

Step 1: Set up your task environment with ${q://QID68/ChoiceTextEntryValue}
and ${q://QID37/ChoiceTextEntryValue}

Step 2: Let the application load and wait for the start screen

Step 3: Please take a picture of the initial set up

Upload the image here
Accepted file types: JPG,PNG,GIF,PDF,MP4,MOV,QT,AVI,WMV,M4V

Step 4: Play the game or use the application for ~ 5 minutes (you can play longer if you wish but be mindful of the session end time)

Step 5: Once you are finished, please take another screenshot/picture to verify task progress or completion

Step 6: Upload the images in the fields below

Step 4: Watch the videos on this page (the first 3 videos at the least)

Step 5: Once you are finished, please take another screenshot to verify the completion or progress

Step 6: Upload the images in the fields below

Upload the image here
Accepted file types: JPG,PNG,GIF,PDF,MP4,MOV,QT,AVI,WMV,M4V

Note: You will not be able to come back to this page once you click next. Please check your work before proceeding.

What is haptics

You have completed the first half of this study! You will now be asked to complete a survey based on the experience you just had with your device. Please click Next to proceed to the survey.

HXI Block 1

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 satisfying	○	○	○	○	○
The haptic feedback distracted me from the task	○	○	○	○	○
The haptic feedback was realistic	○	○	○	○	○
The haptic feedback all felt the same	○	○	○	○	○
The haptic feedback fits well with the other senses (e.g. Sound, visuals)	○	○	○	○	○

HXI block 2

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
I felt engaged with the system due to the haptic feedback	○	○	○	○	○
I like having the haptic feedback as part of the experience	○	○	○	○	○
The haptic feedback was believable	○	○	○	○	○
I like how the haptic feedback itself feels, regardless of its role in the system	○	○	○	○	○
I felt adequate variations in the haptic feedback	○	○	○	○	○

HXI block 3

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 disconnected from the rest of the experience	○	○	○	○	○
I disliked the haptic feedback	○	○	○	○	○

	Strongly disagree	Somewhat disagree	Neither agree nor disagree	Somewhat agree	Strongly agree
The haptic feedback helps me distinguish what was going on	○	○	○	○	○
The haptic feedback was convincing	○	○	○	○	○
The haptic feedback helped me focus on the task	○	○	○	○	○

HXI block 4

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
I would prefer the system without the haptic feedback	○	○	○	○	○
The haptic feedback changes depending on how things change in the system	○	○	○	○	○
The haptic feedback increased my involvement in the task	○	○	○	○	○
The haptic feedback matched my expectation	○	○	○	○	○
The haptic feedback felt appropriate when and where I felt it	○	○	○	○	○

HXI block 5

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	○	○	○	○	○
The haptic feedback felt out of place	○	○	○	○	○

End Survey

We are almost done!

Please answer the following questions about the study

Approximately how long did it take to complete this study (in minutes)?

____

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

Your ${q://QID10/ChoiceGroup/SelectedChoices} ____

The application: ${e://Field/app}	
The haptic feedback in the application	

How noticeable was the haptic feedback?

- Not noticeable at all
- Somewhat noticeable
- Very noticeable

Did the definition of Haptic feedback help you understand the survey questions better?

- Yes
- Somewhat
- No

Has your understanding of haptic technology increased after this study?

- Yes
- Somewhat
- No

Were your survey responses based on your experience with the task/application during this session or your experience with the task/application prior to the session

- My responses were based entirely on this session
- My responses were based mostly on my experience during this session
- My responses were based equally on this session and my previous experience
- My responses were based mostly on my experience prior to this session
- My responses were based entirely on my experience prior to this session

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

- Yes
- Somewhat
- No

Any additional comments on the study

Feedback Letter

Thank you for participating in this study!

We appreciate your participation in our study and thank you for spending the time helping us with our research!

