Factors of Haptic Experience across Multiple Haptic Modalities

Ahmed Anwar
ahmed.anwar@uwaterloo.ca
University of Waterloo
Waterloo, Ontario

Tianzheng Shi
stzz2001@gmail.com
University of Waterloo
Waterloo, Ontario

Oliver Schneider
oliver.schneider@uwaterloo.ca
University of Waterloo
Waterloo, Ontario

ABSTRACT

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

CCS CONCEPTS

• **Human-centered computing → HCI design and evaluation methods**.

KEYWORDS

Haptics, Scale Development, HCI, User experience design

ACM Reference Format:
Ahmed Anwar, Tianzheng Shi, and Oliver Schneider. 2023. Factors of Haptic Experience across Multiple Haptic Modalities. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI '23), April 23–28, 2023, Hamburg, Germany.* ACM, New York, NY, USA, 12 pages. https://doi.org/10.1145/3544548.3581514

1 INTRODUCTION

Haptic technology is increasingly used by designers with evidence that it enhances user experience (UX). For instance, mid-air haptic feedback has shown to make user experience more pleasant, creative, and predictable [32] while motion seats incorporating haptic feedback have shown to invoke better experience when measured by EEG and other psychological signals [39]. Similarly, there is evidence that haptic feedback in virtual environment (VR) can lead to increased presence [2]. However, despite the promising adoption of haptic technology, it is difficult to understand how it influences UX, an important consideration for designers to improve their designs.

Haptic designers currently use qualitative methods to understand the influence of haptic feedback on their designs. Schneider et al.'s exploration of haptic experience design found that haptic designers prefer small in-person tests to evaluate their designs, iterating until it just feels right [44]. However, this approach is time-consuming, costly and not scalable to larger and remote evaluative studies. Although some haptic designers have made use of general scales such as AttrakDiff [12], there is no formal evaluative tool that measures the unique constructs of haptic experience. Therefore, it is pertinent to develop a reliable and scalable instrument that lets haptic designers identify design parameters that require improvement.

In this paper, we report on the development of a novel scale to measure haptic experience (HX), a five-dimensional model proposed by Kim and Schneider [28]. These five dimensions (Autotelics, Harmony, Immersion, Expressivity, Realism) serve as guiding principles for designing our proposed instrument. While Sathiyamurthy et al. [43] explored the HX model with a remote study incorporating vibrotactile feedback only, we explore the model with an in-person study incorporating different haptic feedback (vibrotactile, mid-air, force feedback, surface-haptics). Our aim is to support the development of a measurement instrument that can be used with any type of haptic device, and guide development of novel devices or experiences that work with multiple devices.

We begin by discussing related scales and existing instruments that encapsulate some of the constructs of HX. Next, we outline the process for defining the scale constructs and the corresponding items. Following this, we describe our in-person study design incorporating different haptic feedback (vibrotactile, force-feedback, mid-air, surface haptics) and report the results of our exploratory factor analysis (EFA) and confirmatory factor analysis (CFA). Based on the results, we propose an 11-item model of HX with four experiential dimensions of “Harmony", “Expressivity", “Involvement" and “Realism". Although the obtained model was confirmed using CFA, it needs further validation for use in practice. Our findings contribute:

(1) Evidence of a four factor HX model built using different haptic modalities,
(2) Comparison of vibrotactile model with multi-modal model, and
(3) Guidelines for creating evaluative instruments to measure haptic experiences.

CHI '23, April 23–28, 2023, Hamburg, Germany
https://doi.org/10.1145/3544548.3581514

2 RELATED WORK

2.1 Haptics and UX

User experience (UX) evaluation can be categorized with respect to pragmatic and hedonic qualities. Pragmatic quality refers to the effectiveness and efficiency by which the user is able to achieve their goals, while hedonic quality means the non-pragmatic quality aspects of the product such as enjoyment. The UEQ [30] had been used in previous studies to evaluate UX based on three pragmatic factors (Perspicuity, Efficiency and Dependability) and two hedonic factors (Novelty and Stimulation). Similarly, AttrakDiff2 [25] presents a breakdown of hedonic and pragmatic factors for UX evaluation. Another evaluation framework called meCUE questionnaire [34] evaluates UX based on instrumental and non-instrumental product qualities with a focus on a user's emotional response.

However, existing UX scales are insufficient to measure the unique constructs of HX. HX is focused on the sense of touch, breaking down potential feedback into relevant components; HX is also highly dependent on the context of the interaction and feedback from other modalities such as visual and auditory. Existing UX models evaluate the product as a whole and are not able to measure haptic feedback in isolation, important to support the intentional design of touch feedback. Furthermore, hapticians report a need for better evaluative tools that are standardized [28]. This presents a need for a validated evaluative instrument that is able to quantitatively measure the unique constructs of HX at scale.

2.2 Scale development in HCI

Scale development is often used in HCI studies not just used to produce a quantitative evaluative instrument, but to provide insight into the construct of interest being studied. For instance, Baumgartner et al. developed a pictorial multi-item scale, called PSUS (Pictorial System Usability Scale), which aims to measure the perceived usability of mobile devices [3]. Bentvelzen et al. developed a scale called the Technology-Supported Reflection Inventory (TSRI) that evaluates how effectively a system supports reflection [6]. TSRI enables researchers and practitioners to compare prototypes designed to support reflection. Brühlmann et al. developed the User Motivation Inventory (UMI) to measure a user's motivation to engage with an interactive system [11]. Suh et al. created the User Burden Scale (UBS) to measure the level of burden faced by a user while interacting with a computing system [47]. Similarly, Votipka et al. created and validated secure software development self-efficacy (SSD-SES) scale to measure software developers' belief in their ability to perform vulnerability identification and mitigation as well as security communication tasks [53].

These scales were developed through a systematic scale development process. Boateng et al. [7] outlined the best practices for scale development which involves three stages: 1) item development, 2) scale development, and 3) scale evaluation. 1) The item development step involves generating items for the intended scale from the theoretical construct. 2) The scale development step includes administering the generated questions and extracting a model using Exploratory Factor Analysis (EFA). EFA aims to explore the underlying theoretical structure obtained from the collected response [7]. It does so by clustering similar variables into similar factors to identify underlying latent constructs using the correlation matrix of items. In addition, it also reduces the number of items to a smaller subset to achieve a structure that withstands confirmatory factor and reliability checks.

