POLYPUBLIE

Development and validation of four social scales for the UX evaluation of interactive products

Ehsan Mortazavi, Philippe Doyon-Poulin, Daniel Imbeau, Jean-Marc Robert

Document en libre accès dans PolyPublie
Open Access document in PolyPublie

Date	2024
Type	Article de revue / Article
Version	Version finale avant publication / Accepted version
URL de PolyPublie PolyPublie URL	https://publications.polymtl.ca/55976/
Évaluation par les pairs Peer reviewed	Révisé par les pairs / Refereed
Conditions d'utilisation Terms of use	© Tous droits réservés / All rights reserved
Référence Citation	Mortazavi, E., Doyon-Poulin, P., Imbeau, D., & Robert, J.-M. (2024). Development and validation of four social scales for the UX evaluation of interactive products. International Journal of Human–Computer Interaction, 40(20), 6608-6621. https://doi.org/10.1080/10447318.2023.2258026

Publication officielle
Official publication

Titre de la revue Journal title	International Journal of Human–Computer Interaction (vol. 40, no. 20)
Maison d'édition Publisher	Informa UK (Taylor & Francis)
URL officielle Official URL	https://doi.org/10.1080/10447318.2023.2258026
Mention légale Legal notice	This is an Accepted Manuscript of an article published by Taylor & Francis in International Journal of Human–Computer Interaction (vol. 40, no. 20) on 2024, available online: https://doi.org/10.1080/10447318.2023.2258026

Ce fichier a été téléchargé à partir de PolyPublie, le dépôt institutionnel de Polytechnique Montréal
This file has been downloaded from PolyPublie, the institutional repository of Polytechnique Montréal
https://publications.polymtl.ca

Development and Validation of Four Social Scales for the UX Evaluation of Interactive Products

Ehsan Mortazavi[a]*, Philippe Doyon-Poulin[a], Daniel Imbeau[a], and Jean-Marc Robert[a]

[a] *Department of Mathematics and Industrial Engineering, Polytechnique Montreal, Montreal, Canada*

Ehsan Mortazavi (corresponding author), PhD candidate, Department of Mathematics and Industrial Engineering, Polytechnique Montreal, ehsan.mortazavi@polymtl.ca, (438) 927-6327

Philippe Doyon-Poulin, Assistant Professor, Department of Mathematics and Industrial Engineering, Polytechnique Montreal, philippe.doyon-poulin@polymtl.ca

Daniel Imbeau, Full Professor, Department of Mathematics and Industrial Engineering, Polytechnique Montreal, daniel.imbeau@polymtl.ca

Jean-Marc Robert, Adjunct Professor, Department of Mathematics and Industrial Engineering, Polytechnique Montreal, jean-marc.robert@polymtl.ca

Development and Validation of Four Social Scales for the UX Evaluation of Interactive Products

The social dimension of interactive products covers all aspects of our relationships with others that are impacted by owning and using such products. Although social features are making their way into a growing number of interactive products, there is a lack of an evaluation tool to capture the social dimension of the user experience (UX). This study addressed this shortcoming by developing and validating new social scales based on the UEQ+ framework. We developed four social scales to encompass various aspects within the social dimension. For scale development, 229 participants rated their UX with products having social aspects. Exploratory factor analysis allowed us to identify four sub-dimensions (Identification, Social interaction, Social stimulation, and Social acceptance), each evaluated with four items. For scale validation, 450 participants evaluated the UX of three product categories, using the new social scales, AttrakDiff, and the six UX dimensions of UEQ+. Results of MANOVA showed that the social scales discriminated the three categories ($F(8, 560) = 20.68$, $P < 0.001$, Pillai's trace = 0.456). The four social scales developed in this study can be combined with other UX dimensions of the UEQ+ modular framework to provide a comprehensive overview of user interaction with products.

Keywords: user experience, UX, UX evaluation, social scale, UEQ+

1. Introduction

User experience (UX) is a multidimensional construct that encompasses all aspects of user interaction with products, services, and systems[1] (Robert & Lesage, 2011). It includes pragmatic dimensions, that are associated with products usability and functionality, as well as hedonic dimensions that are associated with users' psychological well-being, like pleasure and stimulation, that together give a comprehensive picture of user-product interaction (Hassenzahl, 2003). As technology advances, new environments and ways of interactions such as virtual reality are

[1] For the sake of brevity, hereafter, we only use product(s) instead of product(s), service(s), and system(s).

introduced, resulting in new UX dimensions that should be considered in UX design and evaluation (Mortazavi, Doyon-Poulin, Imbeau, Taraghi, & Robert, 2021). Despite the fact that people use more and more products that connect them to others, the social aspects of UX have not been adequately investigated in UX studies (Hinderks, Winter, Schrepp, & Thomaschewski, 2019; Mortazavi et al., 2021). Yet, they have major impacts at a personal level for communication and entertainment purposes, as well as at a business level to improve technology adoption and productivity (Li, 2015). An evident example of such products is social media platforms that are being used by 4.7 billion people (Petrosyan, 2023). Social media serves purposes beyond basic individual communication, as businesses design marketing strategies that leverage the potential of customers on these platforms to promote and share information about the launch of new products (Zarnadze, 2020). Moreover, social features can contribute to improving people's wellbeing and preventing isolation. The case of wearable social interaction aid for children with autism is a prominent example of assistive technologies that improve social interaction (Washington et al., 2016). By using wearable glasses, children with autism are provided with real-time social cues during their interactions with others, which would have otherwise been challenging for them to discern. This assistance aids them in sustaining engagement within social interactions with others (Washington et al., 2016). Similarly, a social glasses device and a tactile wristband were used to replicate eye contact during a blind-sighted conversation which can help blind people by giving them more visual information and enhance their communication quality (Qiu, Hu, Han, Osawa, & Rauterberg, 2020). Even products that were not primarily built for social activities are now incorporating them into their core features. For instance, thanks to the influence of social media, social sharing and social competing features are added to fitness trackers to encourage users to exercise more and share their achievements with their community (Zhu, Dailey, Kreitzberg, &

Bernhardt, 2017). Communicating with others, expressing oneself through the possession or use of a product, collaborating with others on a task, and creating an experience with others (i.e., co-experience) are all manifestations of social aspects.

The impact of social aspects on UX has been studied from two perspectives. The first characterizes it as an element of the context of use. Arhippainen and Tahti (Arhippainen & Tähti, 2003) identified social and cultural factors as part of the context in user-product interaction that influence UX. They mentioned time pressure, pressure to succeed or fail, and explicit and implicit requirements as examples of social factors (Arhippainen & Tähti, 2003). Lallemand and Koenig (Lallemand & Koenig, 2020) took a step further and developed the UX context scale (UXCS), a tool that can be used in conjunction with UX evaluation tools to portray the contexts within which users interact with a product. Other people's presence, interactions with them, and how users feel in that environment are items measured in UXCS under social context (Lallemand & Koenig, 2020). However, it does not evaluate the social aspects of the UX with a product.

The second perspective considers social aspect as a UX dimension that should be measured on its own. In order to ensure a common understanding, the terms used in this paper are classified as follows: at the highest level, a "dimension" represents a significant factor explaining the UX of a product (Robert, 2014), "sub-dimensions" are the constituent elements of each dimension and "items" are the constituent elements of a sub-dimension. Identification is an example of a UX sub-dimension of the social dimension present in the AttrakDiff questionnaire (Hassenzahl, Burmester, & Koller, 2003). Identification is the degree to which a user relates to a product and expresses oneself through its possession (Hassenzahl, 2003). It is described as either self-focused or relationship-focused. The former is concerned with the users' self-perception as a result of possessing or using a product, whereas the latter reflects their social identity (Yoon, Kim, & Kang,

2020). Identification has roots in the influence and popularity needs of individuals, i.e., the need to be liked, respected, and regarded as influential to others (Hassenzahl, Diefenbach, & Göritz, 2010), which is in line with Jordan's notion of socio-pleasure (Jordan, 2000). In this study, we followed the second perspective, which considers social aspects as an integral part of UX evaluation tools, treating them as a UX dimension in need of more investigation. Knowing how socially attractive a product is, and how its social features influence the overall UX are of growing importance and deserve more attention from the research community. As presented in the following section, despite the wealth of available UX evaluation tools, there remains a gap in having a single tool that comprehensively measures different aspects of the social dimension without being bounded to a specific field of use. This study aims to bridge this gap by developing and validating social scales that can be used in conjunction with other UX evaluation tools.

2. Literature review

Reviewing the UX evaluation tools shows that social dimension is evaluated under different terms. Table 1 shows subjective UX evaluation tools and the social sub-dimensions they cover. General evaluation tools such as AttrakDiff are applicable to a variety of products, while others, such as the game experience questionnaire (GEQ), are developed for a particular field of use. However, neither covers all sub-dimensions of the social dimension.

