Initial face validation study of an inventory to assess Human-Robot Collaboration User Experience

Ásthildur Lilja Stefánsdóttir

University of Twente

Faculty of Behavioural, Management and Social Sciences

Master's thesis (25 EC) in Human Factors and Engineering Psychology

1st Supervisor: Dr Simone Borsci

2nd Supervisor: Dr Cesco Willemse

February 2nd, 2024

Abstract

Robotics are an ever-growing industry in our world. A field in robotics is Human-Robot Collaboration (HRC). It is a term mainly used in an industrial context and its objective is for both humans and robots to combine their strengths for a successful collaboration. An important factor to consider in HRC is the User eXperience (UX) of the collaborative robot (cobot) to assure that the robot is usable and safe and meets the user’s standards. However, HRC is a field that lacks standardized validation tools. Previous research by Prati et al. (2022) and Borsci et al. (2024) identified 15 dimension to consider when assessing the UX of a cobot. These dimensions were assessed by robotics experts and resulted in a consensus of 10 of the 15 dimensions, whereas 5 of them had much disagreement and needed further research. Subsequently, this thesis consisted of two studies, conducted in parallel: Firstly, a consensus study to further evaluate the 5 problematic dimensions. A panel of robotics experts evaluated the 5 dimensions, giving similar results to the previous study, there is much disagreement about the importance of these 5 dimensions. The problematic dimensions will be retained for the scale as optional dimensions that the researcher can choose to use if they deem the factors important. The second study consisted of an item generation of the UX assessment scale, resulting in 71 items, measuring the 15 dimensions. A card sorting study was then conducted where the item pool was reduced to 44 of the most representative items. After the reduction of items, a cluster analysis of the 15 dimensions was performed, leading to the removal of one dimension and one additional item. The scale in its first edition consists of 40 items measuring 14 dimensions. This scale is the first step towards a validated measurement tool that assesses the UX of HRC and is therefore a valuable addition to the scientific community.

Table of Contents

Introduction

Robotics, HRI and HRC

The field of robotics is an ever-growing industry with increased demand (Ahmed & Hossain, 2019), in Western societies it becomes increasingly common to have some sort of robot interaction daily, e.g., a self-checkout machine at your local grocery store, the Amazon Alexa you use to play music or even the Roomba robot that cleans your house's floors. Moreover, working with robots can lead to increased efficiency and productivity in various industries (Sherwani et al., 2020; Rai et al., 2023). Human-Robot Interaction (HRI) takes place when humans and robots interact through some sort of communication. Specifically, the domain of HRI involves designing, evaluating, and understanding various robotic systems (Murphy et al., 2010). A branch of HRI is Human-Robot Collaboration (HRC). HRC is a term that is mainly used in an industrial context (Berg & Lu, 2020) and it takes place when humans and robots interact or work together in a collaborative space (Matheson et al., 2019; Prati et al., 2021). Therefore, the objective of HRC is for both the human and the robot to contribute their strengths to different parts of a task, thereby complementing each other (Kruger et al., 2017, as cited in Kolbeinsson et al., 2019). Namely, Müller et al. (2017) proposed four categories to look at interaction exchanges between humans and robots:

i) Autarkic/Coexistant: The human and robot are allotted separate workspaces but are in the same environment (e.g., are not separated by a fence or something similar).

ii) Synchronised: The human and robot share a workspace but only one of them can operate in the workspace at a time.

iii) Cooperation: the human and the robot are in the same environment simultaneously working on their separate tasks.

iv) Collaboration: the human and the robot are working on the same task together.

In line with the categorisation above, when robots are designed to perform shared interaction tasks with humans (cobots), we enter the domain of HRC. Cobots' integration into the production industry has been transformative and such systems have been a key part of the fourth industrial revolution (George & George, 2023). Cobots aim to assist humans in tasks that are not easily completed by humans alone (Vicentini, 2020). Compared to their industrial counterparts, the cobots are lighter, more flexible, and easier to move around in various spots of the factories they are located in. Therefore, they are often adaptable and can take part in various tasks which makes them suitable for many different industries (Sherwani et al., 2020). A critical distinction between traditional robots and cobots lies in their interface (El Zaatari et al. 2019). Consequently, the cobot user interface is a crucial factor as it facilitates communication between the user and the cobot (Sriviboon & Jiamsanguanwong, 2022). It is not always guaranteed that the users of the cobot possess expertise in programming, which indicates the user interface needs to go through careful usability evaluation ensuring its ease of use and safety to non-experts (Sriviboon & Jiamsanguanwong, 2022).

User experience, usability, and safety

As technological advancements in our society prioritize enhancing user experiences, the field of user experience and the creation of optimal user interactions become increasingly important (Hassenzahl, 2014; Duarte et al., 2022). The standard definition of User eXperience (UX) is in ISO 9241 – 210 as "A person's perceptions and responses that result from the use and/or anticipated use of a product, system or service." (ISO, 2010). As stated in the definition, UX is the sense of satisfaction a person feels when using a certain product or system. Therefore, it can be challenging to guarantee a specific user experience because of its subjective nature, but a well-designed interaction that is tailored to users and the context of use can positively influence the overall satisfaction and experience for the user (Lindblom &

Andreasson, 2016). To be able to create a positive experience for the user, human-centred design needs to remain a focal point in the design process (Lindblom & Andreasson, 2016).