The final step, 3) scale evaluation, involves testing the extracted model structure using Confirmatory Factor Analysis (CFA) and validity studies. While EFA is used to extract the underlying factor structure, it does not measure the model's goodness-of-fit with respect to new data [45]. Hence, to verify the model structure, CFA is used. CFA is a dimensionality test in which the hypothesized model obtained from EFA is tested at a different time point in a longitudinal study or a new sample. In our work, we build on the previous item development work from Kim and Schneider [28] and Sathiyamurthy et al. [43]. We use a study design similar to Sathiyamurthy et al. [43], who had also conducted scale development and EFA, but only for vibrotactile devices, and did not continue on to 3) scale evaluation.

2.3 Haptics and related scales

There are existing scales in literature that have partial overlap with the experiential dimensions of HX. These are the Need for Touch Scale (NFT), Presence Questionnaire (PQ), Immersive Tendencies Questionnaire (ITQ) and Haptic Fidelity Framework. The (NFT) scale [40] is designed to measure the user's need of obtaining product information by haptic feedback with respect to two dimensions - Instrumental factors and Autotelic factors. Instrumental factors utilize haptics to reflect products' textural properties and measure purpose-driven evaluations based on consumer preferences and the product using haptics. In contrast, Autotelic factors are not purpose driven but mainly for hedonic purposes like enjoyment or sensory stimulation of touching a product. NFT is user-centered and focuses only on measuring the user's desire to touch, not the quality of the HX provided through a system.

Haptic feedback has shown to increase presence in virtual environments [2] and two instruments exist to measure presence - Immersive Tendencies Questionnaire (ITQ) [54] and Presence Questionnaire (PQ) [54]. The ITQ measures individuals' inclination to feel presence in a virtual environment but is not specific to any haptic technology. The PQ measures four factors (control, sensory, distraction and realism) aberrant to the virtual environment that influence the user's immersion in it. The PQ overlaps with two of the experiential dimensions from the HX model, i.e., Immersion and Realism but does not cover Autotelics, Harmony or Expressivity.

Additionally, Muender et al. developed the Haptic Fidelity Framework to define the factors encapsulating realistic haptic feedback for virtual reality [35]. The framework constituted two dimensions: Haptic Fidelity and Versatility. This framework is more of a way for expert categorization and rating of haptic devices, rather than a way to understand the HX of a user. Moreover, this framework is not generalizable to all types of haptic modalities (e.g. force feedback, vibrotactile, mid-air, surface haptics) as it was constructed specifically for haptic feedback in virtual environments (VEs).

2.4 Haptics and Gaming Scales

Haptic feedback is utilized in gaming applications to enhance the user experience and there are many instruments used to measure the game user experience. The Game Engagement Questionnaire (GEQ) was developed to measure the impact of playing games based on engagement but it has not been validated [31]. Similarly, the Player Traits Model (PTM) [51] measures gaming experience based on player traits using five dimensions - challenge orientation, goal orientation, aesthetic orientation, narrative orientation and social orientation. However, this scale is user-centered and measures user preferences for game design but does not assess the HX vis-à-vis the game quality. The Player Experience Inventory (PXI) [1]measures player experience with respect to functional consequences (audiovisual appeal, progress feedback, ease of control, challenge and goals) and psychological consequences (mastery, curiosity, immersion, autonomy, meaning). The PXI has been used to measure the added value of vibrotactile feedback in games [46]. However, this scale also does not measure any constructs of HX as it is aimed for understanding and improving game design parameters based on the user’s game experiences.

3 ITEM DEVELOPMENT

The first step in scale construction is domain identification which is to specify the boundaries of the domain [7]. For this, we leveraged the prior work done by Kim and Schneider [28] in which the researchers outlined five experiential factors (Harmony, Autotelics, Expressivity, Immersion, Realism) and defined 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 multi-sensory experience.”*

The next step involves item generation using the five extracted experiential factors. This step was done in the work of Sathiyamurthy et al. [43], in which the researchers generated 22 questions through establishing face validity (N=8), content validity (N=6), cognitive interviews (N=9) and a pilot study (N=25). These questions and their corresponding experiential dimension are presented in Table 1. While Sathiyamurthy et al. [43] administered these 22 items in user studies involving only vibrotactile haptic feedback, we aim to incorporate multiple haptic modalities in our user study. Using these 22 items we continue the scale development process by survey administration, performing EFA and CFA.

4 STUDY DESIGN

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

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

4.1 Sample Size

Sample sizes for Exploratory Factor Analysis (EFA) in scale development studies can vary depending on the source. Some researchers suggest basing the sample size on the number of items [9, 14, 20] in the preliminary scale with a minimum sample size between 100 and 200 [16, 17, 20]. For instance, Nunnally [37] recommended to have at least 10 respondents per scale item i.e. a ratio of 10:1 for respondents to items. However, Hair et al. [24] recommended having 5 participants for each scale items. On the other hand, some researchers suggest that the sample size should be independent of the number of survey items. For instance, Clark and Watson [15] recommend using a sample size of 300 whereas Guadagnoli and Velicer [22] recommend a sample size of 200-300. Comrey and Lee [17] weighted the sample size (100 = poor, 200 = fair, 300 = good, 500 = very good, ≤ 1000 = excellent).

Additionally, we also needed to take into account the amount of data required for Confirmatory Factor Analysis (CFA). Previous research studies have recommended a sample size of around 100 for CFA [8]. Bentvelzen et al. used a sample size of 507 for EFA and 498 for CFA [6] whereas Votipka et al, used a sample size of 157 for EFA and 162 for CFA. Taking all this together, we targeted N=300 for EFA and N=100 for CFA.

4.2 Procedure

Each participant interacted with a single haptic device for approximately 5 minutes and was asked to complete a small predefined task on it. Figure 1 represents how participants interacted with the devices. Following that, they completed the 22-item questionnaire on a 5-point Likert scale. As remuneration, they were given $2 voucher for a coffee shop. To ensure maximum responses, an option to complete the survey on a smartphone was also included. Participants completed the following steps in an online questionnaire:

- Review and give consent to the study
- Select haptic device from a drop-down menu
- Answer demographic questions
- Complete the task/application using the chosen device
- Complete the questionnaire (22 items in randomized order on a 5-point Likert scale)
- Answer exit survey (rate experience)

---

[1] https://2diy.haply.co/

[2] https://www.ultraleap.com/product/stratos-explore/

[3] https://store.facebook.com/ca/quest/products/quest-2

[4] https://tanvas.co/products/tanvastouch-dev-kit

[5] https://www.3dsystems.com/haptics-devices/touch

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

Table 1: Initial 22 items generated by Sathiyamurthy et al. [43] after content validity for the five dimensional HX model. Here, they are organized by intended construct, denoted by prefix: A=Autotelics, E=Expressivity, I=Immersion, R=Realism, H=Harmony. We conducted a similar study using the same developed items, but using an in-person study with several different haptic modalities.