Table 1. UX evaluation tools with social sub-dimensions

UX evaluation tools	Type	Social sub-dimensions	Number of items
AttrakDiff (Hassenzahl, Burmester, & Koller, 2003)	General	Identification	7
meCUE (Minge, Thuring, Wagner, & Kuhr, 2017)	General	Status	3
Chinese UX questionnaire (Liu, Zheng, Liu, Jian, & Peng, 2013)	General	Conformity, identification	5
EGameFlow (Fu, Su, & Yu, 2009)	Specific (education)	Social interaction	6
Faser LX (Safsouf, Mansouri, & Poirier, 2019)	Specific (education)	Subjective norm, self-image, interaction	5

GEQ (IJsselsteijn, de Kort, & Poels, 2013)	Specific (games)	Empathy, negative feeling, behavioral involvement	17
GUESS (Phan, Keebler, & Chaparro, 2016)	Specific (games)	Social connection	4
MEEGA+ (Petri, von Wangenheim, & Borgatto, 2016)	Specific (games)	Social interaction	3
Smart TV UX (Jang & Yi, 2019)	Specific (smart TV)	Social relatedness	4

In addition to AttrakDiff, meCUE (Minge, Thuring, Wagner, & Kuhr, 2017) is another generally applicable UX evaluation tool that includes a social sub-dimension. It is analogous to AttrakDiff in that it focuses on self-expression and how a product can communicate identity to others under the UX sub-dimension of "status". A more comprehensive view on social dimension of UX is presented by Park et al (Park, Han, Kim, Cho, & Park, 2013). Although they did not develop a UX evaluation tool they defined sociability with three sub-dimensions — social emotion, social value, and friendship — for the UX of mobile phones and services. Sociability was defined as the "degree to which a product satisfies the user's desire of being sociable". The sub-dimensions respectively evaluate product's ability to enable feeling and sharing emotions socially; its ability to support user values such as social issues and problems; and its ability to enable making relationship with other people. In another study, Ryu and Kim (Ryu & Kim, 2019) evaluated sociability with only one item in their questionnaire (without covering the three sub-dimensions) to find whether or not a medical information system satisfied the user social needs. One can notice the limited attention to different social aspects in general UX evaluation tools. AttrakDiff and meCUE only evaluate identification, and the User Experience Questionnaire (UEQ) (Laugwitz, Held, & Schrepp, 2008) — the most commonly used subjective evaluation tool — does not cover any social aspect (Mortazavi et al., 2021). The same goes for UEQ+ (Schrepp & Thomaschewski, 2019), a modular framework developed based on UEQ, including 20 UX dimensions that can be used for the evaluation of different products, yet none relates to social aspect.

Considering the impact of culture on UX, the Chinese UX questionnaire was developed based on AttrakDiff (Liu, Zheng, Liu, Jian, & Peng, 2013). Its authors proposed "conformity" as the new UX dimension that addressed the eastern culture of giving importance to other's opinions. Conformity deals with the prevalence of a product in the market and how widely it is used by people. It is to some extent close to social acceptance, another sub-dimension that has been rarely used in UX evaluation (Weiss & Willkomm, 2013). Social acceptance is "how a user feels when interacting with a system in relation to the social situation, e.g., how uncomfortable or embarrassed they feel with respect to other people or one's own norm" (Weiss & Willkomm, 2013). The negative impact of lack of social acceptability affects not just the UX of a product, but also the user's self and social image, with the risk of stigmatization and misjudgment by others (Koelle, Ananthanarayan, & Boll, 2020). Making an intimate relationship with others when watching smart TVs (Jang & Yi, 2019), sharing information via a mobile app (Sun & May, 2013), collaborating with other students to co-write a poem (Kantosalo & Riihiaho, 2019), and using an interactive public display system for improving social interaction (Kang, Hengeveld, Hummels, & Hu, 2022) are all examples of where different social sub-dimensions of UX need to be addressed.

Social interaction in the form of communicating with others, collaborating with teammates, or expressing oneself through game characters has been widely used in the game industry. Multiplayer online games, gaming conventions, and game streaming services are all examples of social interaction media for players and spectators. As a result, UX evaluation tools developed specifically for games include social dimension, such as GEQ (IJsselsteijn, de Kort, & Poels, 2013) and Game User Experience Satisfaction Scale (GUESS) (Phan, Keebler, & Chaparro, 2016). GEQ features a separate questionnaire for the social presence that covers players' emphatic responses, negative feelings towards others, and behavioral involvement in games (IJsselsteijn et al., 2013).

Another field where social dimension has gained attention is education. Interactions within and between learners and instructors, subjective norm, and self-image are social sub-dimensions evaluated in the FASER LX tool that evaluates learners experience in an e-learning system (Safsouf, Mansouri, & Poirier, 2019). Other researchers used serious games for learning purposes and developed tools like EGameFlow (Fu, Su, & Yu, 2009) and Model for the Evaluation of Educational Games (MEEGA+) (Petri, von Wangenheim, & Borgatto, 2016) with social sub-dimensions. Reviewing the items of these tools highlights the case-specific nature of their development. Moreover, there is no suggestion of the possibility of a modular use of these tools except for GEQ social presence module. As a result, it is unclear whether these tools can be applied to other contexts.

Analyzing the UX evaluation tools showed that the social dimension of UX concerns three categories: product, user, and group (society). The product-related items investigate whether a product can provide social interactions. For example in EGameFlow, the item "the game supports social interaction between players" highlights the product's capability of making interaction. The user-related items evaluate what users can do with the social features of a product. For instance, Smart TV UX questionnaire (Jang & Yi, 2019) uses the item "I can form an intimate relationship with others by using the smart TV" and GEQ uses "I empathized with the other(s)" as actions that can be done by users with social features of products. The group-related items investigate how users/owners of a product are perceived by other people and social norms. The item "I was not worried about other people's judgement" of the flow sub-dimension of UX in IVE questionnaire (Tcha-Tokey, Christmann, Loup-Escande, & Richir, 2016) and the item "Rarely used – widely used" of the Chinese UX questionnaire are examples of this category. Regardless of the naming in

different tools, the social sub-dimensions provided in Table 1 includes at least one of the three categories of items. However, none of the UX tools cover all the social sub-dimensions.

Clearly, the UX community needs a new tool to evaluate the social dimension of UX when interacting with a product that is not bounded to a specific application field (e.g., games or education) and ensure a comprehensive coverage. A promising way of achieving this goal is to develop social scales and include them into the UEQ+ framework. Modularity of the UEQ+ enables it to be used in different contexts, allowing the evaluator the flexibility to tailor the tool and evaluate the set of UX dimensions most relevant to the study (Schrepp & Thomaschewski, 2019). Researchers can construct new UX dimensions and add them to the UEQ+ framework. For instance, response behavior, response quality, and comprehensibility are the new UX dimensions developed for voice assistants (Klein, Schrepp, Hinderks, & Thomaschewski, 2020). Similarly, haptic and acoustic dimensions were developed for the UX evaluation of household devices (Boos & Brau, 2017). In UEQ+, each dimension consists of four items measuring the dimension and a single item that determines the importance of that dimension for the UX evaluation of the product. The evaluation of an item is done on a 7-point Likert scale with two semantic differential anchor points. The consistency of UEQ+ simplifies the UX evaluation compared to using a combination of UX evaluation tools each with different rating scales and possible overlapping dimensions (Schrepp & Thomaschewski, 2019). It also provides an overall UX rating for products that can be used for comparisons. All these advantages make UEQ+ a suitable candidate for hosting UX social dimension that can be used with other UX dimensions for different products.

In this study, we designed, developed, and validated social scales that can be used with other UX dimensions of the UEQ+ framework. To this end, we followed the dimension

development process used in (Schrepp & Thomaschewski, 2019) for other UX dimensions of the UEQ+ framework, focusing on social dimension and present the validated questionnaire.

3. Study Framework

This study followed three phases: identification, integration, and validation (Figure 1). The **identification** phase consisted in extracting social sub-dimensions and corresponding items from the literature and from the analysis of the UX evaluation tools, and in selecting a sample of items for each social sub-dimension by UX experts. These items were used in the **integration** phase to develop social scales by performing Exploratory Factor Analysis (EFA) on the responses of 229 participants who rated their UX with interactive products having social characteristics. The **validation** phase included another survey study with 450 participants who rated their UX with specific products from three different product categories with various levels of social dimension. The results were used to calculate the reliability and validity of the social scales. This study was approved by Polytechnique Montreal's Research Ethics Board (CER-2122-47-D) and participants read and signed an informed consent form before taking part in the study.