Usability and perceived safety are some of the aspects considered crucial when evaluating the UX of robots (Buchner et al., 2012). Undoubtedly, safety is the first thing that should be considered when it comes to the requirements a robot should meet, as it is the most important thing not to cause any harm to humans (Rubagotti et al., 2022). The construct of safety can be divided into two factors: physical safety and perceived safety (Akalin et al., 2022). In HRI, the physical safety of the robot must be guaranteed to prevent either physical or psychological harm to the person using the robot or people in its vicinity (Pervez & Ryu, 2008; Akalin et al., 2022). Furthermore, in addition to being physically safe, a robot should be *perceived* as safe (Rubagotti et al., 2022). To illustrate, Bartneck et al. (2008) defined perceived safety in the field of HRI as "the user's perception of the level of danger when interacting with a robot, and the user's level of comfort during the interaction." (Bartneck et al., 2008, p. 76). Factors that affect the users' sense of perceived safety can be characteristics of the robot, such as its appearance (e.g., the size or shape of it), how it moves (e.g., the speed of movement), and how close it comes to the person (Akalin et al., 2022). Additionally, sense of control, comfort and trust are key factors when it comes to evaluating the perceived safety of robots, which need to be considered in the design processes of robots (Akalin et al., 2022). Essentially, when designing robots, the safety of the robots transcends pure physical safety; additionally, it includes perceived safety which takes the trust, the feelings of control over the robot and the dangers that the user perceives into consideration.

Usability is defined in ISO 9241 - 211 "The effectiveness, efficiency and satisfaction with which specified users achieve specified goals in particular environments" (ISO, 2010). Firstly, effectiveness pertains to the number of errors and how many activities were completed successfully. Secondly, efficiency can be measured in factors such as: time on task, mental workload, and the physical effort required (Chacón et al., 2021). Lastly, satisfaction is mentioned. However, this is a more subjective construct, as it refers to the degree to which users find the product acceptable in fulfilling their objectives or goals (Jordan, 1998). This feeling of satisfaction can vary more from person to person but a widely used approach in satisfaction assessment is with subjective questionnaires (Chacón et al., 2021). Thus, many measurement tools exist to measure usability such as the Usability Metric for User Experience (UMUX; Finstad, 2010), the System Usability Scale (SUS; Brooke, 1996) and the Computer System Usability Questionnaire (CSUQ; Lewis, 1995) to name a few. Measuring usability is important to be able to identify usability problems, in which areas the system or product is not performing as it should (Bevan et al., 1991). Concludingly, measuring how easy and satisfying it is to use something aids in identifying problems leading to a more user-friendly design.

The UX design process can be summed up in the UX wheel proposed by Hartson and Pyla (2012). It is presented as four steps circular process: analyse (understanding the user's needs and how the user works), design (creating the look and feel of the product or system), implement (creating a working prototype, high fidelity, or low fidelity) and evaluate (employing evaluation methods, evaluating if the meets the goals of the product). An easy and cost-effective way to evaluate UX is with questionnaires and rating scales (Hinderks et al., 2019).

Creating an inventory

Different approaches can be adopted when creating an inventory. Borsci et al. (2024) suggested a five-step approach to go from an initial theoretical construct/framework to scale development which this study adhered to:

1. Framework definition: Create an overview of the aspects you want to measure. This can be done using already existing theories, conducting a systematic literature review and/or recruiting experts to review the topic at hand for a consensus analysis e.g., the Delphi method.
2. Item generation and scale specifications: Once the dimensions and constructs have been decided and defined the specifications of the scale should be decided such as, how many items should the scale include, should it be answered in a continuous format (e.g., a Likert-scale) or a categorical format (yes/no). The items that are created should assess the dimensions of the scale.
3. Review of items: When the item list has been formed, it is the time to assess content validity. This can be achieved by consulting experts or doing a pilot test of the items. It is ideal in this process to create a large item pool to be able to choose the best items. The next step would be to reduce the item pool, for example by conducting a card sorting study.
4. Item evaluation: After the item review, the next step is an analysis. Analytic methods can be, for instance, exploratory factor analysis (EFA: it measures which items are necessary to measure what you are evaluating and how they group together into dimensions) and confirmatory factor analysis (CFA: it confirms if the items really measure what they intend to measure and if they belong to the groups identified in the EFA). This process is done by letting people interact with the system being evaluated and filling in the scale.
5. Scale length revision: Once the scale has been validated, it can be shortened by removing items that are not performing as they should.

As mentioned above, a way to establish a consensus on a theoretical construct for scale development is using the Delphi method. A Delphi study is a method of gathering information about a specific topic from a panel of experts. In theory, it can be conducted as often it is needed until a consensus has been established among the expert panel (Hsu & Sandford, 2007). To illustrate, Loo (2002) proposed five major characteristics that sum the Delphi method up:

1. A panel is selected, consisting of carefully selected experts that have a diverse set of opinions on the topic at hand.
2. Respondents are typically kept anonymous.
3. The researcher/team of researchers develop a series of questionnaires for the experts regarding the subject matter, concurrently providing feedback reports throughout the study duration.
4. The process is usually conducted in three or four rounds of questionnaires.
5. A research report is written with the results of the Delphi process.