The purpose of the study was to gain insight into your experience interacting with a system that employed touch feedback using the Haptic Experience Questionnaire. Your responses will allow us to analyze the quality of the scale, thereby helping us to identify and address any necessary revisions. The final product will be a reliable and valid questionnaire ready to be deployed for use by haptic designers and researchers. The Haptic Experience Questionnaire will allow for the identification of a system's strengths and weaknesses, as well as comparisons between devices.

All information you provided is considered completely confidential; indeed, your name will not be included or in any other way associated, with the data collected in the study. Furthermore, because the interest of this study is in the average responses of the entire group of participants, you will not be identified individually in any way in any written reports of this research. Paper records of data and photos collected during this study will be retained for a minimum of three years on ECResearch, in a secure network file server for the University of Waterloo Engineering faculty members. Only researchers associated with this study will have access to the data. All identifying information will be removed from the records prior to storage.

This study has been reviewed and received ethics clearance through a University of Waterloo Research Ethics Committee (ORE #40872). If you have questions for the Committee, contact the Office of Research Ethics, at 1-519-888-4567 ext. 36005 or ore-ceo@uwaterloo.ca. If you have questions about the study or related research, please contact Dr. Oliver Schneider at oliver.schneider@uwaterloo.ca

We really appreciate your participation and hope that this has been an interesting experience for you!

References (related studies that may be of interest to you):

Kim, E., & Schneider, O. (2020). Defining Haptic Experience: Foundations for Understanding, Communicating, and Evaluating HX. Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems. doi:10.1145/3313831.3376280

Guest, S., Dessirier, J. M., Mehrabyan, A., McGlone, F., Essick, G., Gescheider, G., ... & Blot, K. (2011). The development and validation of sensory and emotional scales of touch perception. Attention, Perception, & Psychophysics, 73(2), 531-550

Powered by Qualtrics

C.2 Prolific

The participant pool was screened and recruited through prolific. A sample of how the study was published to the participant pool:

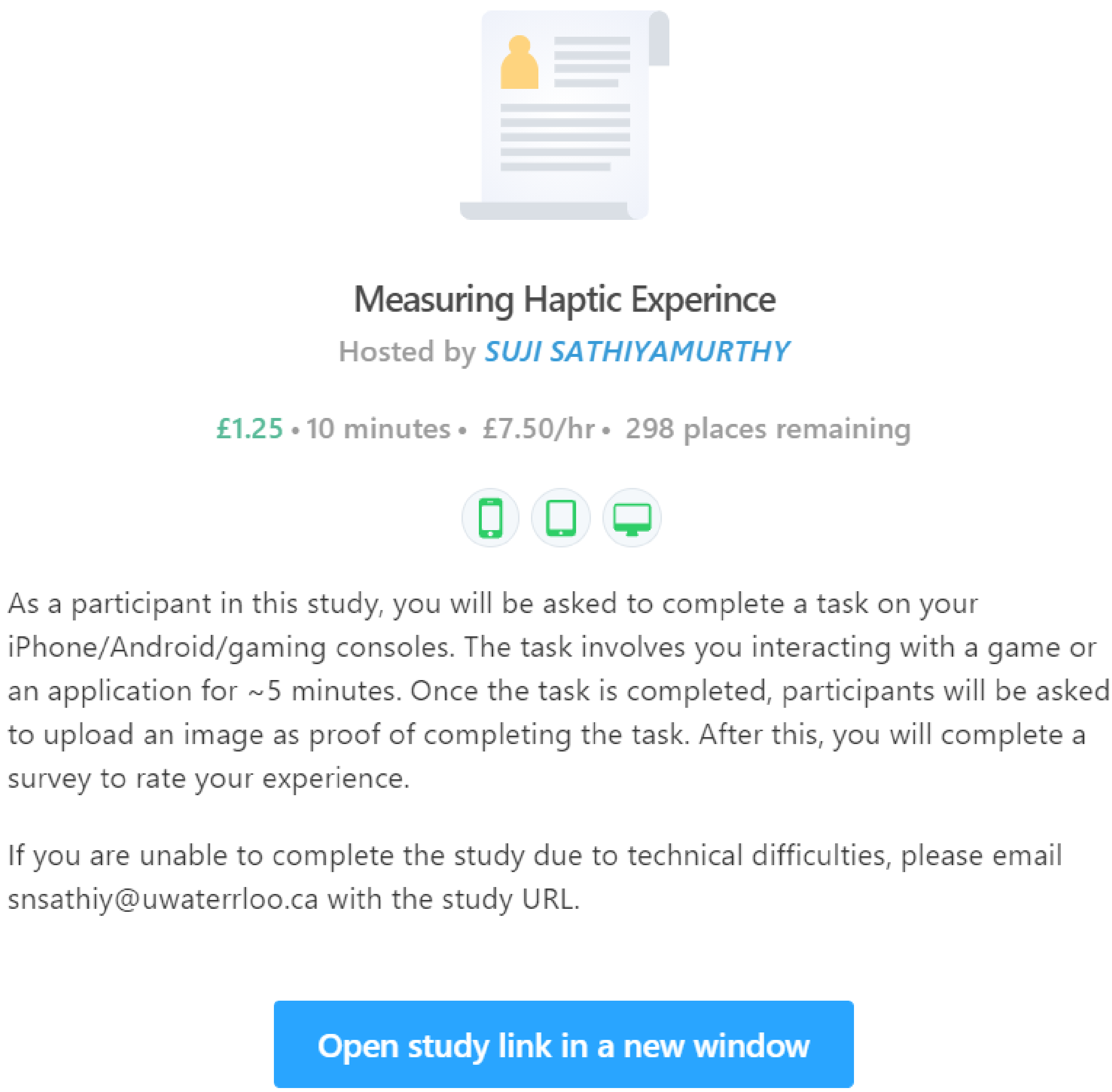


Figure C.1: Participant recruitment poster created on Prolific for the main study

Appendix D

Exploratory Factor Analysis R code

D.1 R Script

```
---
title: "Exploratory␣Factor␣Analysis"
author: "Suji␣Nivedita␣Sathiyamurthy"
date: "02/09/2020"
output:
  slidy_presentation: default
  powerpoint_presentation: default
  beamer_presentation: default
---

```{r setup, include=FALSE}
knitr::opts_chunk$set(echo = FALSE)
library(psych)
library(ggplot2)
library(corrplot) #plotting correlation matrices
library(GPArotation) #methods for factor rotation
library(nFactors)  #methods for determining the number of
factors
library(xtable)
```
```

```
## Procedure

- Data Cleaning
    - Assumptions:
      - Presence of latent variables underlying the measured variables
      - Correlation between factors
    - Missing Data

- Select Extraction Method
- Decide on how many factors
- Select Rotation method
- Interpret Results and Group factors
- Replication or evaluation of robustness


## Extraction Methods

- Maximum likelihood (MLAF) - Most Robust and
asymptotically significant
- Principal Axis factoring (PAF) - Recovers weak factors
- Ordinary Least Squares (OLS) - Recovers weak factors, works
well with TETRACHORIC/POLYCHORIC CORRELATIONS


## Rotations

- Orthogonal : Forces factors to be uncorrelated
    - Types: varimax, quartimax, bentlerT, equamax,
             varimin, geominT,bifactor
- Oblique : Factors can be correlated if it yeilds optimal solution
    - Types: Promax, oblimin, simplimax,
             bentlerQ, geominQ, biquartimin,cluster

## DATA
- Number of varibales: 22
- Number of Latent variables: 5
```

````
- Number of responses: 303
- Number of valid responses: 261
```{r , echo = TRUE,include=FALSE}
setwd("C:/Users/ssuji/Documents/Studies&Pilot/Main_Study/Analysis/")
HXIV2 <- read.csv(file="Main␣Study␣Prolific_HXI_V1_old.csv",
         sep=',',header=T)
HXIV1.2 <- HXIV2[,c(1,10,14,16,21,2,7,13,18,22,3,8,12,20,4,9,15,
           19,5,6,11,17)]
HXIV1.3 <- HXIV1.2