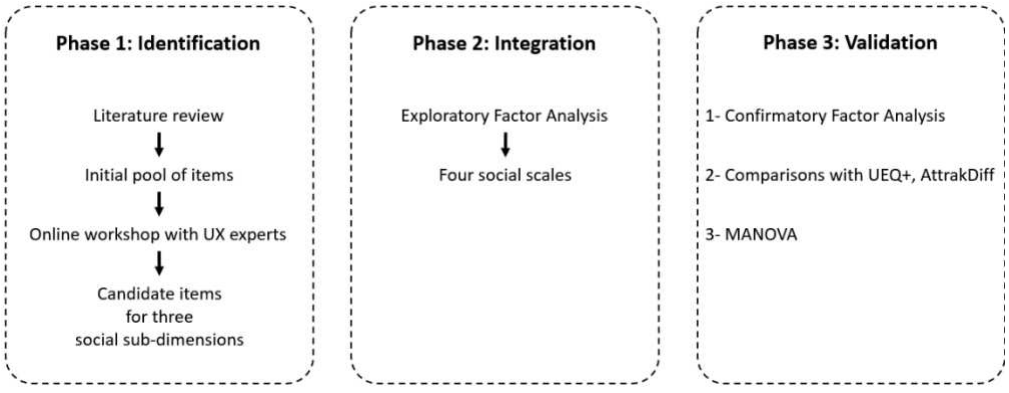


Figure 1. Three phases of the study

This paper is organized into eight sections. Section 4 presents the identification phase, followed by section 5 including the procedure, methodology, and results of the integration phase.

Then, section 6 presents the validation phase and its main results with the validated social scales. Section 7 puts our results into perspective by comparing them with those of other social scales and presents limitations of our work. Section 8 concludes the paper with the main take-aways and some propositions for future research work.

3.1. Participants

Two types of participants took part in this study, UX experts in the identification phase and end users for the two rounds of data collection (integration and validation phases). Seven UX experts (two from the industry and five from academia) were brought together to work in an online workshop during the identification phase. They had from 5 to 25 years of work experience in the fields of human-computer interaction (HCI), UX, cognitive ergonomics, or digital accessibility.

End users who filled out questionnaires in the two rounds of data collection were from Canada and recruited through Amazon Mechanical Turk (AMT). The first survey included a sample of 229 participants who had experience in using an interactive product with social aspects. The second data collection captured users' experience with specific products from three product categories: Social network, Online shopping, and Online banking. To that end, three separate questionnaires were prepared, and 150 participants responded to each category (total n=450). Table 2 presents the demographic data of experts and end user participants for each questionnaire in terms of age, gender, and level of education.

Table 2. Demographic data

Criteria		**Online workshop**	**1st data collection**	**2nd data collection**		
		UX experts (n=7)	Interactive products (n=229)	Social network (n=150)	Online shopping (n=150)	Online banking (n=150)
Age	18-29	0	70	23	32	33
	30-39	4	93	60	53	58
	40-49	1	38	38	31	29
	50-59	1	23	17	24	25

	60-69	1	5	11	10	4
	70+	0	0	1	0	1
Gender	Male	4	121	87	78	81
	Female	3	106	62	70	67
	Other	0	2	1	2	2
Education	High school or equivalent	0	41	39	35	43
	Bachelor	0	135	83	82	77
	MSc.	4	39	18	26	22
	Ph.D.	3	7	4	2	4
	Other	0	7	6	5	4

The 30-39 age group had the highest number of participants. For all questionnaires, participants were well balanced on gender. Overall, more than 70% of each group had a university-level education.

4. Phase 1 Identification

In a previous study, we performed a systematic literature review and an analysis of the UX evaluation tools and their dimensions (Mortazavi et al., 2021). As a result, we found that social dimension can be translated into three main sub-dimensions: identification, sociability, and social acceptance. **Identification** is the degree to which a user relates to a product and expresses oneself through its use or possession (Hassenzahl, 2003). **Sociability** is the degree to which a product enables communication with others to meet user's social need (Park et al., 2013). **Social acceptance** is "how a user feels when interacting with a system in relation to the social situation, e.g., how uncomfortable or embarrassed they feel with respect to other people or one's own norm" (Weiss & Willkomm, 2013).

An online workshop with seven UX experts took place in December 2021 on the collaborative whiteboard platform Miro. They reviewed the three social sub-dimensions and 24 items, discussed the possibility of merging them, and proposed 12 new items in a 90-minute session. Results of the workshop were reassessed by four of these experts and the number of items was reduced from 36 to 27. Similar-meaning items were removed, and some modifications were

made to the wording of the semantic differential poles. Overall, the 27 items were grouped under three social sub-dimensions: identification, sociability, and social acceptance.

5. Phase 2 Integration

5.1. Procedure

The first data collection with end users was done via an online questionnaire in a one-month period (March-April 2022). The questionnaire was designed on the QuestionPro website and data collection was done through AMT with participants receiving $2 compensation after having completed their response. The questionnaire was written in English and contained 34 questions. The first seven questions covered end users' demographic data, frequency of using interactive products with social aspects, name of their selected product, and a brief description of the main usage of the product. The remaining 27 questions were on social items, among which there were seven items on identification, eight on social acceptance, and 12 on sociability. The 27 items on a 7-point semantic differential scale were presented in 5 sections, each with a short introductory sentence, similar to (Schrepp & Thomaschewski, 2019). Participants were asked to evaluate their experience with interactive products having social aspects. They were given examples of such products like social networks, messengers, forums, online games, and online collaboration tools. However, they were given the freedom to choose any social product with which they would frequently interact. Those who rarely or never used any of the selected products did not qualify to fill out the questionnaire. The average response time was 4 minutes and responses that took less than 2 minutes were removed. In addition to eliminating incomplete answers, red-herring

questions[2] were used in the questionnaires to catch participants who seemed to have randomly answered the questions. 148 participants (39%) were removed including 22 (6%) who did not meet the qualifications to answer the questionnaire, 57 (15%) who left the questionnaire incomplete, and 69 (18%) who failed the red-herring questions. In total 229 responses were kept for the analysis. Table 3 shows the variety of interactive products with social aspects that were evaluated by participants, including both software (e.g., websites or application) and hardware (e.g., handheld devices or gaming consoles).

Table 3. Frequency of the evaluated products of the 1st data collection

Product	Frequency (n=229)	Percentage (%)
Facebook	**55**	**24.0**
Instagram	**25**	**10.9**
Twitter	**17**	**7.4**
Reddit	**15**	**6.6**
Online games	**12**	**5.2**
Mobile devices	**12**	**5.2**
WhatsApp	**7**	**3.1**
YouTube	**7**	**3.1**
Discord	**5**	**2.2**
LinkedIn	**5**	**2.2**
Microsoft teams	**5**	**2.2**
Other	**64**	**27.9**

5.2. Methodology

Exploratory factor analysis (EFA) was used for refining and reducing the items of the first survey. Following the review by seven UX experts, we assumed that the candidate items describe only the sub-dimension they relate to. With a modular approach in mind for the social scales, we conducted 3 separates EFAs on identification (7 items), social acceptance (8 items), and sociability (12 items) instead of a single EFA on 27 items. However, the structural model is tested in the validation

[2] In questionnaires, red-herring or attention check questions are used to detect participants who do not read carefully or answer randomly.

phase. As the initial step of EFA, we assessed the adequacy of sample size and the strength of correlations between items (Field, 2013). There are different suggestions for the suitable sample size for factor analysis. Some assert that the total number should be at least 300, while others suggest to use the ratio of participants to items such as 10 to 1 or 5 to 1 (Pallant, 2020). We adopted the conservative approach of having at least 200 respondents (Hinkin, 1998). Suitability of data for factor analysis was explored by using Bartlett's test and the Kaiser-Meyer-Olkin measure of sampling adequacy (KMO). Multicollinearity was tested by checking whether the determinant of the correlation matrix of items is above 0.00001 and no pair of items has a correlation coefficient greater than 0.9. There should be a reasonable correlation between items measuring the same construct. Therefore, by investigating the correlation matrix, we removed the items displaying several correlations below 0.3 with other items (Field, 2013).

We did factor extraction and rotation as the main analysis of EFA (Field, 2013). We applied principal axis factoring as the extraction method, and we used Kaiser's criterion, Scree plot, and parallel analysis to determine how many factors to retain. We used three criteria for factor extraction: factors with eigenvalues above Kaiser's criterion of 1, clear break on the Scree plot, and eigenvalues greater than the corresponding values for a randomly generated data matrix of the same size (Pallant, 2020). Factor analysis is an exploratory tool that assists researchers in making decisions based on different tests (Field, 2013). As a result, when criteria differ, decisions should be based on the knowledge of the researchers in respective fields. We applied the Varimax rotation technique and analyzed the rotated factor matrix for items with factor loadings above 0.3 (Field, 2013). Following the four items per UX dimension format of the UEQ+ framework (Schrepp & Thomaschewski, 2019), we kept the four items with highest factor loadings for each social sub-dimension. Lastly, the reliability of each scale was tested with Cronbach's alpha. Reliability refers

to the internal consistency of the items of each factor. We used IBM SPSS 28 for statistical analyses.

5.3. Results

The result of the online workshop with experts was a sample of 27 items organized under five introductory sentences as shown in Table 4. They are at the basis of the questionnaire submitted to the end-user participants. The sub-dimension column shows the sub-dimension to which items used in each EFA belong. The sample size of 229 participants was adequate (i.e., n>200) for performing EFAs.