Figure 1: Devices used in the study

4.3 Data Cleaning

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

4.4 Sample Description

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

5 EXPLORATORY FACTOR ANALYSIS

EFA aims to explore the underlying theoretical structure obtained from the collected responses. It does so by clustering similar items into the same factor to identify underlying latent constructs using the correlation matrix. In addition, it also reduces the number of items to a smaller subset to achieve a structure that withstands

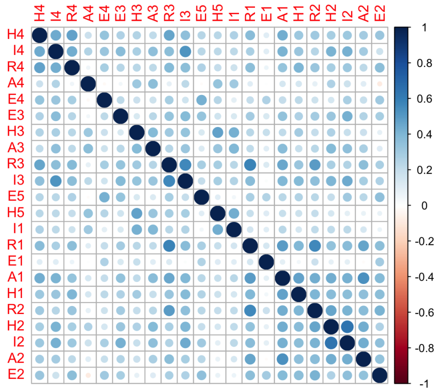

Figure 2: Correlation matrix for initial 22 items with N=291. The presence of "clusters" indicates the presence of factors.

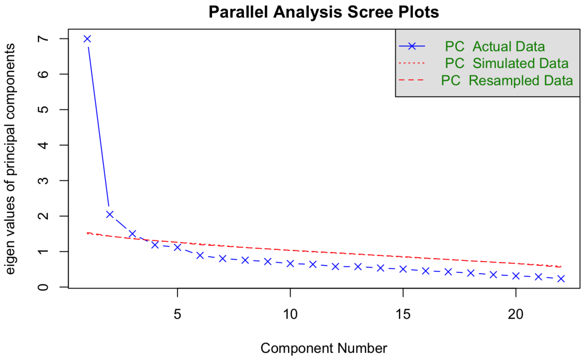

Figure 3: Scree plot with decreasing eigenvalues with respect to the number of possible factors. A steep drop at the 3rd component and a more subtle drop at the 5th component indicates that there could be between 3-5 underlying factors in the 22 items

confirmatory factor analysis and reliability checks. There are three main steps involved in EFA. These are:

- Assessment of the suitability of the data for factor analysis,
- Factor extraction by determining extraction method and rotation type, and
- Item refinement and reliability analysis.

5.1 Assessment of the suitability of the data

To determine the suitability of collected data for factor analysis, we analyzed our data based on three different criteria. First, we determined the strength of relationship among the items. In general, there must be presence of correlation coefficients > .30 in a correlation matrix of all items for adequate factor analysis. The correlation matrix in Figure 2 show significant items with inter-item correlations above 0.30. More specifically, 105 item pairs had inter-item correlation > 0.30. This means that there is evidence of correlation in our data which makes it suitable for factor analysis.

Next, we performed Kaiser-Meyer-Olkin (KMO) test which is a metric to measure the adequacy of the data for factor analysis. It does so by calculating the proportion of variance among variables that might become variance. In general, lower proportion of inter-item variance makes the data more suitable for EFA. The KMO value ranges from 0 to 1 and is ranked such that a value from 0.8 to 1.0 is considered excellent, 0.7 to 0.79 is considered adequate, 0.6 to 0.69 is considered mediocre. KMO values less than 0.6 indicate that the sample is not appropriate for factor analysis [23, 27, 48]. Our test results (KMO = .88) indicated adequate factorability.

Lastly, we applied the Bartlett's test of Sphericity [50]. It is used to test the hypothesis that the variables are not orthogonal i.e. the variables are sufficiently correlated that the correlation matrix is significantly different from an identity matrix). Based on our results ($\chi^2 = 2335, p < 0.05$), we conclude that the items are sufficiently correlated.

5.2 Factor Extraction

After establishing sampling adequacy, we determined the ideal number of underlying factors using the scree plot shown in Figure 3. The plot starts to level off at around the 3rd component and again at the 5th component which indicates the presence of 3-5 factors in the 22 items. Though the HX model constitutes of five factors, we also iteratively experimented with 3-factor, 4-factor, and 5-factor models. We compared the models based on their conceptual interpretability and number of items per factor and finalized the 4-factor model.

Next, we determined the factor extraction method and factor rotation type. We tested our data for multivariate normality using Mardia Tests and the results of both tests ($\chi_s^2 = 4861.42, p < 0.01; Z_k = 34.39, p < 0.01$) indicated that data is non-normal. Hence, "Principal Axis Factoring" was used for factor extraction method. To determine which item loaded on which factor, we needed to specify the rotation type. There are two rotation types in general i.e. oblique and orthogonal. Oblique rotations are appropriate when the factors are expected to be correlated and orthogonal is used otherwise [20]. We found that the factors are correlated (inter-factor correlation scores ranging from 0.3 to 0.5) which makes oblique rotation appropriate. Within oblique rotation, we experimented with 'promax', 'oblimin' and 'simplimax' and achieved the best results with 'promax' rotation.

5.3 Item Refinement and Reliability

After performing factor analysis, we refined our item set based on two inclusion criteria. Firstly, we included an item only if its loading was above the recommended threshold of 0.30 [7]. This ensures that the extracted item has a significant association with the underlying factor [36]. This resulted in removing items H4, I4, R4, E3 and E2. Secondly, we considered an item if it did not have significant cross loading i.e. the item did not load significantly on other factors. This step affirms that variances in items are uniquely associated to their respective factor only and not to other factors. As a result items, I3, A1, H1, A2, A3 and A4 were removed.

These 11 items were then subjected to a second analysis. The results of Kaiser-Meyer-Olkin (KMO) test (overall KMO = 0.77, none below 0.70) and Batlett's test ($\chi^2 = 884, p < 0.01, df = 55$) indicated adequate factoribility of the remaining items. The inspection of the scree plot suggested the presence of four factors and the results of Mardia tests ($\chi_s^2 = 1293.27, p < 0.01; Z_k = 24.5, p < 0.01$) indicated that the remaining 11-item data is non-normal. Using principal axis factoring and Oblimin (Promax) rotation, we obtained a final refined model as shown in Figure 4. All the items loaded significantly on their corresponding factors with no substantial cross-loading. The extracted items and their correlations are shown in Table 2.

To confirm that our 11 extracted items maintained their reliability (internal consistency), we calculated Cronbach's α. A scale is considered reliable if the overall α exceeds 0.6 and the majority of sub-scale αs exceed 0.7 [33]. Our test results (overall $\alpha = 0.78$, majority subscale's $\alpha > 0.70$) indicate adequate reliability of the refined model, meaning that the results are expected to be consistent when repeated under identical conditions [41].