Another method frequently used in scale development is card sorting. It is used to establish validity in scale development (Cohen et al., 2018; Dong et al., 2016; Singh et al., 2016), specifically face validity (Beerlage-de Jong et al., 2020). It is a method used to empirically access a person’s mental model of a certain concept (Schmettow & Sommer, 2016). When using the card sorting method, the stimuli are shown on cards, the participant is then asked to sort similar stimuli together either based on pre-existing groups or just as they feel appropriate (Lantz et al., 2019). The sorting can be performed as an open card sort or a closed card sort. Firstly, an open card sorting method asks the participants to group the cards together in a way they deem appropriate. Secondly, a closed card sorting method asks the participant to sort the cards into a pre-established set of groups (Spencer & Warfel, 2004). The following study will use a closed card sorting method. Furthermore, card sorting can be conducted both as an in-person activity or as an online activity. It has been shown that the methods are equally effective (Bussolon et al., 2006). Finally, when the card sorting has been completed the data is analysed, which can be done in a couple of ways, such as factor analysis, cluster analysis or a proximity matrix (Lantz et al., 2019; Beerlage-de Jong et al., 2020). In summary, card sorting, whether conducted in person or online, can provide valuable insights into how people think and the data that come from card sorting tasks can be a helpful tool in scale development.

Previous phases of the research

Two previous phases have been conducted as a part of this ongoing research. Firstly, a systematic literature review of UX evaluation scales for HRC was published by Prati et al. (2022). Their research evaluated 95 papers that related to the use of scales for evaluation of HRI. The literature review demonstrated that while these studies provided valuable insights, a notable gap persisted; namely a lack of standardized evaluation tools specifically designed for assessing UX in industrial HRC scenarios. Therefore, this gap signified the need for an inclusive and comprehensive UX evaluation scale for the industrial HRC domain. This literature review contributed to closing the identified gap by proposing an initial list of 40 aspects that were often associated with the design or use of a cobot for the creation of a new scale.

The next phase of the research was conducted by Borsci et al. (2024). It began with an expert brainstorming session evaluating and adding to the 40 aspects identified by Prati et al. (2022). This led to a final list of 85 aspects that are deemed important in the UX assessment of cobots. These 85 aspects were then grouped into 15 dimensions that can be considered relevant to evaluate the User Experience in Human-Robot Interaction/Collaboration (See Table 1).

Table 1. The 15 dimensions along with their descriptions.

Dimension name	Dimension description
1. Easiness of robot regulation.	The easiness of the robot's physical regulation (e.g., robot's components positioning).
2. Robot physical appearance.	How the physical features of the robot can affect the user's judgment. In particular, the dimension considers aspects such as e.g., Level of Anthropomorphism (e.g., Machinelike, Humanlike), Dimension of the robot (i.e., High, Width, Length, Weight), Type of robot (e.g., robotic arm, humanoid robot), Form and Material, Perceived robustness.
3. Robot's emotional appearance.	How the robot's physical and behavioral characteristics delineate the "robot's emotional profile" and how it can affect the user's judgment. It considers e.g., Robot's Likeability (e.g., happy, kind), Warmth (e.g., social, friendly), Disturbance (e.g., creepy, scary), Discomfort (e.g., awkward, dangerous), Attractiveness.
4. Robot's competence features.	The user’s judgment of the robot's competencies (e.g., reliability, responsiveness) and perceived intelligence (e.g., knowledgeable, responsible) based on its behavior during the interaction.
5. Robot's physical behavior.	The user’s judgment of the robot’s physical behavior during the interaction, considering parameters such as, e.g., Movement mode (e.g., rigid, elegant), Autonomy (e.g., no autonomy, full autonomy), Noise produced while it is moving, Adaptability, Animacy (e.g., alive, natural), Interactivity (e.g., no causal behavior, fully causal behavior).
6. Robot's social behavior.	The user’s judgment of the robot’s social behavior considering parameters such as e.g., Companionship, Initiative (e.g., not giving orders, not being intrusive), Social relationship (e.g., telling its story, having a real exchange of opinion), Social norms (e.g., no knowledge, full knowledge), Communication.
7. Robot task performance.	The user’s judgment of the robot during a specific performance, considering the efficiency (e.g., time on task), Effectiveness (e.g., task completeness, number of errors), and Utility.
8. Human judgment before the interaction with a cobot.	The user's perception of the robot before the interaction, based on. Perception and effect, anxiety (e.g., toward communication capability, toward behavioral characteristics), Attitudes toward use, Expectation (e.g., performance expectancy, effort expectancy), Acceptance, Perceived safety (e.g., speed), Trust (e.g., Reliability), Intention to use.
9. Human judgment of the performance with a cobot.	The user judgment of the robot during a specific performance task, considering. This includes aspects such as e.g., Acceptance, Perceived Safety, Trust, Control (e.g., the robot always listens), Comfort, Intention to use again, Enjoyment (e.g., pleased, bored), Satisfaction, Usability, Frustration, Stress, Cognitive workload.