Table 4. List of items that were used in the first survey

No.	Items		Sub-dimensions
Q1	In my opinion, using this product in social situations makes me feel		
Q1-1	uncomfortable	comfortable	Social acceptance
Q1-2	embarrassed	proud	Social acceptance
Q1-3	rejected by others	accepted by others	Social acceptance
Q1-4	judged by others	approved by others	Social acceptance
Q2	In my opinion, using/owning this product makes me look		
Q2-1	cheap	stylish	Identification
Q2-2	out of date	modern	Identification
Q2-3	unprofessional	professional	Identification
Q2-4	old-fashioned	fashionable	Identification
Q2-5	odd	normal	Social acceptance
Q3	In my opinion, this product is		
Q3-1	unknown	widespread	Social acceptance
Q3-2	rarely used	widely used	Social acceptance
Q3-3	disliked by my peers	liked by my peers	Social acceptance
Q3-4	non-reflective of my values	reflective of my values	Identification
Q3-5	diminishing my standing (status)	enhancing my standing (status)	Identification
Q3-6	socially exclusive	socially inclusive	Sociability
Q3-7	dissuasive	persuasive	Sociability
Q3-8	socially neutral	socially empowering	Sociability
Q3-9	socially discouraging	socially encouraging	Sociability
Q3-10	socially disengaging	socially engaging	Sociability
Q3-11	non-valuable	valuable	Identification
Q4	In my opinion, using/owning this product		
Q4-1	does not support learning from interaction with others	supports learning from interaction with others	Sociability
Q4-2	rejects communities	supports communities	Sociability
Q5	In my opinion, by using this product I can		
Q5-1	be ignorant to others	be helpful to others	Sociability
Q5-2	not share my feelings	share my feelings	Sociability
Q5-3	isolate from others	communicate with others	Sociability
Q5-4	not collaborate with others	collaborate with others	Sociability
Q5-5	not empathize with others	empathize with others	Sociability

Investigating the correlation matrix of items of each EFA resulted in excluding two items (Q3-1 and Q3-2) for social acceptance and one item (Q3-7) for sociability. For these items, more than half of the correlation coefficients with other items were below 0.3. After their removal, the correlation matrices were recalculated. Results showed determinant values above the recommended value of 0.00001 for all the new correlation matrices. The KMO values exceeded the recommended value of 0.6 (Pallant, 2020). The Bartlett's test of Sphericity was statistically significant for all sub-dimensions (at the alpha level of 0.01). Overall, all measures confirmed the suitability of the sample data and the absence of multicollinearity (Table 5).

Table 5. Measures for determining the suitability of sample data for EFAs

Measure		Identification	Social acceptance	Sociability
Determinant of correlation matrix		0.052	0.1	0.01
Kaiser-Meyer-Olkin measure of sampling adequacy		0.850	0.822	0.898
Bartlett's Test of Sphericity	Approx. Chi-Square	662.898	519.534	1025.175
	Degree of freedom	21	15	55
	Significance	<0.001	<0.001	<0.001

The Scree plot, parallel analysis, and Kaiser's criterion retained one factor for identification and one factor for social acceptance. It explained 55.2% of the variance for identification and 54.3% for social acceptance. However, there was not a consensus on the number of factors among the three criteria for sociability. Parallel analysis and scree plot showed one factor while Kaiser's criterion suggested retaining two. Considering the items that were loading on each factor, sociability was divided into two factors of social stimulation and social interaction. We defined **social stimulation** as qualities of a product that foster social interaction. **Social interaction** was also defined as the ways through which users make contact with others. The two-factor solution explained a total of 48.4 % of the variance. Social interaction accounted for 27.7% of the variance, whereas social stimulation contributed to 20.7 %.

Table 6 shows the rotated factor matrices of the 3 EFAs with corresponding factor loadings. The factor loadings represent the contribution each item makes to a factor (Field, 2013). The four items with the highest factor loading were selected for each social sub-dimension. These items had factor loading greater than 0.6, except for Q3-6. We tested content validity by investigating how theoretically relevant each item was to its factor which resulted in keeping Q3-6. This item was kept regardless of its relatively low loading on both factors, so as to conform with the 4-item format of the UEQ+ framework.

Table 6. Factor loading of items for the 3 EFAs

Identification		**Social acceptance**		**Sociability**		
Item number	Factor	Item number	Factor	Item number	Factor 1	Factor 2
Q2-4	**0.772**	Q1-4	**0.853**	Q5-3	**0.676**	
Q2-2	**0.750**	Q1-3	**0.827**	Q5-1	**0.649**	
Q2-1	**0.750**	Q1-2	**0.679**	Q5-5	**0.636**	
Q2-3	**0.681**	Q1-1	**0.646**	Q5-4	**0.618**	
Q3-5	0.663	Q2-5	0.567	Q4-2	0.607	
Q3-4	0.649	Q3-5	0.421	Q4-1	0.542	
Q3-11	0.559			Q5-2	0.533	
				Q3-9		**0.872**
				Q3-10	0.363	**0.652**
				Q3-8		**0.605**
				Q3-6	0.387	**0.388**

Cronbach' alpha exceeded the suggested cutoff of 0.7 confirming the reliability of all four social scales of identification, social acceptance, social interaction, and social stimulation as shown in Table 7.

Table 7. Reliability statistics of the four social scales

	Identification	**Social acceptance**	**Social interaction**	**Social stimulation**
Cronbach's alpha	0.841	0.846	0.802	0.792
Cronbach's alpha based on standardized items	0.844	0.845	0.803	0.793
Number of items	4	4	4	4

6. Phase 3 Validation

The goal of this phase was three-fold. First, to investigate the relationship between the social sub-dimensions using confirmatory factor analysis (CFA) and to test the fitness of this new data to the structural model we proposed. Second, to validate the scales by comparing the scores with those of other common evaluation tools. Third, to evaluate the social aspects of three product categories with the new social scales.

6.1. Procedure

The data collection was done over one month (May 2022) on the same survey and data collection websites and participants received $2 compensation after having completed their response. This time, three separate questionnaires, all written in English, were created with specific software products to be evaluated from three different product categories. These products were Facebook, Instagram, and LinkedIn for the Social network product category; Amazon, Walmart, and eBay for the Online shopping product category; and CIBC, RBC, TD, and Scotiabank for the Online banking product category. We chose these three categories as they represented a range of socially engaging products, with social networks being highly social (e.g., messaging, community interactions), online shopping being moderately social (e.g., products posting, reviews, shared cart), and online banking being low on that continuum. We used software products during the validation phase to facilitate comparison. We aimed to avoid introducing a new variable, such as the type of product (i.e., physical and software), that could potentially impact the outcomes. Each questionnaire contained the AttrakDiff questionnaire, six UX dimensions of the UEQ+ framework, and our newly developed social scales. This was done to investigate the validity of our social scales and the extent to which it is consistent with existing UX evaluation tools. Following the format of

the UEQ+ framework, participants rated the importance of UX dimensions for the UX evaluation of each product category. Overall, participants answered a total of 89 questions. Due to the similar format, we maintained the same order for both the new scales and the UEQ+ scales for all participants. However, AttrakDiff was administered either before or after the other two questionnaires to mitigate any potential bias arising from respondent boredom and fatigue due to answering many questions. The average response time was 7 minutes for each questionnaire, and we removed responses that took less than 3 minutes. Similar data cleaning approach to the first survey was taken and 290 participants (i.e., 39%) were removed from which 49 (7%) participants did not meet the qualifications to answer the questionnaire, 90 (12%) left the questionnaire incomplete, and 151 (20%) failed red-herring questions resulting in a total of 450 responses. Table 8 shows the frequency of selection of the products evaluated.

Table 8. Frequency of selection of the products evaluated during the 2nd data collection (n=150 for each product category)

Product category	Name	Frequency	Percentage (%)
Social network	Facebook	98	65.3
	Instagram	44	29.3
	LinkedIn	8	5.4
Online shopping	Amazon	135	90
	Walmart	11	7.3
	eBay	4	2.7
Online banking	CIBC	25	16.7
	TD	30	20
	RBC	23	15.3
	Scotiabank	21	14
	Other	51	34

6.2. Methodology

Prior to conducting CFA for each product category, we used Mahanalobis distances to identify and remove multivariate outliers from the three samples collected in the second round of data collection (Field, 2013). Mahanalobis distances have a chi-square distribution and we used the recommended threshold value of $p<0.001$ for removing outliers (Tabachnick & Fidell, 2019). As a result, 146,

142, and 144 valid responses were used for Social network, Online shopping, and Online banking product categories, respectively. We ran CFA for two models. The first model included four factors of identification, social stimulation, social interaction, and social acceptance. Considering that two factors of social stimulation and social interaction were two measures of the same underlying construct (i.e., sociability), large correlations between them were expected, which is why we used second-order CFA for sociability (Wang & Wang, 2020). As a result, we did the CFA for the second model on three latent variables of identification, sociability, and social acceptance. We also used modification indices to improve the fitness of the model. CFA uses different goodness-of-fit indices to assess the quality of model fit to the data (Hinkin, 1998). In this study, we reported: Chi-square statistical significance test, PCMIN/DF, Comparative Fit Index (CFI), Tucker Lewis Index (TLI), and Root Mean Square Error of Approximation (RMSEA). Composite Reliability (CR) was calculated, with values greater than 0.7 indicating acceptable reliability. Convergent validity that shows to what extent items of a factor are interrelated were tested by calculating Average Variance Extracted (AVE), with values above 0.5 deemed acceptable. The discriminant validity shows the extent to which latent factors are different and was tested by calculating AVE and Maximum Shared Variance (MSV) values (Jang & Yi, 2019).