names_of_col<- c("Fits␣well(H1)",
"Like␣as␣part␣of␣experience(H2)",
"disconnected(H3)",
"appropriate(H4)",
"out␣of␣place(H5)",
"All␣same(E1)",
"Adequate␣Variations(E2)",
"distinguishable(E3)",
"consistency(E4)",
"Reflects␣variation(E5)",
"Realistic(R1)",
"Believable(R2)",
"Convincing(R3)",
"Matched␣expectation(R4)",
"Distraction(I1)",
"engagement(I2)",
"focus(I3)",
"Involvement(I4)",
"Satisfying(A1)",
"Like␣by␣itself(A2)",
"dislike(A3)",
"prefer␣no␣haptics(A4)")

names_of_col2<- c("H1","H2","H3","H4",
"H5","E1","E2","E3","E4","␣E5","R1","R2","R3","R4","I1",
"I2","I3","I4","A1","A2","A3","A4")
````
````

```
names(HXIV1.2) <- names_of_col2
```
```{r , echo = TRUE}
head(HXIV1.2)
describe(HXIV1.2)

#Data Adequacy

KMO(HXIV1.2)

```


Correlation Matrix
```{r , echo = TRUE, include=FALSE}
hxicor<-round(cor(HXIV1.2),2)
upper<-hxicor
upper[upper.tri(hxicor)]<-""
upper<-as.data.frame(upper)
#Hide lower triangle
lower<-hxicor
lower[lower.tri(hxicor, diag=TRUE)]<-""
lower<-as.data.frame(lower)
print(xtable(upper), type="latex")
```
```{r results = "asis"}
upper
```


Cor Plot
```{r }

corrplot(cor(HXIV1.2, use="complete.obs"),
order = "hclust", tl.col='black', tl.cex=.75)