10. Human-Factors personality-based.	The user's self- description regarding their own personality characteristics, like e.g., ethics (e.g., social impact, social acceptance), Personality traits, Self-confidence, and Personality to trust.
11. Human-Factors ability-based.	The user's self- description regarding their own work characteristics, like e.g., self-efficacy (e.g., a robot setup, technology familiarity), Expertise, and Competence.
12. Task performed.	The characteristics of the specific performed task during the interaction e.g., Type of task, Perceived usefulness of the robot, Physical effort, Task difficulty, and Task criticality.
13. The environment of interaction.	The specific characteristics of the environment where the task was performed. This dimension refers, for instance, to the Workstation layout, Workstation elements, Environment aspects (e.g., illumination, noise, dust), and Application context (e.g., industry, healthcare).
14. Team involved during the task performance.	The members involved in the specific task performed, considering e.g., the number of humans and robots, Members' roles.
15. Interaction aspects.	The interaction aspects of the specific performed task, in terms of, for instance, knowledge of the robot's status, Situation awareness (e.g., feedback), Functionality, Ease of use, Learnability, Memorability, Interface type (e.g., physical-based interface, graphical-based interface, vocal-based interface, gesture-based interface)

The next step in the research was a Delphi study where 81 experts in the field of robotics evaluated the 15 dimensions and their importance in a UX scale evaluating HRI/HRC. The survey led to a common consensus regarding 10 of the dimensions. Furthermore, the survey revealed disagreement by the expert panel on the five dimensions (D3, D6, D8, D10 and D14) indicating that they need further research.

Research goals

For this thesis, two phases were conducted in parallel. Firstly, a consensus study was performed, it involved connecting to experts and asking them to evaluate the five problematic dimensions that were identified by Borsci et al. (2024). This study was an important step in evaluating if the problematic dimensions should be retained for the scale or if they should be removed, therefore excluded from the evaluation tool being developed.

The second study consisted of creating items for the 15 dimensions, one item per aspect that each dimension measures, resulting in a large group of items. These items were then evaluated in a card sorting study, where participants were asked to attribute an item to a certain dimension. By doing this the most representative items were retained for the evaluation tool. Finally, a cluster analysis was conducted to identify which items were frequently clustered together to see if the items were appropriate for the dimensions.

Phase 1 – Consensus study

Method

Participants

Prior to participant recruitment, ethical approval was granted by The University of Twente ethics committee (request nr. 231176). A total of 27 people participated in the study, however 5 participants did not finish the questionnaire in its entirety. All the participants were experts in the field of HRC. Among them were 5 women (23%) and 16 men (73%) and 1 person who preferred not to answer (4%). The participants had a varying level of experience working in the field of HRC: two participants had less than one year of experience, 10 participants had 1-5 years of experience, eight participants had 5-10 years of experience and two participants had more that 10 years of experience. Furthermore, most of the participants were researchers, along with Human-factors specialists, a software developer, and a robot user (i.e., they use the robot for its final scope, e.g., assembly, physical support). They had experience from various domains of HRC, both in industrial domains (industry and warehouse) and commercial domains (cobots for healthcare, domestic, social, education, agriculture). The participants were resided in various countries, Italy, Denmark, Jamaica, Mexico, Norway, Portugal, Spain, Greece, France, and the USA.

Recruitment strategies involved the research team reaching out to their networks asking for participation. Participants from the previous research phase that had agreed to be contacted again were contacted via email. 14 participants had taken part in the first phase of the research, 6 had not and 2 preferred not to say.

Materials

The study comprised a survey constructed in Qualtrics. The survey was adapted from the survey used in the previous research phase (Borsci et al., 2024). The survey aimed to delve deeper into expert opinion on the five dimensions that were deemed problematic in the previous research phase, why they could be considered problematic and if there is a reason to remove the dimensions from the UX scale. The survey required a consent form to participate, five rounds of questions about each of the dimensions that were deemed problematic, three questions each round, one mandatory question asking the participants if they agreed with the results from the previous survey, that the dimension should be removed. This mandatory question was measured on a nine-point Likert-scale (1; strongly disagree — 9; strongly agree). Additionally, two optional open-ended question were presented for each dimension, one asking why the experts thought there was disagreement with the dimension, whereas the second open-ended question asked how the experts thought the dimension description could be improved (see Appendix A for exact phrasing).

The final section of the questionnaire assessed participant characteristics relating to their work within the domain of HRC and general demographic questions. The full survey can be found in Appendix A.

Procedure

Participants opened the survey via a link they received through email or through their network. The survey opened to a front page where information on the purpose of the research, the results from the previous research, instructions for the survey (see Appendix A), the expected time spent on the survey (15-20 mins) and information on how the participant data would be used (only used for a scientific purpose, the data will be kept anonymous in a secure server) as well as the names and email addresses of the research team. The participant was asked to consent to participation in the study, if consent was not given, they were sent to the end of the survey. The next step was answering the survey questions about the five dimensions, shown in a random order. When the questionnaire was completed, the participant was presented with a closing page where they were thanked for their participation along with the email of two members of the research team, allowing participants to contact them for more information or if any further clarification was needed.

Data analysis

The data analysis was conducted in Microsoft Excel and R Studio. The first step in the data analysis was cleaning the data by removing the participant data from those that did not finish giving their opinions on the five dimensions. Data was retained for participants that gave their expertise on the five dimensions but did not answer the questions pertaining participant information.