To further investigate the validity of our scales, we analyzed the correlations of the four social scales with the scales of AttrakDiff and six scales of the UEQ+ framework to verify their redundancy with a threshold value of 0.8. As stated earlier, in addition to rating different items on a semantic differential scale, participants rated the importance of each dimension in the UX evaluation for the three product categories. We studied the ratings of social sub-dimensions to see whether they met our expectations of descending importance rating for Social network, Online shopping, and Online banking product categories.

The mean differences in ratings on the social scales between product categories were then examined using Multivariate Analysis of Variance (MANOVA). In this study the three product categories were considered as independent variables and the four social scales as the dependent variables. The data was checked to confirm that it met the MANOVA assumptions. The minimum sample size for conducting MANOVA is 20 responses per independent variable (Hair, Black, Babin, & Anderson, 2014). We tested the normality of the sample using the Shapiro-Wilk test. Multivariate normality was checked by calculating Mahanalobis distances (Pallant, 2020). We kept equal sample sizes across all independent variables (Hair et al., 2014). We checked for the linearity between each pair of dependent variables that was tested using a scatterplot. Finally, multicollinearity was tested by measuring correlations among the dependent variables where values above 0.9 show potential for multicollinearity (Pallant, 2020). We used IBM SPSS 28 for statistical analyses and IBM SPSS Amos 28 for CFA.

6.3. Results

6.3.1. CFA

The results of the three CFAs of the first model showed a high correlation (0.92) between social stimulation and social interaction for the social network product category (Figure 2-left). According to EFA results, these two dimensions could be represented by a higher-level factor of sociability. Thus, using sociability as a second-order factor, we kept both social interaction and social stimulation in the second model (Figure 2-right). The three latent variables had correlation coefficients above 0.3 but below 0.9, indicating that they were reasonably related, but not multicollinear. Although below the threshold, the correlation coefficient of 0.77 was observed between sociability and social acceptance.

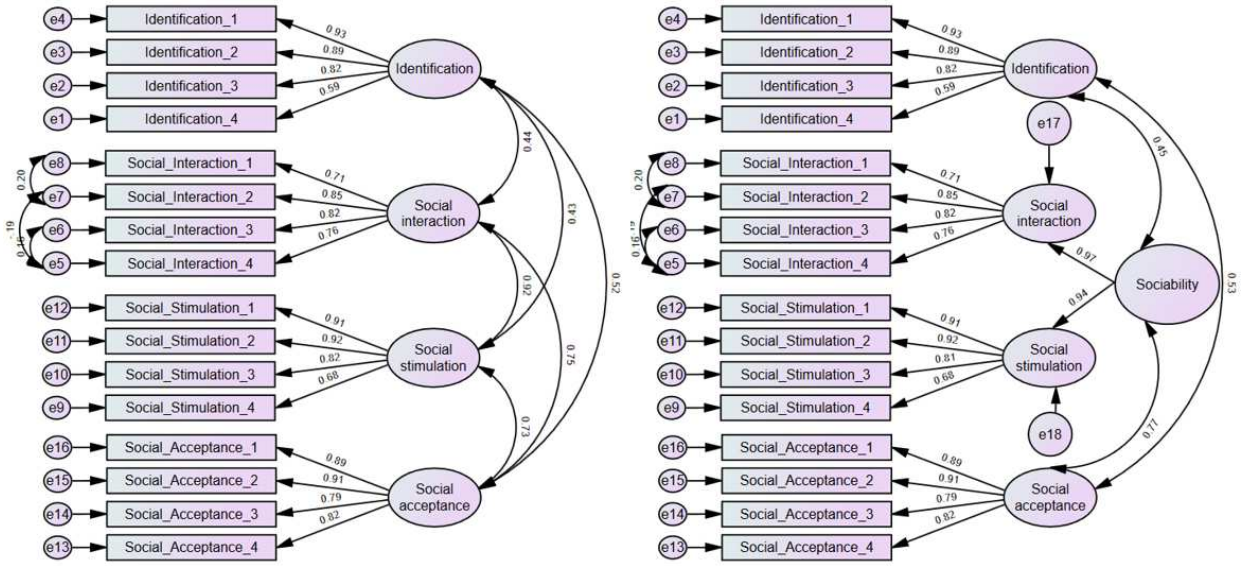


Figure 2. CFA for the first model with four factors (left) and the second model with three factors (right) for social network category.

Table 9 shows the goodness of fit measures calculated for the three CFAs on the second model after the modifications. The thresholds shown in the table are from (Hu & Bentler, 1999; Wang & Wang, 2020). Results confirm the model fit on the data for the three product categories.

Table 9. Goodness of fit measures for the 3 CFAs of the second model

Measure	Threshold	Online banking	Online shopping	Social network
Chi square	NA	154.385	162.808	208.459
Degree of freedom (df)	NA	95	93	96
Chi square/df	1-3	1.625	1.751	2.171
CFI	>0.95 (great), >0.90 (traditional)	0.959	0.956	0.939
TLI	>0.9	0.948	0.943	0.924
RMSEA	<0.5 (good) 0.5-.10 (moderate),>0.1 (bad)	0.066	0.073	0.090

Table 10 shows factor loadings greater than 0.5 for all items of each factor, which is better than the results of the EFAs (Table 6). The average variance extract (AVE) of all factors exceeded the cutoff value of 0.5, confirming the convergent validity of the second model (Hair et al., 2014).

The discriminant validity of the three factors of identification, social acceptance, and sociability was confirmed because maximum shared variance (MSV) was greater than AVE for each factor, and the squared root of AVE (bold in Table 11) was greater than the correlation of the factors (Gefen & Straub, 2005).

Table 10. Factor loadings of the CFAs for the 3 product categories

Factor	Item	Online banking	Online shopping	Social network
Identification	Identification_1	0.630	0.752	0.926
	Identification_2	0.670	0.849	0.890
	Identification_3	0.813	0.768	0.817
	Identification_4	0.784	0.844	0.588
Social interaction	Social_interaction_1	0.753	0.770	0.707
	Social_interaction_2	0.734	0.788	0.850
	Social_interaction_3	0.825	0.754	0.821
	Social_interaction_4	0.732	0.685	0.762
Social stimulation	Social_stimulation_1	0.841	0.910	0.911
	Social_stimulation_2	0.943	0.953	0.920
	Social_stimulation_3	0.653	0.665	0.815
	Social_stimulation_4	0.777	0.588	0.676
Social acceptance	Social_acceptance_1	0.929	0.951	0.891
	Social_acceptance_2	0.935	0.930	0.911
	Social_acceptance_3	0.704	0.716	0.787
	Social_acceptance_4	0.641	0.591	0.822
Sociability	Social interaction	0.993	0.932	0.972
	Social stimulation	0.776	0.872	0.942

Table 11. Model validity measures of the CFAs for the 3 product categories (AVE squared root in bold)

Category	Factor	Social acceptance	Identification	Sociability	CR	AVE	MSV
Online banking	Social acceptance	**0.813**			0.884	0.661	0.360
	Identification	0.600	**0.728**		0.817	0.530	0.360
	Sociability	0.427	0.347	**0.891**	0.884	0.794	0.182
Online shopping	Social acceptance	**0.815**			0.885	0.664	0.394
	Identification	0.562	**0.824**		0.894	0.679	0.316
	Sociability	0.628	0.499	**0.886**	0.880	0.785	0.394
Social network	Social acceptance	**0.854**			0.915	0.730	0.596
	Identification	0.525	**0.816**		0.886	0.666	0.276
	Sociability	0.772	0.454	**0.957**	0.956	0.916	0.596

Table 12 shows the final four social scales with the corresponding items used for the UX evaluation of the three product categories.