```

```
## Scree Plot
```{r }
scree(hxicor,factors = FALSE)
```


Parallel Analysis
```{r , echo = TRUE}
# Maximum Likelihood
fa.parallel(HXIV1.2,fm='ml',fa='fa')
#Principal Axis
fa.parallel(HXIV1.2,fm='pa',fa='fa')
# Minimum residual
fa.parallel(HXIV1.2,fm='minres',fa='fa')

```

Maximum Likelihood
3 factor model
```{r , echo = TRUE}
Ml_3f=fa(HXIV1.2,nfactors = 3,rotate = "oblimin",fm="ml")
print(Ml_3f$loadings,cutoff = 0.33)
Ml_3f_2=fa(HXIV1.2,nfactors = 3,rotate = "promax",fm="ml")
print(Ml_3f_2$loadings,cutoff = 0.33)
```

4 factor model
```{r , echo = TRUE}
Ml_4f_1=fa(HXIV1.2,nfactors = 4,rotate = "none",fm="ml")
print(Ml_4f_1$loadings,cutoff = 0.33)
Ml_4f_2=fa(HXIV1.2,nfactors = 4,rotate = "oblimin",fm="ml")
print(Ml_4f_2$loadings,cutoff = 0.33)
Ml_4f_3=fa(HXIV1.2,nfactors = 4,rotate = "promax",fm="ml")
print(Ml_4f_3$loadings,cutoff = 0.33)
Ml_4f_4=fa(HXIV1.2,nfactors = 4,rotate = "varimax",fm="ml")
print(Ml_4f_4$loadings,cutoff = 0.33)
```
```

```
### 5 factor model
```{r , echo = TRUE}
Ml_5f_1=fa(HXIV1.2,nfactors = 5,rotate = "none",fm="ml")
print(Ml_5f_1$loadings,cutoff = 0.33)
Ml_5f_2=fa(HXIV1.2,nfactors = 5,rotate = "oblimin",fm="ml")
print(Ml_5f_2$loadings,cutoff = 0.33)
Ml_5f_3=fa(HXIV1.2,nfactors = 5,rotate = "promax",fm="ml")
print(Ml_5f_3$loadings,cutoff = 0.33)
Ml_5f_4=fa(HXIV1.2,nfactors = 5,rotate = "varimax",fm="ml")
print(Ml_5f_4$loadings,cutoff = 0.33)
```

Principal Axis Factoring
3 factor model
```{r , echo = TRUE}
Paf_3f_1=fa(HXIV1.2,nfactors = 3,rotate = "none",fm="pa")
print(Paf_3f_1$loadings,cutoff = 0.33)
Paf_3f_2=fa(HXIV1.2,nfactors = 3,rotate = "oblimin",fm="pa")
print(Paf_3f_2$loadings,cutoff = 0.33)
Paf_3f_3=fa(HXIV1.2,nfactors = 3,rotate = "promax",fm="pa")
print(Paf_3f_3$loadings,cutoff = 0.33)
Paf_3f_4=fa(HXIV1.2,nfactors = 3,rotate = "varimax",fm="pa")
print(Paf_3f_4$loadings,cutoff = 0.33)
```

4 factor model
```{r , echo = TRUE}
Paf_4f_1=fa(HXIV1.2,nfactors = 4,rotate = "none",fm="pa")
print(Paf_4f_1$loadings,cutoff = 0.33)
Paf_4f_2=fa(HXIV1.2,nfactors = 4,rotate = "oblimin",fm="pa")
print(Paf_4f_2$loadings,cutoff = 0.33)
Paf_4f_3=fa(HXIV1.2,nfactors = 4,rotate = "promax",fm="pa")
print(Paf_4f_3$loadings,cutoff = 0.33)
Paf_4f_4=fa(HXIV1.2,nfactors = 4,rotate = "varimax",fm="pa")
print(Paf_4f_4$loadings,cutoff = 0.33)
```
```

```
### 5 factor model
```{r , echo = TRUE}
Paf_5f_1=fa(HXIV1.2,nfactors = 5,rotate = "none",fm="pa")
print(Paf_5f_1$loadings,cutoff = 0.33)
Paf_5f_2=fa(HXIV1.2,nfactors = 5,rotate = "oblimin",fm="pa")
print(Paf_5f_2$loadings,cutoff = 0.33)
Paf_5f_3=fa(HXIV1.2,nfactors = 5,rotate = "promax",fm="pa")
print(Paf_5f_3$loadings,cutoff = 0.33)
Paf_5f_4=fa(HXIV1.2,nfactors = 5,rotate = "varimax",fm="pa")
print(Paf_5f_4$loadings,cutoff = 0.33)
```


Ordinary Least Squares

3 factor model
```{r , echo = TRUE}
minres_3f=fa(HXIV1.2,nfactors = 3,rotate = "promax",fm="minres")
print(minres_3f$loadings,cutoff = 0.33)
```

4 factor model
```{r , echo = TRUE}
minres_4f_1=fa(HXIV1.2,nfactors = 4,rotate = "none",fm="minres")
print(minres_4f_1$loadings,cutoff = 0.33)
minres_4f_2=fa(HXIV1.2,nfactors = 4,rotate = "oblimin",fm="minres")
print(minres_4f_2$loadings,cutoff = 0.33)
minres_4f_3=fa(HXIV1.2,nfactors = 4,rotate = "promax",fm="minres")
print(minres_4f_3$loadings,cutoff = 0.33)
minres_4f_4=fa(HXIV1.2,nfactors = 4,rotate = "varimax",fm="minres")
print(minres_4f_4$loadings,cutoff = 0.33)
```

###5 factor model
```{r , echo = TRUE}
minres_5f_1=fa(HXIV1.2,nfactors = 5,rotate = "none",fm="minres")
print(minres_5f_1$loadings,cutoff = 0.33)
minres_5f_2=fa(HXIV1.2,nfactors = 5,rotate = "oblimin",fm="minres")
```
```

```
print(minres_5f_2$loadings,cutoff = 0.33)
minres_5f_3=fa(HXIV1.2,nfactors = 5,rotate = "promax",fm="minres")
print(minres_5f_3$loadings,cutoff = 0.33)
minres_5f_4=fa(HXIV1.2,nfactors = 5,rotate = "varimax",fm="minres")
print(minres_5f_4$loadings,cutoff = 0.33)
# Checking what happens if we increase the factor further more
# SS loading shows factor 6 is not significant
minres_6f_4=fa(HXIV1.2,nfactors = 6,rotate = "promax",fm="minres")
print(minres_6f_4$loadings,cutoff = 0.33)
```

```
##Model Fit
- RMSR :Root mean square of residuals should be close to  0 (<0.05)
- RMSEA:Root mean square of Error Approximation should be <0.05
- TLI  :Tucker Lewis Index >0.90

All models have RMSR value of 0.03
All 3f models have RMSEA as ~0.07
All 4f models have RMSEA as ~0.06
All 5f models have RMSEA ~0.05