6 CONFIRMATORY FACTOR ANALYSIS

Next, we verified the structure of the extracted model using Confirmatory Factor Analysis (CFA) with our testing set (N=112). While EFA is used to extract the underlying factor structure, it does not measure the model's goodness-of-fit with respect to new data [45]. Hence, to verify the model structure, CFA is used. CFA is a dimensionality test in which the hypothesized model obtained from EFA is tested at a different time point in a longitudinal study or a new sample [10].

6.1 Goodness of Fit

Our hypothesized model demonstrated adequate goodness-of-fit with its ($\chi^2 = 47.07$) below the conservative limit of double the degrees of freedom (DoF = 38) [13]. In addition, using ANOVA, our theoretical model demonstrated better fit than the null model ($\chi^2 = 371.42, p < 0.001$).

We also computed other goodness-of-fit metrics. Firstly, we calculated the Comparative Fit Index (CFI), which measures the model's fit with respect to a more restrictive baseline model [4], and the Tucker-Lewis Index (TLI), a more conservative version of CFI, penalizing overly complex models [5]. Our model achieved scores over the recommended threshold of 0.90 in both metrics (CFI = 0.971, TLI = 0.959) which indicates its good performance.

Afterward, we calculated Standardized Root Mean Square Residual (SRMR) which measures the mean absolute difference between observed and predicted correlations [48]. Our model had a sufficient SRMR of 0.070 which is below the recommended threshold of 0.080 [48] indicating adequate model fitness.

Next, we calculated the Root Mean Square Error of Approximation (RMSEA), which measures how well the model produces item covariances, instead of a baseline model comparison [53]. To interpret RMSEA, Cudeck [19] recommended RMSEA ≤ 0.05 as a close fit, $0.05 \leq$ RMSEA <= as a fair fit, values > 0.10 as indicative of a poor fit between the hypothesized model and the observed data. Our model's RMSEA (0.046) could be interpreted as a 'close fit'.

6.2 Reliability

To further confirm the internal consistency of our model, we recalculated Cronbach's α using the sample (N=112) collected for CFA. The results are represented in Table 3. The overall Cronbach's α exceeded the recommended threshold of 0.70 [18, 33]. However, two of the sub-scales i.e. EXP and HAR were less than the recommended threshold of 0.7 which is an indication of inadequate internal consistency.

7 DISCUSSION

In this section, we discuss the interpretation of the extracted factors, including the practical considerations for haptic designers and researchers. We also discuss the limitations and future work.

7.1 Factor Interpretation

Our findings refine the existing understanding of hedonic factors underlying HX by showing which items load, and how it compares to prior investigations with vibrotactile feedback in a remote study [43]. Two of our factors (Realism, Expressivity) closely match the intended construct and the measured outcome with vibrotactile feedback. One of our factors (Harmony) seems to align with an existing construct from the HX model. The final factor (Involvement), with only two loading items, is one we had to introduce to explain our results.

7.1.1 *Realism.*

Realism is the strongest factor in the model with items R1, R2 and R3, consistent with the original proposal. Realism is defined as whether the haptic effect convincingly exhibits what someone expects to feel in reality [28]. These results are consistent with the results obtained by Sathiyamurthy et al. [43]. "Realism" in this context remains heavily linked to believability, both of which load highly. As such, we believe these items do not necessarily refer to photorealism, but rather a convincing, believable sensation, and that this construct is appropriate whether focused on vibrotactile experience or general HX with diverse modalities.

The variable R4 (*"The haptic feedback matched my expectations"*) did not load significantly. This could be due to the fact that R4 is expecting the user to have some prerequisite expectations about the haptic feedback which might be perplexing the user. It could also be due to a mismatch of intended construct. In the formulation of Presence, and in the Presence Questionnaire [54], Control Factors are said to involve expectations, e.g., through expected immediacy of control, or expectations of being able to modify an environment. Control Factors are expected to influence immersion, but not involvement; ultimately, it could be that this item is not appropriate to measure realism, and is more appropriate to measure more specific control or immersion factors.

7.1.2 *Harmony.*

This factor has two items intended to measure Harmony (H3, H5) and one item intended to measure Immersion (I1). While I1 (*"The haptic feedback distracted me from the task"*) was intended for Immersion, we believe distraction from the task could be an outcome of low Harmony, and thus argue that all three could be related to Harmony. The key difference between H3, H5 and I1 are that H3, H5 refer to the haptic feedback's connection to other parts in the system, while I1 refers to the users' connection to the system.

The items in this factor are all negatively-phrased and could be interpreted as "disruption," a possible opposite to Harmony. However, we decided to represent this as Harmony for two reasons. First, all the negatively-phrased items were reverse-coded at the data