The scores of whether the experts agreed with the statement that each of the five dimensions should be removed from the scale, ranged from 1-9. The mean score was calculated along with the standard deviation. Subsequently, the percentage of participants advocating for removal was then computed, by giving the experts who scored higher than 5 (indicating agreement for removal) a value of one, while scores of 5 or lower were assigned a zero.

Lastly, the interquartile range (IQR) was calculated using R Studio, indicating the variability of the answers between the experts.

The criteria for removing the dimension were the following: an IQR lower than 2 (indicating that there is little variability among the experts) and an agreement for removal higher or equal to 75%.

Results

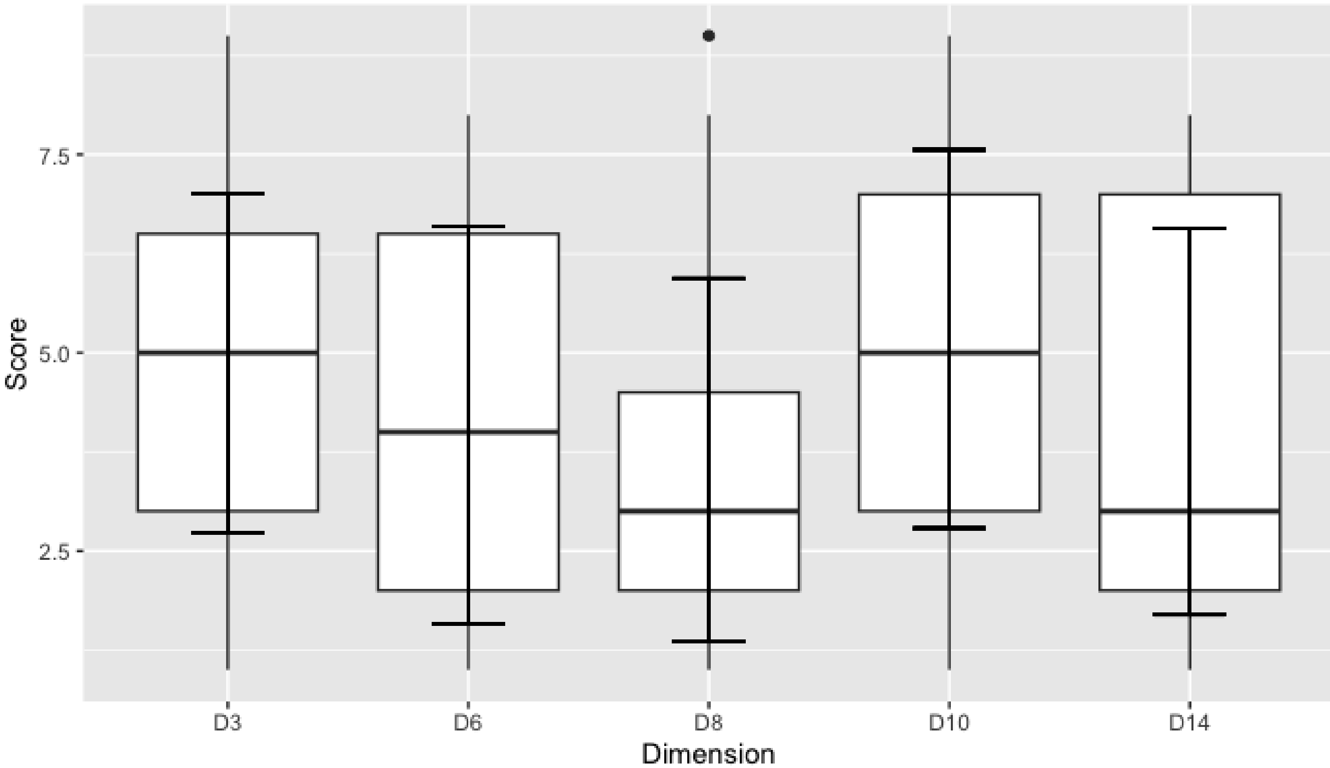
We asked experts to agree on the idea of removing the five problematic dimensions. Table 2 presents the descriptive statistics of the data, along with the agreement of experts regarding the 5 dimensions that experts were not able to agree upon.

Table 2. Expert agreement on the 5 dimensions that experts in a previous studies were not able to agree on.

Dimensions	Mean	SD	Median	IQR	% of agreement to remove
3. Robot's emotional appearance	4.87	2.09	5	3.5	39%
6. Robot's social behavior	4.23	2.41	4	4.5	30%
8. Human judgment before the interaction with a robot.	3.73	2.26	3	2.5	22%
10. Human-Factors personality-based.	5.27	2.34	5.5	4	48%
14. Team involved during the task performance	4.23	2.39	3.5	5	35%

The results suggest moderate disagreement among experts and none of the dimensions meet the criteria for removal (IQR lower than 2, over 75% agreement of removal). As observed, the average scores across all dimensions fall closely around the mid-point (5), indicating a moderate disagreement regarding the potential removal of these dimensions. However, there is a notable variation in the expert opinions, as indicated by the standard deviation and IQR, which adds complexity to this assessment. See Figure 1 for data visualisation.

Figure 1. Boxplots indicating the responses along each of the five dimensions.