```{r , echo = TRUE}


diagram(Ml_5f_3)
plot(Ml_5f_3)

```
```

Appendix E

Study Scripts for User Studies

E.1 Face Validity Script

Face Validity of Questionnaire

Welcome the participants

Present the Power Point

State purpose of the gathering: Get feedback from the group for the initial draft of the questionnaire

Consent form: Hand over consent form and get it signed

Demo the example stimuli to add context: Play the ball game and let group members try it out if they want to.

Go over the HX model: Briefly explain the constructs

Display the questions: One by one go through the questions and discuss the following -

- Individual perception of the question
- Complexity of the questions
- What construct it is trying to measure
- Do the stimuli serve the purpose?

Record the feedback

Ask for any final thoughts/ additional feedback

- What are your impressions given the purpose of this scale?

Thank the participants for their time.

E.2 Content Validity Script

Content Validity - Expert Review Script

Pre-Session Set-up

Pre-session email package: Send the following materials using the email below

1. Doodle link to select time slot

Hi <NAME>,

> Thank you for agreeing to participate in this study. Please use the following link to choose your time slot. You will receive a confirmation mail with a survey link that you can use at the end of the session.

Send **confirmation mail** with the following:

1. Brief Instructions for the meeting
2. Confirmed time slot
3. Zoom / webex/ skype
4. Survey link

Audio Recording

- ☐ Initiate FlashBack Express Recorder
- ☐ Once session is completed, save as a WMV file and convert to audio file using: https://online-audio-converter.com/

Group Discussion

Introduction

Hello Everyone,

Thank you so much for participating in the study with me today.

I am a graduate student at the University of Waterloo, and I've been working with Dr.Oliver since the beginning of this year. Since then I've been working to develop a user experience questionnaire to evaluate haptic experiences. We want to create this tool to quantify haptic interactions, giving users a simple means to communicate about their experiences and allowing designers to easily identify the strength and weaknesses of the haptic portion of their system.

Before getting further into the session, I would like to mention that this session is being recorded and its mostly for internal use only. Some feedback obtained in this session maybe anonymized and quoted in our paper. Does that sound good? Stop me if you have any questions and concerns.

Thank you.

Self-Introductions

Before we get started, I was wondering if you could share a little bit about your background and experience with haptics

Goal of the Research

The goal of my research is to create an evaluation tool for haptics experience. Or a first cut draft of it at the least. Using an existing theoretical framework (Called the HX model – I will shortly be talking about it in detail), we have come up with a set of items to build a haptics experience questionnaire. The idea is to administer this questionnaire as a survey to a large number of participants after they complete a task that includes experiencing a haptics application. The end goal is to propose a statistically valid scale to measure Haptics experience and also prove the HX model.