Table 12. Developed social scales and their items

Identification								
In my opinion, using/owning this product makes me look:								
Old-fashioned	1	2	3	4	5	6	7	Fashionable
Cheap	1	2	3	4	5	6	7	Stylish
Out of date	1	2	3	4	5	6	7	Modern
Unprofessional	1	2	3	4	5	6	7	Professional
Social interaction								
In my opinion, by using this product I can:								
Isolate from others	1	2	3	4	5	6	7	Communicate with others
Not collaborate with others	1	2	3	4	5	6	7	Collaborate with others
Be ignorant to others	1	2	3	4	5	6	7	Be helpful to others
Not empathize with others	1	2	3	4	5	6	7	Empathize with others
Social stimulation								
In my opinion, the use of this product is:								
Socially discouraging	1	2	3	4	5	6	7	Socially encouraging
Socially disengaging	1	2	3	4	5	6	7	Socially engaging
Socially neutral	1	2	3	4	5	6	7	Socially empowering
Socially exclusive	1	2	3	4	5	6	7	Socially inclusive
Social acceptance								
In my opinion, using this product in social situations makes me feel:								
Judged by others	1	2	3	4	5	6	7	Approved by others
Rejected by others	1	2	3	4	5	6	7	Accepted by others
Embarrassed	1	2	3	4	5	6	7	Proud
Uncomfortable	1	2	3	4	5	6	7	Comfortable

6.3.2. Comparisons

Table 13 shows the correlations between the four social scales developed in this study and the four scales of the AttrakDiff questionnaire for the three product categories. The highest correlations were observed with identification and attractiveness scales, particularly for the social network product category. All the statistically significant correlations are marked with a star in the table and those greater than 0.5 are highlighted in black.

Table 13. Correlations between the social scales and those of AttrakDiff (* significant with P<0.05)

Category	AttrakDiff scales	Identification	Social interaction	Social stimulation	Social acceptance
Social network	Pragmatic	0.331*	0.503*	0.549*	0.426*
	Identification	0.542*	0.648*	0.696*	0.584*
	Stimulation	0.534*	0.390*	0.427*	0.427*
	Attractiveness	0.622*	0.605*	0.665*	0.621*
Online shopping	Pragmatic	0.385*	0.326*	0.178*	0.433*
	Identification	0.502*	0.532*	0.456*	0.514*
	Stimulation	0.476*	0.263*	0.148	0.477*
	Attractiveness	0.521*	0.419*	0.295*	0.586*
Online banking	Pragmatic	0.326*	0.077	0.036	0.385*
	Identification	0.472*	0.347*	0.243*	0.479*

	Stimulation	0.412*	0.206*	0.137	0.401*
	Attractiveness	0.534*	0.188*	0.207*	0.507*

Table 14 shows the correlations between the four social scales developed in this study and the six scales of UEQ+. The four social scales were not redundant given that all the correlation coefficients were below 0.8. The highest correlations were observed with the Value and Attractiveness scales of the UEQ+. However, the social network product category displayed strong correlations with the pragmatic dimensions of Usefulness, Efficiency and Intuitive usage.

Table 14. Correlations between the social scales and the UEQ+ scales (* significant with P<0.05)

Category	UEQ+ scales	Identification	Social interaction	Social stimulation	Social acceptance
Social network	Efficiency	0.445*	0.508*	0.548*	0.457*
	Intuitive usage	0.296*	0.527*	0.555*	0.440*
	Perspicuity	0.276*	0.322*	0.393*	0.334*
	Value	0.694*	0.470*	0.522*	0.495*
	Usefulness	0.461*	0.696*	0.752*	0.619*
	Attractiveness	0.624*	0.658*	0.670*	0.645*
Online shopping	Efficiency	0.421*	0.296*	0.224*	0.404*
	Intuitive usage	0.347*	0.241*	0.169*	0.431*
	Perspicuity	0.339*	0.290*	0.095	0.361*
	Value	0.566*	0.471*	0.322*	0.542*
	Usefulness	0.476*	0.247*	0.206*	0.417*
	Attractiveness	0.544*	0.420*	0.306*	0.608*
Online banking	Efficiency	0.403*	0.043	0.028	0.429*
	Intuitive usage	0.331*	0.005	-0.020	0.456*
	Perspicuity	0.276*	-0.014	-0.033	0.407*
	Value	0.530*	0.206*	0.185*	0.434*
	Usefulness	0.384*	0.131	0.083	0.488*
	Attractiveness	0.459*	0.263*	0.213*	0.589*

Table 15 shows the average importance rating of each social scale. The highest ratings for all social scales went to social network and the lowest to online banking product category, as expected.

Table 15. Average importance rating of social scales for the three product categories (from 1: completely irrelevant to 7: very important)

Category	Identification	Social interaction	Social stimulation	Social acceptance
Social network	4.94	5.92	5.76	5.26
Online shopping	4.69	4.46	4.08	4.09
Online banking	4.02	3.22	3.06	3.27

6.3.3. MANOVA

Before conducting MANOVA to test the social scales ratings differences between product categories, participants who responded to more than one product category were removed yielding three independent samples of 106, 98, and 96 participants for social network, online shopping, and online banking product categories, respectively. Following Mahanalobis distance calculations, multivariate outliers were eliminated such that further analyses were performed on three independent samples of 95 participants.

Results of Shapiro-Wilk test of normality was significant (P<0.001), showing that the data was not normality distributed. However, the sample size was greater than 30 for each independent variable (i.e., product category), which is robust to the violation of normality or equality of variance (Pallant, 2020). The general pattern of scatterplot showed a linear relationship for each pair of dependent variables (i.e., four social scales). Table 16 shows that there was no multicollinearity between dependent variables because none of them had a correlation coefficient larger than 0.9 with another one. Moreover, being greater than 0.2, they were all at least moderately correlated. Box's test of equality of covariance matrices was significant (P <0.001) meaning that the observed covariance matrices of the dependent variables were not equal for all categories. Assumptions testing showed no violations for conducting MANOVA except for normality of data and homogeneity of variance-covariance matrices. As our sample size exceeds 30, violation of normality is not a concern. Moreover, the results of Pillai's trace were used for multivariate test because it is more robust when there is violation of assumptions (Tabachnick & Fidell, 2019).

Table 16. Correlation between the dependent variables (** correlation is significant at the 0.01 level)

Dependent variable	Identification	Social interaction	Social stimulation	Social acceptance
Identification	1	0.350**	0.399**	0.525**
Social interaction	0.350**	1	0.792**	0.485**

Social stimulation	0.399**	0.792**	1	0.480**
Social acceptance	0.525**	0.485**	0.480**	1

The results of one-way MANOVA showed significant difference between the three product categories on the combined social scales, F (8, 560) = 20.68, P< 0.001, Pillai's trace = 0.456, partial eta squared = 0.228. More investigation on each dependent variable showed statistical significance for social interaction and social stimulation using a Bonferroni adjusted alpha level of 0.0125 (highlighted in Table 17). Bonferroni adjustment is used to reduce the risk of a Type 1 error when several separate analyses are performed. Table 17 shows the multiple comparisons between product categories for each social scale.

Table 17. Significance of mean differences of dependent variables for three product categories (*significant at the 0.05 level and highlighted significant at 0.0125 level)

Dependent variable	Category (I)	Category(J)	Mean difference (I-J)	Std. Error	Sig.
Identification	Social network	Online shopping	0.684	0.1477	0.889
	Social network	Online banking	0.360*	0.1477	0.040
	Online shopping	Online banking	0.292	0.1477	0.120
Social interaction	Social network	Online shopping	1.455*	0.1532	<0.001
	Social network	Online banking	1.989*	0.1532	<0.001
	Online shopping	Online banking	0.534*	0.1532	0.002
Social stimulation	Social network	Online shopping	0.981*	0.1581	<0.001
	Social network	Online banking	1.639*	0.1581	<0.001
	Online shopping	Online banking	0.657*	0.1581	<0.001
Social acceptance	Social network	Online shopping	0.086	0.1460	0.823
	Social network	Online banking	0.402*	0.1460	0.017
	Online shopping	Online banking	0.315	0.1460	0.080

7. Discussion

This study covered the design, development, and validation of social scales for the UX evaluation of interactive products. UEQ+ scales used EFA to identify the four most representative items. However, scale development in this study was different in that we also performed CFA on three different samples to provide more validity measures for the social scales. Results showed good

internal consistency (Table 7), convergent validity, and discriminant validity (Table 10 and Table 11) for the social scales. Although we only have a good model fit for the second model, the CFA confirmed our structural model of social dimension with the three main sub-dimensions of identification, sociability, and social acceptance.

Identification, social stimulation, social interaction, and social acceptance are the four social scales that were developed in this study (Table 12). The identification scale is similar to that of AttrakDiff and status dimension of meCUE. It deals with users' personal and social image. Users meet their influence and popularity need by expressing their identity, influence, and power to others through the use or possession of a product. The social interaction and social stimulation scales are complementary. The social stimulation scale represents the social incentives provided by a product to enable users to be socially active. A socially engaging, encouraging, empowering, and inclusive product enables users to achieve their social interaction goals in the form of communication, collaboration, and sharing of emotions with others. Finally, the social acceptance scale entails the feeling of being accepted by others. It can evoke self-conscious emotions like pride that are fulfilled when one is recognized and approved by others.