The boxplots in Figure 1 show the five dimensions; the boxes in the boxplot show the range of the IQR and the line in each box represents the median of each dimension. The lines around the boxplots represent the error bars, indicating the size of the standard deviation. As can be in the boxplot, both the error bars and the IQR cover a quite large area for each dimension, indicating that there is much variability in the data, highlighting that the experts seem to differ in whether they thought that the dimension should be removed or not.

The survey had two open-ended questions for each dimension which asked the participants why they thought there was disagreement within the dimension or how they would suggest the dimension could be rephrased:

Dimension 3, Robot's emotional appearance

The answers to the optional question gave some insights into the experts' opinions on this dimension:

Participant 10: "Honestly talking, you can not ignore this dimension even 1 percent! If anyone follows up these days articles in the field of HRI, especially in the field of social robotics, can not ignore the emotional profile of a robot…"

Participant 4: "Can not be ignored as we move into real human-robot collaboration."

Participant 1: "Sometimes it is felt that, in specific contexts, robots' appearance is irrelevant for the purpose. However, it is well established that acceptance and trust are strongly related to their appearance"

Participant 14: "Several studies highlight controversial results on the "adequate" level of anthropomorphism of the robotic system. I agree that the discussed features have a strong impact on the user. However, for an industrial context, overdoing human-like features, especially if they are aesthetic or communicative, should be avoided. Based on this assumption, I believe that experts want to avoid supporting this dimension."

The arguments that the experts bring up are quite varied, participants 10 and 4 emphasize that this dimension cannot be ignored, indicating they do not agree with the removal of this dimension. However, participants 1 and 14 both mention that the importance of the dimension is context dependent, while participant 14 highlights that the dimension might be unnecessary in an industrial context.

Dimension 6, Robot's social behavior

The expert opinions on the disagreement of this dimension were the following:

Participant 1: "As for the appearance, in specific contexts, social competencies of robots might appear irrelevant."

Participant 2: "…Additionally, interest in social robot behaviour may not be relevant in every aspect of HRI / HRC situation. E.g., it's more relevant for service robots / social robots compared to industrial collaborative robots, which often is more viewed as a tool compared to a companion."

Participant 4: "Can not be ignored as we move into real human-robot collaboration."

Participant 13: "Robot's social behaviour is an important dimension. It is particularly relevant towards robots with social roles such as robots with virtual assistants and companion robots"

Participant 14: "I think these issues are very important … I do not know why experts did not consider this dimension as relevant."

Many of the experts bring up that this dimension is context specific, it is not relevant in every case of HRI/HRC, it might be more relevant in the interaction with service robot to industrial collaborative robots. Both participant 4 and 14 highlight the dimensions importance, indicating it should not be removed.

Dimension 8, Human judgement before the interaction with the cobot

The expert opinions on the disagreement of this dimension were the following:

Participant 1: "Lack of knowledge in the field if I may say. Stereotype activation is well established when it comes to HRI, and it heavily affects the user's intention to use or trust a robot (e.g. I think the robot is bad, I wont use it)"

Participant 3: "Not really sure but it is a crucial aspect. Human attitude towards the adoption of robot based technologies is of crucial importance and therefore, working towards providing friendly UX approaches is very much needed"

Participant 6: "again there are differences between professional interaction and service interaction…. Expectations and judgement in advance are very important in all these cases, but need to be adjusted from case to case."

Participant 13: "Expectation and perceived safety are the most likely parameters to emotionally skew a user's evaluation of HRI/HRC. This factor should be very relevant to the assessment of UX in HRI/HRC"

The experts highlighted the significance of human judgement before the interaction in HRC. They emphasized that preconceived notions, stereotypes, and perceived safety significantly influence users' intentions and evaluations of robot-based technologies. While acknowledging differences between professional and service interactions, they underscored the importance of adjusting expectations case by case.

Dimension 10, Human-Factors personality based

The expert opinions on the disagreement of this dimension were the following:

Participant 3: "It opens the door to high subjective considerations, whether some personality features may be relevant to be considered."

Participant 10: "Some personality traits of end-users such as affinity to technology can form the users' experience of using a new technology. So it can not be ignorable."

Participant 13: "I believe it impacts the assessment of the UX, as more trusting person will naturally be more at ease with a robot than an insecure person. However, this dimension seems to be less relavant than most other proposed dimensions"

Participant 6: "In case of professional robot the best behaviour of the robot is a predictable one. personality characteristics are not considered in relation to hardware tools and are thus no use. In case of non-professional end-users of service collaborative robotics, some aspects might get relevant, in particular Self-confidence, and Personality to trust in order to evaluate if the UX evaluation might be jeopardized because of fear of the technology."

The experts pointed out that traits like comfort with technology and trust levels significantly influence how people interact with robots. While some highlighted the impact of these traits on user experience and their potential role in determining compatibility between humans and robots, others suggested that this dimension might be less critical compared to others. Furthermore, participant 6 emphasizes that the importance of this dimension might depend on the type of robot being used, i.e., this dimension might be more relevant when assessing service robots.

Dimension 14, Team involved during the task performance

The expert opinions on the disagreement of this dimension were the following:

Participant 2: "Perhaps a lack of understanding of how much the environment/group dynamics influence the interaction with robots. Typically, research on HRI / HRC only consider 1 person interactions, which in many cases does not resemble real world environments in which the robot needs to be placed."