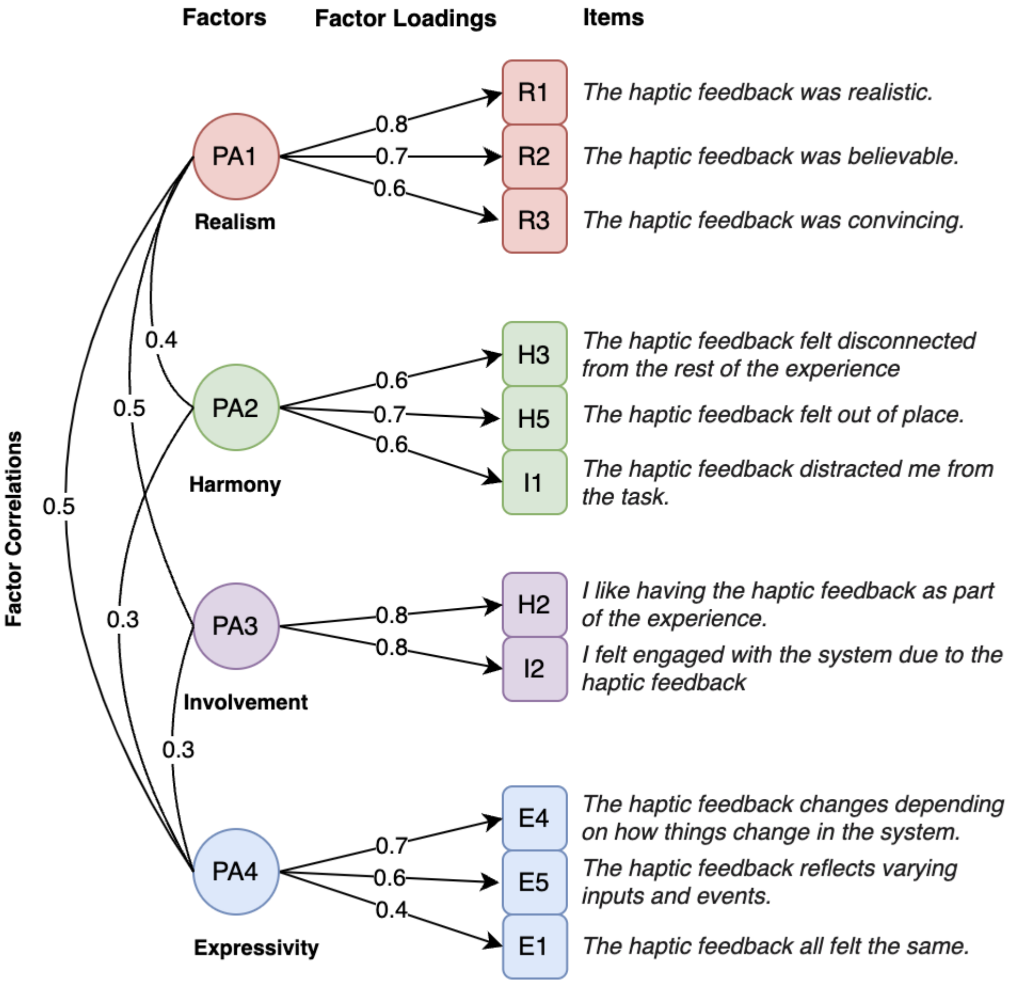

Figure 4: Path diagram for the 4-factor 11-item model obtained using promax rotation and principal axis factoring from exploratory factor analysis (N=291)

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

Table 2: 11 items finalized after refinement alongside their loading score on their respective factors. PA1=Realism, PA2=Harmony, PA3=Involvement, PA4=Expressivity

	REL	INV	EXP	DIS	α
REL					0.79
INV	0.43				0.76
EXP	0.48	0.35			0.53
HAR	0.53	0.28	0.32		0.59

Table 3: Factor correlations and internal consistency (Cronbach's alpha) for 11 items from confirmatory factor analysis (CFA) (N=112) where REL=Realism, ENG=Engagement, INV=Involvement and HAR=Harmony. Note that here, in contrast to the EFA findings, majority of the subscale's Cronbach's α is below the recommended threshold of 0.70 [7]

cleaning stage, meaning that, low scores on negative items were changed to higher ones leading to a positive correlation. Therefore, this factor could be interpreted as measuring the absence of disruption or Harmony in the system. Second, items H1 (*"The haptic feedback fits well with the other senses"*) and H4 (*"The haptic feedback felt appropriate when and where I felt it"*) did not load onto the model at all. This could reinforce what hapticians have long claimed: that poor haptics are noticeable (and here, potentially measurable), but good haptics are subtle and less prominent [28, 44]. As such, our resulting model might measure when Harmony is not achieved.

7.1.3 Involvement. Involvement is the only newly-defined factor that emerged from our investigation. Although the items (H2 and I2) in this factor correspond to different intended constructs (Harmony and Immersion) from the hypothesized model, taken together they could be interpreted as measuring the user's "Involvement". Involvement is a psychological state experienced as a consequence of focusing one's energy and attention on a coherent set of stimuli or meaningfully related activities and events [54]. Involvement depends on the degree of significance or meaning that the individual attaches to the stimuli, activities, or events. The fact that both items start with "I" indicates it is focused on the subjective experience. In this case, we suppose that Involvement, as measured by these two items, could represent engagement with the system due to meaningful haptics.

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

7.1.4 Expressivity. The fourth factor consists of items E1, E4 and E5 where E1 is a negatively-phrased item (reversed-coded in data cleaning stage). Expressivity has been defined as such that it allows users to feel the haptics distinguishably reflect varying user input and system events [28]. This construct is captured by the three items. E4 and E5 measures the extent to which a user's different interactions with the system result in different forms of haptic feedback. E1 being a negatively-phrased item, measures the opposite of that and after reverse-coding it could be interpreted as the haptic feedback being not constant. The fact that E2 and E3 didn't load significantly may indicate the expressivity here demands correspondence with inputs and other parts of the system, rather than just variations.

Compared to the vibrotactile evaluation by Sathiyamurthy et al. [43], the only difference is that E2 (*"I felt adequate variations in the haptic feedback"*) loaded with vibrotactile feedback, and did not load with multimodal feedback. This particular item has its roots in the "variation in feedback intensity/vibration intensity" cluster in the originally proposed HX model [28]. We suggest that designers of vibrotactile feedback should focus on having adequate variation, while other types of feedback do not have this as central of a concern.

7.2 Differences with the vibrotactile model and absence of autotelics

Our extracted model differs from the model extracted by Sathiyamurthy et al. [43], who followed a similar protocol in a remote study with a variety of vibrotactile devices. Our model has four factors with the absence of Immersion and Autotelics as independent factors and has lower items per factor. Autotelics is a purely hedonic factor measuring the experience of touch in and of itself [38].

Alternatively, the absence of Autotelics could mean that our participants, and thus similar populations, are not able to reliably notice (or care about) Autotelics. This could again be because haptic feedback is subtle, and might require focused attention on it to know what a user likes. In addition, it could also be due to the fact that novel devices and uncommon haptic feedback were used in the study which led to novelty effect. This could mean that the novelty effect made it difficult for participants to focus on or assess what felt good or bad in and of itself.

Differences between our developed model and the vibrotactile one could be due to the fact that for developing the initial draft of the HX model, novice input relied heavily on recognizable commercial devices (e.g., smartphones and gaming consoles), which overwhelmingly employ vibrotactile feedback. However, we used less common commercial devices (TanvasTouch, Ultraleap, 3D Systems Touch, Oculus Quest 2, and Haply) with different forms of haptic feedback (mid-air, vibrotactile, force, surface-haptics). We thus wonder whether HX might vary depending on the modality. Namely, vibrotactile experience (VX) might differ from a general sense of HX. Both share constructs of Realism, Expressivity and Harmony, although Expressivity might be represented slightly differently, i.e., with "adequate variations" being important with VX, but the HX model may be more tuned to vibrotactile feedback than initially intended.

7.3 HX may vary depending on the modality, but there are general similarities

We believe that HX may have a different structure depending on the modality, but with some elements that transcend modalities. There is evidence for a general model structure that is similar to a vibrotactile-focused model. Namely, the structure that we found (Realism, Harmony, Involvement, Expressivity) emerged when we studied several devices with varying modalities. An overlapping model was found by Sathiyamurthy et al. [43], with the same Realism items and almost the same Expressivity and Harmony items. However, in the present model, Involvement was introduced, and Autotelics and Immersion dispersed. Thus, we suggest that there could be a structure for HX that is different for each haptic modality, and a separate general structure that could be useful across devices and modalities.