The social scales were subject to validation studies to find their relationship with other UX dimensions. Identification, the only social dimension of the AttrakDiff questionnaire, was expected to have strong correlations with the four social scales developed in this study. Our results confirmed this expectation by finding correlation coefficients between 0.45 and 0.69 with identification, except for social interaction ($r=0.34$) and social stimulation ($r=0.24$) for the online banking product category (Table 13). In addition to identification, we observed high correlations between the social scales and the attractiveness dimension of AttrakDiff. Hassenzahl (Hassenzahl, 2004) found a strong correlation between identification and beauty. He stated that beauty is social

because it communicates identity and can be shared with and approved by others. Considering that beauty is an item in the attractiveness dimension of AttrakDiff, we can justify finding high correlations of social scales with attractiveness.

In addition to AttrakDiff, we measured the correlations of the four social scales with six UX dimensions of UEQ+. In the AttrakDiff, UEQ, and UEQ+ questionnaires, the attractiveness dimension is used for the overall assessment of a product. Boos and Brau (Boos & Brau, 2017) calculated the correlations of the acoustic and haptic dimensions with the attractiveness dimension of UEQ+ to demonstrate that their new UX dimensions measured the same construct (i.e., UX). They concluded that the strong correlations with attractiveness validated the two new UX dimensions for the UEQ+ modular framework. Following the same approach, we found correlation coefficients ranging from 0.42 to 0.67 between the four social scales and the attractiveness dimension of UEQ+ (Table 14). However, there were three exceptions: social interaction (r=0.34) and social stimulation (r=0.24) for online banking, and social stimulation (r=0.30) for online shopping product category. We observed higher correlations between the social scales and the value dimension of UEQ+. This suggests that social scales are closer in nature to the hedonic dimensions of UX, such as value, than the pragmatic dimensions of efficiency or intuitive usage. However, high correlations with pragmatic dimensions were found in the social network product categories. Other studies have presented the influence of social features on both hedonic and pragmatic dimensions. For instance, (Guo & Li, 2021) found that interactivity, recommendations, and feedback as social features of social commerce platforms have an impact on both utilitarian and hedonic values that ultimately influence consumers' repurchasing intention. The influence of perceived co-experience on enjoyment in social live streaming services (Bründl, Matt, & Hess, 2017) and the impact of social image on the use of

hedonic systems like games (Lin & Bhattacherjee, 2010) are additional examples of social aspects that affect user experience.

Validation of the social scales was tested by evaluating the UX of three product categories. These categories were selected to represent different levels of social aspects. The average importance rating of each scale in the UX evaluation of each product category followed our expectations, with social network having the highest importance ratings and online banking having the lowest across the four social scales (Table 15). Validation with three product categories addressed Lallemand and Koenig's (Lallemand & Koenig, 2017) concerns on standardized scales development performing only a single validation study.

The MANOVA results showed a significant difference between the three product categories for the combination of social scales. Further analysis of each pair of categories revealed significant differences for social interaction and social stimulation (Table 17). The greatest difference was found in the social interaction scale between the social network and online banking product categories. It was to be expected, given that the former is designed as a communication platform, whereas the latter is exclusively task-related and rarely incorporates social interaction aspects. The mean differences for the two scales of identification and social acceptance were not significant for any product category when using a Bonferroni adjusted alpha level of 0.0125. The fact that all of the evaluated products were software applications could have had an impact on the results. It can be expected that the identification and social acceptance dimensions of software products do differ from those of physical interactive products that attract more attention from users and spectators in a social situation, such as virtual reality glasses, or even non-interactive products such as jewelry or designer handbags. However, at the alpha level of 0.05, significant differences

could be observed between social network and online banking, which are at the opposite extremes regarding the social features.

Overall, the four social scales developed in this study can be used in conjunction with other dimensions of the UEQ+ modular framework.

The first limitation of this study is that the selected products were all interactive products. Evaluating other products that can develop social UX such as designer clothes, watches, physical board games, and museum pieces could be considered in future studies. The second limitation of this study is that we relied on the users' reported experience with the selected products. Using laboratory studies and doing specific tasks with better control over the hardware and context could provide more reliable results. However, this would have been possible only at a significantly increased cost with a likely much smaller sample.

8. Conclusion

In this study, we developed and validated four social scales for the UX evaluation of interactive products: identification, social stimulation, social interaction, and social acceptance. These scales are each measured with four items, following the same format as the other scales of the UEQ+ modular framework. They can be used in conjunction with the other dimensions of the framework for the UX evaluation of product with social aspects. Therefore, they have the benefits of UEQ+, such as being freely available, rapid to use, and providing a global score for easier comparisons. We used the results of an online workshop with UX experts to prepare the candidate items for each social scale. Then we did two rounds of data collection, the first used for scale development and the second for calculating reliability and the validity of scales. The scales were then validated through correlation analysis with the dimensions of the AttrakDiff questionnaire and the six UX dimensions of the UEQ+ framework. Results showed a high correlation with the identification and

attractiveness dimensions of these two UX evaluation tools. Finally, using the social scales developed in this study we successfully discriminated among three product categories with different levels of social features. This study highlighted the importance of paying more attention to social dimensions. We recommend that future studies evaluate physical interactive products with social dimensions to better validate the social scales.

Disclosure statement: The authors report there are no competing interests to declare.

Funding: This research was funded by NSERC (Natural Sciences and Engineering Research Council of Canada) Individual Research Grant awarded to Dr. Jean-Marc Robert (#RGPIN-2018-06733).

Reference

Arhippainen, L., & Tähti, M. (2003). *Empirical evaluation of user experience in two adaptive mobile application prototypes.* Paper presented at the Mum 2003. proceedings of the 2nd international conference on mobile and ubiquitous multimedia.

Boos, B., & Brau, H. (2017). Erweiterung des UEQ um die Dimensionen Akustik und Haptik. In S. Hess & H. Fischer (Eds.), *Mensch und Computer 2017 - Usability Professionals*: Gesellschaft für Informatik e.V.

Bründl, S., Matt, C., & Hess, T. (2017). Consumer use of social live streaming services: The influence of co-experience and effectance on enjoyment.

Field, A. (2013). *Discovering statistics using IBM SPSS statistics* (4th ed.): Sage.

Fu, F.-L., Su, R.-C., & Yu, S.-C. (2009). EGameFlow: A scale to measure learners' enjoyment of e-learning games. *Computers & Education, 52*(1), 101-112.

Gefen, D., & Straub, D. (2005). A Practical Guide to Factorial Validity Using PLS-Graph: Tutorial and Annotated Example, Communications of the AIS. In (pp. 91-109).

Guo, J., & Li, L. (2021). Exploring the Relationship Between Social Commerce Features and Consumers' Repurchase Intentions: The Mediating Role of Perceived Value. Frontiers in Psychology, 12, 775056. doi:10.3389/fpsyg.2021.775056

Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2014). *Multivariate data analysis* (7th ed.): Pearson.

Hassenzahl, M. (2003). The Thing and I: Understanding the Relationship Between User and Product. In M. A. Blythe, K. Overbeeke, A. F. Monk, & P. C. Wright (Eds.), *Funology: From Usability to Enjoyment* (pp. 31-42). Dordrecht: Springer Netherlands.

Hassenzahl, M. (2004). The interplay of beauty, goodness, and usability in interactive products. *Human–Computer Interaction, 19*(4), 319-349.

Hassenzahl, M., Burmester, M., & Koller, F. (2003). AttrakDiff: Ein Fragebogen zur Messung wahrgenommener hedonischer und pragmatischer Qualität. In G. Szwillus & J. Ziegler (Eds.), *Mensch & Computer 2003: Interaktion in Bewegung* (pp. 187-196). Wiesbaden: Vieweg+Teubner Verlag.

Hassenzahl, M., Diefenbach, S., & Göritz, A. (2010). Needs, affect, and interactive products–Facets of user experience. *Interacting with Computers, 22*(5), 353-362.

Hinderks, A., Winter, D., Schrepp, M., & Thomaschewski, J. (2019). Applicability of User Experience and Usability Questionnaires. *Journal of Universal Computer Science, 25*(13), 1717-1735. Retrieved from <Go to ISI>://WOS:000508640800004

Hinkin, T. R. (1998). A brief tutorial on the development of measures for use in survey questionnaires. *Organizational research methods, 1*(1), 104-121.

Hu, L. t., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal, 6*(1), 1-55. doi:10.1080/10705519909540118

IJsselsteijn, W. A., de Kort, Y. A., & Poels, K. (2013). The game experience questionnaire. *Eindhoven: Technische Universiteit Eindhoven, 46*(1).

Jang, J., & Yi, M. Y. (2019). Determining and validating smart TV UX factors: A multiple-study approach. *International Journal of Human Computer Studies, 130*, 58-72. doi:10.1016/j.ijhcs.2019.05.001

Jordan, P. W. (2000). *Designing pleasurable products: An introduction to the new human factors*: CRC press.