Participant 10: "Simply, when we are talking about a user and the experience of a user, we think of the experience of only one person and do not think about a team of users. So I think it's better to omit this dimension. …"

Participant 23: "Because robot is generally a one to one communication"

Participant 27: "Probably this is relevant is some scenarios only"

The expert opinions on this dimension are varied, it is highlighted that there is not always a team present, often it is a one-on-one task. However, participant 2 mentions that is more common thinking in research and is not always the case in a real-world situation. Participant 27 mentions that this dimension is not always relevant, it might be based on context.

The results are in line with the previous study, suggesting that there is a certain division among experts regarding the five dimensions we investigated. Overall, the experts did not have a consensus on the idea of removing these dimensions, as none of dimensions met the criteria for removal. The experts showed varying opinions in their comments, however, the comments highlighted a common theme across all five problematic dimensions—each measure's importance is context-dependent. Because of this we decided to retain all these dimensions, but to present these as optional to assess the UX with cobots.

Discussion of phase 1

As previously stated, the results of the consensus study led to similar results as in the Borsci et al. (2024) study: there is disagreement between experts with the importance of these dimensions. There is clearly variability in what experts in robotics think is important when it comes to measuring the UX when interacting with a cobot. None of the dimensions met the criteria for removal, so none of them were excluded from the scale. One of the possible reasons for disagreement with these dimension that the experts mentioned was that the importance depended on the context/case of use. Namely, this was mentioned for all five dimensions. We therefore suggest a new approach relating to these five dimensions moving forward, we suggest keeping them in the scale as optional dimensions. Therefore, by keeping these dimensions as optional, the researchers putting the scale to use can estimate the need for these dimensions based on the context of use and type of robot and decide to measure them or not.

These results could have the implication that there is need for nuance when assessing the UX of cobots, that what is important when assessing the UX of cobots might be dependent on the context of use of the cobot. Maybe there is not a need for evaluating the robot's emotional profile or social behavior when assessing a robot that is being used in an industrial context but there might be a need for that assessment when it comes to evaluating a social robot. The same could be said for evaluating the team involved in the task, there is not always a team that is operating the robot so it could be argued that it is not necessary to measure the teams task performance in all situations. This is however something that will become clearer in future research when the scale will be put to the test in a real-word scenario.

Phase 2 – Item generation and card sorting for face validity

Method

Participants

Ethics approval was applied for and accepted by the Ethics committee of the University of Twente (request nr. 231176) prior to participant recruitment. For phase 2, two participants were recruited for a pilot test, then 53 participants were recruited in total for the formal card sorting study. Among them, 16 were male (30%), and 37 were female (70%) as assigned at birth, with the mean age of 24.75 years.. The participants were recruited in various ways, ads were posted in multiple Facebook groups and Reddit communities that are specifically for individuals to advertise surveys in. Many of the groups/communities offered participation in exchange for participation in other member's surveys/projects. Additionally, an account was created on SurveyCircle where it was advertised for 40 days.

Materials

The first step was the item generation. The item generation was a collaboration of the research team, five experts in Human-Factors, three of the members with expertise in HRC. In the primary stage of creating the items we looked for inspiration from other questionnaires, scales, and models that measure UX of HRI in some way such as; The framework for assessing Social Human-Robot Interaction (Bartneck & Forlizzi, 2004), The USUS evaluation framework for HRI (Weiss et al., 2009), The RoSAS (Carpinella et al., 2017), the PYTHEIA scale (Koumpouros et al., 2016), The Model of Human-Robot Trust (Sanders et al., 2011), and The design guidelines for cognitive ergonomics in human robot collaborative systems (Gualtieri et al., 2022).

The research team decided that the items of the scale would mostly use first person language which is commonly used in UX and usability assessment tools (Lewis, 1995;

Brooke, 1996; Phan et al., 2016). The statements primarily start out with I believe, I think, I expected, I found etc. The research team decided that the items should be measured using a Likert scale, which are a commonly used method in psychology research that aims to measure a person's attitude (Schrum et al., 2020).

In the description of each dimension there were a variety of aspects to measure, one item was created per aspect that the dimensions aim to measure (e.g. Dimension 7 - The user's judgment of the robot during a specific performance, considering the efficiency (e.g., time on task), Effectiveness (e.g., task completeness, number of errors), and Utility. This dimension measures three main aspects, efficiency, effectiveness, and utility, resulting in three items being created for this dimension). This resulted in a total of 71 items (see appendix B). The items were created by two members of the research team and then underwent two rounds of review by the three other members of the research team (those with an expertise in HRC).

The original 71 item scale had a varying number of items per dimension, as some dimensions had more aspects to measure than others, e.g., dimension 9 had 10 items compared to dimension 14 with only two items. See Table 3 for the 15 dimensions and the items that belong to each one of the dimensions.

Table 3. The table contains all 71 items that were created for the scale, and which dimension the item belongs to.