Practitioners and researchers can user this general model to guide their work. The emergent factors from our model cover a wide range of complementary considerations. Realism and believability are common goals of many haptic systems, but may complement Involvement (engagement and enjoyment) as a goal. Harmony (many senses being synchronized and working together) is relatively independent from the richness of the haptics (Expressivity). Practitioners might be able to amend the original HX model [28] with these constructs when working across different haptic devices, making it easier to articulate their goals and communicate elements of their design. Meanwhile, researchers can further explore modality specific evaluative models and compare them with the obtained generalized HX model to help us understand how we should design haptics with specific devices or modalities. Ultimately, though, we believe that the overlap between this model and the vibrotactile-focused model suggests that some elements of experience might be similar across different devices. If this is the case, it will make design easier and more streamlined if designers don't need to learn each modality completely in isolation.

7.4 Expertise in training

The HX model was developed with input from novices and expert hapticians. Novices typically had experience with vibrotactile devices. Although novices were told about other types of devices, they did not have the option to try non-vibrotactile devices. Kim and Schneider [28] had to rely on expert input to provide generality to the model. As such, it could be that the HX model includes constructs that are meaningful to expert hapticians generally, but non-hapticians only for vibrotactile feedback.

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

Unfortunately, we did not have a large enough sample size to conduct factor analysis of each modality separately to see if we could replicate the results of Sathiyamurthy et al. [43], as we only had one vibrotactile device (Oculus Quest 2) with 41 responses in our sample.

7.5 Hygiene and motivators

In our resulting model, we found that items intended for Harmony and Immersion (H2, H3, H5, I1, I2) were intertwined into two other factors: our resulting Harmony (H3, H5, I1) and Involvement (H2, I2). In the theoretical HX model, Harmony and Immersion are highly related: poor Harmony would almost certainly break Immersion, while Immersion could also be produced by attaining Harmony if desired for the system [28].

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

This could possibly relate to the notions of *hygienes* and *motivators*, as adapted to UX [52]. In this formulation, adapted from Herzberg's work on job satisfaction [26], low levels of hygienes contribute to poor experience, but high levels of hygienes are not enough to produce a good experience; motivators are the factors that contribute to good experience once hygienes have been met. In other words, the resulting structure might measure disruptive features that negatively impact the experience, and then positive features that mean it is a good experience.

7.6 Implications for measurement and evaluation

Our goal was to support a scale that can be used to measure HX across different types of haptic devices. Measuring someone's HX, distinguishing it from the overall user experience, and generalizing it across all haptic devices using a scale is a challenging task. However, we believe using a scale can complement qualitative feedback by being more accessible, affordable and efficient.

While the obtained model from this work needs further validation, we believe that it is an important milestone towards the objective of creating an evaluative scale for measuring HX. Haptic designers can then incorporate a future validated scale in the form of a survey into their design evaluation process with the end users. Users could then be required to interact with the haptic device and provide feedback by completing the survey. Subsequently, designers can analyze and compare the survey scores against different iterations and benchmarks set by other research studies. Based on this comparison, designers can manipulate the design parameters, and achieve a better scale score. For example, a low score obtained for the Harmony sub-scale could indicate that the haptic feedback is not well integrated with other modalities (visual or auditory). To improve the score, designers could probe causes (hardware, software, physical design) and improve the integration of the haptic feedback with the overall system. The next steps for scale development across devices will be to develop new items based on this structure in an effort to create new, more reliable means of measuring HX, and hopefully produce a scale that can support hapticians' needs for evaluation.

Even without such a scale, however, we believe this model can help guide existing evaluation and design practices. Designers can use the four factors we found as guiding criteria for their designs. The factors and language from the loaded items might help a designer articulate problems or elicit feedback during qualitative feedback. For example, in an interview, a designer might specifically ask someone if they felt engaged or involved with a system to get an idea of Involvement, then ask if the feedback was believable or realistic to get a rough idea of realism. Meanwhile, a software developer on the team might be asked to make the haptic feedback more Expressive, with guidance by suggesting that the developer makes the feedback reflect varying inputs and outputs and change when things change in the system.

7.7 Limitations

While this study gives evidence of a generalized structure for a 4-factor HX model, it is not without limitations. Although the scale's reliability was established in EFA, it did not hold true during CFA as two sub-scales had lower Cronbach's α. Therefore, we do not advocate the use of this version of the scale in user studies.

Additionally, there is novelty bias in the sample as the devices used in the study are not common devices for people outside of haptic research and development, and not easily accessible to respondents; the haptic feedback provided by these devices was unique and quite different from the more readily accessible vibrotactile haptic feedback. Koch et al. [29]. has defined novelty effect as "an increased motivation to use something, or an increase in the perceived usability of something, on account of its newness. When novelty eventually fades, usage patterns and/or perceived usability changes". In addition, the work of Rutten et al. [42] involving mid-air haptic feedback has shown the existence of novelty effect in UX research studies.

The presence of researchers on the study site along with the respondent might have led to social desirability bias. Social desirability bias occurs when a respondent opts for responses that they consider more socially acceptable rather than choosing responses that are a reflection of their true responses [21]. This leads to over-reporting of socially desirable responses and under-reporting of less socially desirable responses. Since our survey included negatively-phrased questions, we believe that it might be prone to social desirability bias.

We had a limited number of devices and applications in our study. It would be impossible to try every kind of haptic device, so we prioritized commercial haptic devices that would have a level of polish, which we felt was most appropriate to investigate studying the experience of a developed haptic system. These devices had a limited number of applications associated with them, but we did aim for variability by having some games and some end-user applications. We suggest that future work can continue to incorporate more devices and applications to help validate or correct our results.

Our obtained model does not consider the construct of user's agency. When designing for interactive systems, the sense of agency is important to achieve a user experience that grants a sense of control to the user [49]. In the context of haptic interactions, agency is encapsulated by two factors: haptic timing (when the device starts to move the device) and expected outcome of an action [49]. While our extracted model does contains the construct of Expressivity, which links richness of the response to inputs, it does not directly consider agency. Agency can either be separately studied for correlation to HX, or possibly incorporated into future models of HX.

7.8 Future Work

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

Convergent validity confirms whether the obtained model correlates with other instruments that measure the same or similar construct. For establishing convergent validity, we would find existing instruments through literature review that measure the four experiential dimensions of Harmony, Expressivity, Involvement and Realism. Subsequently, we would administer these instruments alongside our obtained model and calculate the Pearson correlation between their scores. A high correlation will affirm that our extracted model is indeed measuring what it was intended to measure. For example, a high correlation between our Involvement sub-scale and existing items related to Involvement from the Presence questionnaire would validate that our Involvement sub-scale indeed measures Harmony.