Kang, K., Hengeveld, B., Hummels, C., & Hu, J. (2022). Enhancing Social Interaction among Nursing Homes Residents with Interactive Public Display Systems. *International Journal*

of Human–Computer Interaction, 38(17), 1701-1717. doi:10.1080/10447318.2021.2016234

Kantosalo, A., & Riihiaho, S. (2019). Quantifying co-creative writing experiences. *Digital Creativity, 30*(1), 23-38. doi:10.1080/14626268.2019.1575243

Klein, A. M., Schrepp, M., Hinderks, A., & Thomaschewski, J. (2020). Measuring User Experience Quality of Voice Assistants Voice Communication Scales for the UEQ plus Framework. In A. Rocha, B. E. Perez, F. G. Penalvo, M. D. Miras, & R. Goncalves (Eds.), *2020 15th Iberian Conference on Information Systems and Technologies*.

Koelle, M., Ananthanarayan, S., & Boll, S. (2020). Social Acceptability in HCI: A Survey of Methods, Measures, and Design Strategies. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (pp. 1–19): Association for Computing Machinery.

Lallemand, C., & Koenig, V. (2017). *"How could an intranet be like a friend to me?" - Why standardized UX scales don't always fit.* Paper presented at the 35th Annual Conference of the European Association of Cognitive Ergonomics, ECCE 2017, September 20, 2017 - September 22, 2017, Umea, Sweden.

Lallemand, C., & Koenig, V. (2020). *Measuring the Contextual Dimension of User Experience: Development of the User Experience Context Scale (UXCS).* Paper presented at the Proceedings of the 11th Nordic Conference on Human-Computer Interaction: Shaping Experiences, Shaping Society.

Laugwitz, B., Held, T., & Schrepp, M. (2008). *Construction and evaluation of a user experience questionnaire.* Paper presented at the Symposium of the Austrian HCI and usability engineering group.

Li, H. (2015). *Enhancing user experience of enterprise systems for improved employee productivity: A first stage of case study.* Paper presented at the 2nd International Conference on HCI in Business, HCIB 2015 Held as Part of 17th International Conference on Human-Computer Interaction, HCI International 2015, August 2, 2015 - August 7, 2015, Los Angeles, CA, United states.

Lin, C.-P., & Bhattacherjee, A. (2010). Extending technology usage models to interactive hedonic technologies: a theoretical model and empirical test. Information Systems Journal, 20(2), 163-181. doi:https://doi.org/10.1111/j.1365-2575.2007.00265.x

Liu, S., Zheng, X. S., Liu, G., Jian, J., & Peng, K. (2013). Beautiful, usable, and popular: Good experience of interactive products for Chinese users. *Science China Information Sciences, 56*(5), 1-14. doi:10.1007/s11432-013-4835-4

Minge, M., Thuring, M., Wagner, I., & Kuhr, C. V. (2017). *The meCUE questionnaire: A modular tool for measuring user experience.* Paper presented at the International Conference on Ergonomics Modeling, Usability and Special Populations, AHFE 2016, July 27, 2016 - July 31, 2016, Walt Disney World, FL, United states.

Mortazavi, E., Doyon-Poulin, P., Imbeau, D., Taraghi, M., & Robert, J.-M. (2021). *Systematic Literature Review on UX dimensions and UX subjective evaluation tools*. Manuscript submitted for publication.

Pallant, J. (2020). *SPSS survival manual: A step by step guide to data analysis using IBM SPSS* (7th ed.): Routledge.

Park, J., Han, S. H., Kim, H. K., Cho, Y., & Park, W. (2013). Developing elements of user experience for mobile phones and services: Survey, interview, and observation approaches. *Human Factors and Ergonomics In Manufacturing, 23*(4), 279-293. doi:10.1002/hfm.20316

Petri, G., von Wangenheim, C. G., & Borgatto, A. F. (2016). MEEGA+: an evolution of a model for the evaluation of educational games. *INCoD/GQS, 3*, 1-40.

Petrosyan, A. (2023). Worldwide digital population 2023. Retrieved from https://www.statista.com/statistics/617136/digital-population-worldwide/

Phan, M. H., Keebler, J. R., & Chaparro, B. S. (2016). The development and validation of the game user experience satisfaction scale (GUESS). *Human Factors, 58*(8), 1217-1247.

Qiu, S., Hu, J., Han, T., Osawa, H., & Rauterberg, M. (2020). An Evaluation of a Wearable Assistive Device for Augmenting Social Interactions. IEEE Access, 8, 164661-164677. doi:10.1109/ACCESS.2020.3022425

Robert, J.-M. (2014). *Defining and structuring the dimensions of user experience with interactive products.* Paper presented at the 11th International Conference on Engineering Psychology and Cognitive Ergonomics, EPCE 2014, Held as Part of 16th International Conference on Human-Computer Interaction, HCI International 2014, June 22, 2014 - June 27, 2014, Heraklion, Crete, Greece.

Robert, J.-M., & Lesage, A. (2011). Designing and evaluating user experience. In G. A. Boy (Ed.), *The Handbook of Human-Machine Interaction: A Human-Centered Design Approach* (pp. 321-338). UK: Ashgate.

Ryu, H., & Kim, J. (2019). Evaluation of User Experience of New Defense Medical Information System. *Healthcare Informatics Research, 25*(2), 73-81. doi:10.4258/hir.2019.25.2.73

Safsouf, Y., Mansouri, K., & Poirier, F. (2019). *Design of a new scale to measure the learner experience in e-learning systems.* Paper presented at the International Conference on e-Learning 2019, EL 2019, July 17, 2019 - July 19, 2019, Porto, Portugal.

Schrepp, M., & Thomaschewski, J. (2019). Design and Validation of a Framework for the Creation of User Experience Questionnaires. *International Journal of Interactive Multimedia & Artificial Intelligence, 5*(7).

Sun, X., & May, A. (2013). A comparison of field-based and lab-based experiments to evaluate user experience of personalised mobile devices. *Adv. in Hum.-Comp. Int., 2013*, Article 2. doi:10.1155/2013/619767

Tabachnick, B. G., & Fidell, L. S. (2019). *Using Multivariate Statistics* (7th ed.): Pearson.

Tcha-Tokey, K., Christmann, O., Loup-Escande, E., & Richir, S. (2016). Proposition and Validation of a Questionnaire to Measure the User Experience in Immersive Virtual Environments. *International Journal of Virtual Reality, 16*(1), 33-48. doi:10.20870/IJVR.2016.16.1.2880

Wang, J., & Wang, X. (2020). *Structural equation modeling: Applications using Mplus* (2nd ed.): John Wiley & Sons.

Washington, P., Voss, C., Haber, N., Tanaka, S., Daniels, J., Feinstein, C., Winograd, T., Wall, D. (2016). A Wearable Social Interaction Aid for Children with Autism. Paper presented at the Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems, San Jose, California, USA. https://doi.org/10.1145/2851581.2892282

Weiss, B., & Willkomm, S. (2013). *User experience and social attribution for an embodied spoken dialog system.* Paper presented at the Workshop on Computers As Social Actors, CASA 2013 - Co-located with 13th International Conference on Intelligent Virtual Agents, IVA 2013, September 28, 2013, Edinburgh, United kingdom.

Yoon, J., Kim, C., & Kang, R. (2020). Positive User Experience over Product Usage Life Cycle and the Influence of Demographic Factors. *International Journal of Design, 14*(2), 85-102. Retrieved from <Go to ISI>://WOS:000576790900006

Zarnadze, G. (2020). Social interactions impact on product and service development. Proceedings of the International Conference on Business Excellence, 14(1), 324-332. doi:10.2478/picbe-2020-0031

Zhu, Y., Dailey, S. L., Kreitzberg, D., & Bernhardt, J. (2017). "Social Networkout": Connecting Social Features of Wearable Fitness Trackers with Physical Exercise. Journal of Health Communication, 22(12), 974-980. doi:10.1080/10810730.2017.1382617

About the authors

Ehsan Mortazavi is a PhD candidate in the Department of Mathematics and Industrial Engineering at Polytechnique Montreal. His research focuses on UX subjective evaluation tools, UX dimensions, and cognitive ergonomics.

Philippe Doyon-Poulin is an assistant professor in the Department of Mathematics and Industrial Engineering at Polytechnique Montreal. His research focuses on human factors in aviation, error prevention and decision-making with automated systems. He authored more than 25 aircraft certification reports to show compliance on flight deck usability and pilot error.

Daniel Imbeau, Eng, PhD is a Full professor in the Department of Mathematics and Industrial Engineering at Polytechnique Montreal. His research focuses primarily on occupational ergonomics, continuous improvement, and human-centered engineering design to improve worker health & safety at work, and productivity.

Jean-Marc Robert, Doctorate, is an Adjunct professor in the Department of Mathematics and Industrial Engineering at Polytechnique Montreal. He retired in 2020. He taught and conducted research in the fields of Cognitive Ergonomics, Prospective Ergonomics, User Experience, and Human-Computer Interaction.