Goal of this Study:

Now that we have generated a pool of items, we are in the process of refining its validity and reliability. We have already conducted one round of validity test where we conducted group discussion with participants from other research labs in the university. We have learnt a lot from it and improved our item pool. Similarly, we would like to repeat the same with experts like you and do some rigorous grooming to our questionnaire.

So essentially, I'm going to ask for your feedback, first through a brief semi-formal interview followed by a quick survey

Now I'm going to be sharing my deck. It has details on the HX model I mentioned earlier and the questions that we have come up with so far.

The HX Model:

<Display the slide with the HX model's pictorial representation>

The HEQ: Now, on to the questionnaire we have built based on this model. I would like to go around the table for each question and collect feedback and thoughts.

Harmony:

<Definition Slide>

<Questions Slide>

Notes:

Realism:

<Definition Slide>

<Questions Slide>

Autotelic:

<Definition Slide>

<Questions Slide>

Notes:

Immersion:

<Definition Slide>

<Questions Slide>

Notes:

Expressivity:

<Definition Slide>

<Questions Slide>

Notes:

Follow-up

Now that you have completed the review. There are some general questions I would like to get your opinion on.

- Does the order in which these questions are asked matter?
- Do we need to have the questions in sections? (eg. Harmony, Realism)
- Are the questions generic?
    - If so, does it sound vague
    - If not, should it be more general or customized to a experience/system.

The reason we began this project was because the field currently lacks a consistent evaluation tool.

- How would you use the scale we are creating?

Thanks, and Survey

I really appreciate all of you, for taking this time today and sharing your insights with us. We have got some valuable feedback that will help us make our questionnaire better. Additionally, we are also aiming to get a validity score for each of these items. So, I hope you can fill out a quick survey rating these questions you just saw, based on its relevance to haptics experience and the level of clarity with which its worded.

I have pinged the survey link in the chat. You can also find it in the confirmation email sent earlier.

One more question: I will be sending you an amazon gift card. So could you please let me know/post in the chat which email address you want me to send it to?

I can stay on the call while you get started if you need me around.

Thank you so much once again!

Wrap-up

- ☐ Send remuneration (via Amazon or Pay Pal, based on request) and Money Form
- ☐ Send Feedback Letter

E.3 Cognitive Interview Script

Cognitive Interview – Study Script

Requirements

of Participants: ~10 (+2 pilot participants)

Experiences:

- Fitbit:
  Task:
  - Set an alarm
  - set a step goal and achieve it!
- Android:
  Task:
  - Open Chrome
  - go to - http://gleemlabs.haptter.com/website/haptic-video-marketing/haptic-videos-examples/
  - play the first 2 videos
  (or)
  - Download this fidget spinner https://play.google.com/store/apps/details?id=com.zairong.fidgetspinner
  - Spin, stop and play with the digital fidget spinner
- IOS
  - Task:
  - Download iTranslator Converse
  - Skip subscription
  - Choose language
  - Tap & Hold – Speak
  (or)
  - Open Clock app
  - Go to Timer
  - Set a timer for 5 seconds
  - Feel the timer go off
- Game Console (ps4)

Pre-session set up:

~~Set up https://lookback.io/: create project and enter information. Generate link.~~

~~Share this link with participants - https://participate.lookback.io/3dypbn?live~~

Create survey on Qualtrics – Consent form + HXI

Share this link with participants

Consent form: https://uwaterloo.ca1.qualtrics.com/jfe/form/SV_3lRpU1zdcncJvyR

Email:

Hi <name>,

Thank you for agreeing to participate in this study. You will be using your (Android device/IOS/FITBIT/) to complete a task.

Here are some links that you will need during the study.