In contrast, divergent validity establishes that two unrelated constructs are indeed measuring different constructs. For establishing divergent validity, we would administer items from an unrelated construct such as Presence or Engagement, alongside our existing items. Afterwards, we would calculate Pearson correlation coefficient between the unrelated constructs. A low correlation coefficient would establish divergent validity of our proposed instrument. For example, if Expressivity has a low correlation with existing UX constructs not related to richness of feedback, then Expressivity is measuring a new construct that has not previously been measured.

8 CONCLUSION

In this paper, we present the development of a multidimensional scale using the experiential dimensions of HX. Through, two in-person user studies incorporating diverse haptic modalities and involving 430 participants we extracted and evaluated a 4-factor model that is generalizable. Using our 4-factor model, hapticians can obtain insights into their designs regardless of the type of haptic feedback. The obtained model has theoretical and practical implications for haptics research and also paves the way for future research on the measurement of haptic experience.

ACKNOWLEDGMENTS

We thank the University of Waterloo HCI community, especially the Haptic Experience Lab and Games Institute, for their continued guidance and feedback. We acknowledge the support of the Natural Sciences and Engineering Research Council of Canada (NSERC), [funding reference number 2019-06589], Canada Foundation for Innovation (CFI), Ontario Research Foundation (ORF) and Mitacs.

REFERENCES

[1] Vero Vanden Abeele, Katta Spiel, Lennart Nacke, Daniel Johnson, and Kathrin Gerling. 2020. Development and validation of the player experience inventory: A scale to measure player experiences at the level of functional and psychosocial consequences. *International Journal of Human-Computer Studies* 135 (2020), 102370. https://doi.org/10.1016/j.ijhcs.2019.102370

[2] Mahdi Azmandian, Mark Hancock, Hrvoje Benko, Eyal Ofek, and Andrew D Wilson. 2016. Haptic retargeting: Dynamic repurposing of passive haptics for enhanced virtual reality experiences. In *Proceedings of the 2016 chi conference on human factors in computing systems*. 1968–1979. https://doi.org/10.1145/2858036.2858226

[3] Juergen Baumgartner, Naomi Frei, Mascha Kleinke, Juergen Sauer, and Andreas Sonderegger. 2019. Pictorial System Usability Scale (P-SUS) Developing an Instrument for Measuring Perceived Usability. In *Proceedings of the 2019 chi conference on human factors in computing systems*. 1–11. https://doi.org/10.1145/3290605.3300299

[4] Peter M Bentler. 1990. Comparative fit indexes in structural models. *Psychological bulletin* 107, 2 (1990), 238. https://doi.org/10.1037/0033-2909.107.2.238

[5] Peter M Bentler and Douglas G Bonett. 1980. Significance tests and goodness of fit in the analysis of covariance structures. *Psychological bulletin* 88, 3 (1980), 588. https://doi.org/10.1037/0033-2909.88.3.588

[6] Marit Bentvelzen, Jasmin Niess, Mikołaj P Woźniak, and Paweł W Woźniak. 2021. The Development and Validation of the Technology-Supported Reflection Inventory. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–8. https://doi.org/10.1145/3411764.3445673

[7] Godfred O Boateng, Torsten B Neilands, Edward A Frongillo, Hugo R Melgar-Quiñonez, and Sera L Young. 2018. Best practices for developing and validating scales for health, social, and behavioral research: a primer. *Frontiers in public health* 6 (2018), 149. https://doi.org/10.3389/fpubh.2018.00149

[8] Kenneth A Bollen. 1989. *Structural equations with latent variables*. Vol. 210. John Wiley & Sons. https://doi.org/10.1002/9781118619179

[9] Richard W Brislin. 1980. Cross-cultural research methods. In *Environment and culture*. Springer, 47–82. https://doi.org/10.1007/978-1-4899-0451-5_3

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

[11] Florian Brühlmann, Beat Vollenwyder, Klaus Opwis, and Elisa D Mekler. 2018. Measuring the "why" of interaction: Development and validation of the user motivation inventory (umi). In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. 1–13. https://doi.org/10.1145/3173574.3173680

[12] M Burmester, M Hassenzahl, and F Koller. 2003. AttrakDiff: Ein Fragebogen zur Messung wahrgenommener hedonischer und pragmatischer Qualität. In *Mensch & Computer*, Vol. 57. 2003.

[13] Edward G Carmines. 1981. Analyzing models with unobserved variables. *Social measurement: Current issues* 80 (1981). https://doi.org/10.117/0049124183011003003

[14] Raymond B Cattell. 1952. Factor analysis: an introduction and manual for the psychologist and social scientist. (1952).

[15] Lee Anna Clark and David Watson. 2016. Constructing validity: Basic issues in objective scale development. (2016). https://doi.org/10.1037/1040-3590.7.3.309

[16] Andrew L Comrey. 1978. Common methodological problems in factor analytic studies. *Journal of consulting and clinical psychology* 46, 4 (1978), 648. https://doi.org/10.1037/0022-006X.46.4.648

[17] Andrew L Comrey and Howard B Lee. 2013. *A first course in factor analysis*. Psychology press. https://doi.org/10.4324/9781315827506

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

[19] Robert Cudeck. 1993. Alternative ways of Assessing Model Fit. *Testing structural equation models* 154 (1993), 136. https://doi.org/10.1177/0049124192021002005

[20] Richard L Gorsuch. 2014. *Factor analysis: Classic edition*. Routledge. https://doi.org/10.4324/9781315735740

[21] Pamela Grimm. 2010. Social desirability bias. *Wiley international encyclopedia of marketing* (2010). https://doi.org/10.1002/9781444316568.wiem02057

[22] Edward Guadagnoli and Wayne F Velicer. 1988. Relation of sample size to the stability of component patterns. *Psychological bulletin* 103, 2 (1988), 265. https://doi.org/10.1037/0033-2909.103.2.265

[23] Louis Guttman. 1954. Some necessary conditions for common-factor analysis. *Psychometrika* 19, 2 (1954), 149–161. https://doi.org/10.1007/BF02289162

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

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

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

[27] Henry F Kaiser. 1970. A second generation little jiffy. (1970). https://doi.org/10.1007/BF02291817

[28] Erin Kim and Oliver Schneider. 2020. Defining haptic experience: Foundations for understanding, communicating, and evaluating HX. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–13. https://doi.org/10.1145/3313831.3376280

[29] Michael Koch, Kai von Luck, Jan Schwarzer, and Susanne Draheim. 2018. The novelty effect in large display deployments–Experiences and lessons-learned for evaluating prototypes. In *Proceedings of 16th European conference on computer-supported cooperative work-exploratory papers*. European Society for Socially Embedded Technologies (EUSSET). https://doi.org/10.18420/ecscw2018_3