Task links:

http://gleemlabs.haptter.com/website/haptic-video-marketing/haptic-videos-examples/

or

https://play.google.com/store/apps/details?id=com.zairong.fidgetspinner

*Add notes to download/update necessary apps

~~Research Tool:~~

~~Lookback.io : https://participate.lookback.io/3dypbn?live~~

Survey Links:

Consent form : https://uwaterloo.ca1.qualtrics.com/jfe/form/SV_3lRpU1zdcncJvyR

Survey: https://uwaterloo.ca1.qualtrics.com/jfe/form/SV_0ecbS8aFfq6huLP

Introduction:

Hello <participant-name>

My name is Suji. Im a research student in University of Waterloo. I'm doing my master's in management science here. my research is based on evaluating user experience in haptic devices and a applications with haptic feedback

Thank you for joining this study with me today. This study is going to be 30-45 mins long and will require you to use <one-of-this-devices>.

- Fitbit
- Apple Watch
- Android phone
- IOS phone
- Ps4
- Laptop/desktop (for video call and surveys)

Before we begin, please take a few minutes to go through the consent form in the following link –

https://uwaterloo.ca1.qualtrics.com/jfe/form/SV_3lRpU1zdcncJvyR

if you are not able to find it you can check the webex chat, I have posted it there.

Feel free to ask me any questions you may have regarding this step.

The session will be recorded for internal purposes and may be anonymized and shared in our paper if need be. Does this sound good? If not, you may let me know!

Now that we have completed the initial procedure, let me give you a brief idea of what this study is going to look like.

The study will have 2 parts.

Part 1: you will complete a task (or play a game (in your case it is <game name>) using Fitbit/Apple watch/Android/IOS phone/ps4.

Part 2: This is basically a think-aloud interview. Based on the task completed you will be asked to "think aloud" and answer a set of survey questions. I will not be explaining any questions if they are unclear but will make sure to note your confusions down. I will explain more about this part while we get to it.

~~Now lets setup lookback~~

~~*Help/Guide with lookback setup*~~

Once setup, guide participant through one of the tasks described in the "Requirement" section

*After completion of the task, proceed to part 2"

Now you can open the survey link that I provided earlier.
https://uwaterloo.ca1.qualtrics.com/jfe/form/SV_0ecbS8aFfq6huLP

If you don't mind, can you share your screen so that we can go through the statements together

Hope you can see the "Definition of Haptics". Read it and try to recall(and match) the haptic feedback in the task you just performed. Because the statements that follow need to be answered based on it.

Let us go through every statement one by one.

I will the statement aloud.

You will explain your interpretation of the question; Choose the answer you see fit; Briefly explain why you went with that answer.

I would like to remind you again that I will not be clarifying any ambiguous statements because I I'm trying to avoid biased answers and understand how you interpret these statements (or misinterpret in some cases)

*insert probing questions like: "What does <TERM> mean to you?"

*insert probing questions like: "Can you repeat the question I just asked in your own words"

*insert probing question like: "How did you arrive at that answer? Was that easy or hard to answer? I noticed that you hesitated - tell me what you were thinking"

Please hit the submit for the survey. And that brings us to the end of the session.

I really appreciate your time. Thank you once again. I will be sending you an amazon gift card worth $10 for your time shortly.

Post session

Save and import recording to next cloud.

Send feedback letter

Send Remuneration

Send Remuneration acknowledgement form

Appendix:

Consent form:

Welcome to the research study!

We are interested in testing the validity of the Haptic Experience Inventory (HXI) we are building in the haptic lab at the University of Waterloo. For this study, you will be presented with information relevant to Haptics Experience and its applications. Then, you will be asked to answer some questions about it. Your responses will be kept completely confidential.

The study should take you around 30 minutes to complete. You will receive $10(CAD) for your participation. Your participation in this research is voluntary. The study will be recorded for internal purposes only. You have the right to withdraw at any point during the study. The Principal Investigator of this study can be contacted at Oliver Schneider <oliver.schneider@uwaterloo.ca>; Student Investigator at Suji Sathiyamurthy <snsathiy@uwaterloo.ca>.

By clicking the button below, you acknowledge:
Your participation in the study is voluntary.
You are 18 years of age. You agree to Audio and Video record this session for internal purposes.
You are aware that you may choose to terminate your participation at any time for any reason.