[30] Bettina Laugwitz, Theo Held, and Martin Schrepp. 2008. Construction and evaluation of a user experience questionnaire. In *Symposium of the Austrian HCI and usability engineering group*. Springer, 63–76. https://doi.org/10.1007/978-3-540-89350-9_6

[31] Effie L-C Law, Florian Brühlmann, and Elisa D Mekler. 2018. Systematic review and validation of the game experience questionnaire (geq)-implications for citation and reporting practice. In *Proceedings of the 2018 Annual Symposium on Computer-Human Interaction in Play*. 257–270. https://doi.org/10.1145/3242671.3242683

[32] Emanuela Maggioni, Erika Agostinelli, and Marianna Obrist. 2017. Measuring the added value of haptic feedback. In *2017 ninth international conference on quality of multimedia experience (QoMEX)*. IEEE, 1–6. https://doi.org/10.1109/QoMEX.2017.7965670

[33] Robert K McKinley, Terjinder Manku-Scott, Adrian M Hastings, David P French, and Richard Baker. 1997. Reliability and validity of a new measure of patient satisfaction with out of hours primary medical care in the United Kingdom: development of a patient questionnaire. *Bmj* 314, 7075 (1997), 193. https://doi.org/10.1136/bmj.314.7075.193

[34] Michael Minge, Manfred Thüring, and Ingmar Wagner. 2016. Developing and validating an English version of the meCUE questionnaire for measuring user experience. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, Vol. 60. SAGE Publications Sage CA: Los Angeles, CA, 2063–2067. https://doi.org/10.1177/1541931213601468

[35] Thomas Muender, Michael Bonfert, Anke Verena Reinschluessel, Rainer Malaka, and Tanja Döring. 2022. Haptic Fidelity Framework: Defining the Factors of Realistic Haptic Feedback for Virtual Reality. In *CHI Conference on Human Factors in Computing Systems*. 1–17. https://doi.org/10.1145/3491102.3501953

[36] Richard G Netemeyer, William O Bearden, and Subhash Sharma. 2003. *Scaling procedures: Issues and applications*. sage publications. https://doi.org/10.4135/9781412985772

[37] Jum C Nunnally. 1978. Psychometric Theory, McGraw-Hill Book Company, New York, NY. (1978). https://doi.org/10.1177/014662169501900308

[38] Marianna Obrist, Sriram Subramanian, Elia Gatti, Benjamin Long, and Thomas Carter. 2015. Emotions mediated through mid-air haptics. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. 2053–2062. https://doi.org/10.1145/2702123.2702361

[39] Horea Pauna, Pierre-majorique Léger, Sylvain Sénécal, Marc Fredette, François Courtemanche, Shang-Lin Chen, Élise Labonté-Lemoyne, and Jean-François Ménard. 2018. The psychophysiological effect of a vibro-kinetic movie experience: The case of the D-BOX movie seat. In *Information systems and neuroscience*. Springer, 1–7. https://doi.org/10.1007/978-3-319-67431-5_1

[40] Joann Peck and Terry L Childers. 2003. Individual differences in haptic information processing: The "need for touch" scale. *Journal of Consumer Research* 30, 3 (2003), 430–442. https://doi.org/10.1086/378619

[41] Miquel Porta. 2008. *A dictionary of epidemiology*. Oxford university press. https://doi.org/10.1093/aje/kwv031

[42] Isa Rutten, Lawrence Van den Bogaert, and David Geerts. 2020. From initial encounter with mid-air haptic feedback to repeated use: the role of the novelty effect in user experience. *IEEE Transactions on Haptics* 14, 3 (2020), 591–602. https://doi.org/10.1109/TOH.2020.3043658

[43] Suji Sathiyamurthy, Melody Lui, Erin Kim, and Oliver Schneider. 2021. Measuring Haptic Experience: Elaborating the HX model with scale development. In *2021 IEEE World Haptics Conference (WHC)*. IEEE, 979–984. https://doi.org/10.1109/WHC49131.2021.9517220

[44] Oliver Schneider, Karon MacLean, Colin Swindells, and Kellogg Booth. 2017. Haptic experience design: What hapticians do and where they need help. *International Journal of Human-Computer Studies* 107 (2017), 5–21. https://doi.org/10.1016/j.ijhcs.2017.04.004

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

[46] Tanay Singhal and Oliver Schneider. 2021. Juicy haptic design: Vibrotactile embellishments can improve player experience in games. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–11. https://doi.org/10.1145/3411764.3445463

[47] Hyewon Suh, Nina Shahriaree, Eric B Hekler, and Julie A Kientz. 2016. Developing and validating the user burden scale: A tool for assessing user burden in computing systems. In *Proceedings of the 2016 CHI conference on human factors in computing systems*. 3988–3999. https://doi.org/10.1145/2858036.2858448

[48] Barbara G Tabachnick, Linda S Fidell, and Jodie B Ullman. 2007. *Using multivariate statistics*. Vol. 5. pearson Boston, MA. https://doi.org/10.5555/1213888

[49] Daisuke Tajima, Jun Nishida, Pedro Lopes, and Shunichi Kasahara. 2022. Whose Touch is This?: Understanding the Agency Trade-Off Between User-Driven Touch vs. Computer-Driven Touch. *ACM Transactions on Computer-Human Interaction* 29, 3 (2022), 1–27.

[50] Sigmund Tobias and James E Carlson. 1969. Brief report: Bartlett's test of sphericity and chance findings in factor analysis. *Multivariate behavioral research* 4, 3 (1969), 375–377. https://doi.org/10.1207/s15327906mbr0403_8

[51] Gustavo F Tondello, Karina Arrambide, Giovanni Ribeiro, Andrew Jian-lan Cen, and Lennart E Nacke. 2019. "I don't fit into a single type": A Trait Model and Scale of Game Playing Preferences. In *IFIP Conference on Human-Computer Interaction*. Springer, 375–395. https://doi.org/10.1007/978-3-030-29384-0_23

[52] Alexandre N Tuch and Kasper Hornbæk. 2015. Does Herzberg's notion of hygienes and motivators apply to user experience? *ACM Transactions on Computer-Human Interaction (TOCHI)* 22, 4 (2015), 1–24. https://doi.org/10.1145/2724710

[53] Daniel Votipka, Desiree Abrokwa, and Michelle L Mazurek. 2020. Building and validating a scale for secure software development self-efficacy. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–20. https://doi.org/10.1145/3313831.3376754

[54] Bob G Witmer and Michael J Singer. 1998. Measuring presence in virtual environments: A presence questionnaire. *Presence* 7, 3 (1998), 225–240. https://doi.org/10.1162/105474698565686