D1	D2	D3	D4	D5	D6	D7	D8	D9	D10	D11	D12	D13	D14	D15
1	4	10	15	18	22	27	30	36	46	51	54	59	64	66
2	5	11	16	19	23	28	31	37	47	52	55	60	65	67
3	6	12	17	20	24	29	32	38	48	53	56	61		68
	7	13		21	25		33	39	49		57	62		69
	8	14			26		34	40	50		58	63		70
	9						35	41						71
								42						
								43						
								44						
								45						

The card sorting study was then constructed in Qualtrics, it contained a practise card sorting task, the actual card sorting task and a few demographic questions relating to age, sex as assigned at birth and previous experience with cobots (see appendix D). Finally, participants were invited to comment on the clarity of the dimensions or the items they were shown.

Procedure

A pilot test was conducted to see if the instructions were clear and how much time would be spent on the study (10-15 mins per 10 items), during the pilot test the participants conducted the card sorting study while a member of the research team observed, this was to note down any irregularities or elements needing refinement.

The card sorting study was accessed via link, which led to the Qualtrics page. The participants were shown an opening page, then a consent form, if they said to the consent form the survey ended. Next, the participants received instructions by the interface for a practise sorting task (see Appendix C). Participants were asked to sort (drag and drop) two example items ("Try to put this item under the dimension "Robot's competence features"" and "Try to put this item under the dimension "NONE of the ABOVE"") into one of the 16 groups (the 15 dimensions and a group named "None of the Above"). Once this was completed the participant moved on to the next step, the actual card sorting. The participants were instructed to sort 10 items into the dimensions they thought were the most appropriate. The card sorting survey used a randomized option where 10 of the 71 items were shown to the participants. Next, the participants were asked to answer a demographic questionnaire (see appendix D). Lastly, the participants were thanked for their participation and offered a code to use for SurveyCircle.

Data analysis

Data Analysis was performed in R. The first step was removing all the data from participants that did not finish the card sorting experiment. After the data was cleaned, it was calculated how many times each item was sorted into each dimension in line with expectations.

To establish face validity, we calculated the percentage of participants that attributed an item to a certain (expected or unexpected) dimension, and to retain or remove items that participants were not able to correctly categorised. A criterion was established that items that would be retained for the scale should have an attribution percentage of 50% or higher, as that means a majority of the participants attributed it to a certain dimension (Roth et al., 2011).

To inspect if the remaining items were able to provide a reliable fit with the 15-dimension model, a k-means cluster analysis was conducted, along with an Elbow method analysis to calculate the optimal number of clusters that should be retained to fit our data. Finally, a dendrogram was used to observe the hierarchical organization of the clusters. After the dendrogram analysis another k-means cluster analysis was performed, along with a plot that visualises the clusters.

Results

Face Validity

The results of the card sorting study led to 43 out of the 71 item receiving an agreement of 50% or higher with a certain dimension (See appendix E for the item list), 39 of these items were sorted into their expected dimensions, 4 were not, (e.g. item 69 was expected to be sorted in dimension 15, however it was sorted into dimension 1 by a majority of the participants), see table 5.

Table 4. The table contains the items that were assigned to a certain dimension by 50% or more of the participants. It shows the percentage of participants that assigned an item to certain dimension. The item numbers marked with an asterisk (*) were sorted to an unexpected dimension.

Dimension	Item and attribution percentage					
D1	i1 83%	i2 60%	i3 57%	i69* 50%		
D2	i4 100%	i5 80%	i6 67%	i7 63%	i13* 75%		
D3	i10 100%	i11 50%	i14 57%				
D4	i17 86%						
D5	i18 88%	i20 67%					
D6	i22 88%	i23 75%	i24 100%	i25 50%			
D7	i8* 63%	i28 63%					
D8	i30 57%	i31 57%	i32 57%	i33 50%	i34 60%	i35 86%	i48* 60%
D9	i15* 50%	i29* 67%	i50 50%	i45 75%			
D10	i49 86%	i50 86%					
D11	i46* 50%	i51 71%	i52 57%				
D12	i55 64%						
D13	i59 100%	i60 83%	i61 88%				
D14	i64 63%	i65 50%					
D15	x	x					

In order to design a scale with a manageable number of items the top 3 items were retained per dimension (with the exception of dimension 14 that consists of only two items). To make sure there are enough items to work with for the next steps of the research we make an exception to the 50% or higher level of attribution criterium when only one or two items survived in a dimension, in such cases the item with the highest percentage of attribution of those remaining will be retained for the scale.

Dimension 1 – Easiness of Robot Regulation: Dimension 1 originally consisted of three items (1-3). However, four items seem to be attributed to this dimension by the participants (≥50%):

- **Item 1.** It was easy to physically regulate the robot. (83%) (D1)
- **Item 2**. I found it/ it appears to be easy to position the robot's components correctly. (60%) (D1)
- **Item 3.** I believe that from a physical point of view it appears to be easy to manipulate and put the robot into position. (57%) (D1)
- **Item 69.** I found the robot's interface and interaction modality easy to learn. (50%) (D15)

Items 1-3 all received a moderate or high percentage of attribution and will therefore all be retained for the scale as they represent this dimension well. Item 69 has the word easy in it, the same as the name of the dimension, "easiness of robot regulation", which could possibly be the reasoning behind it being considered belonging to this dimension. Item 69 will not be retained for the scale as items 1-3 all had higher attribution percentages.

Dimension 2 – Robot Physical Appearance: Dimension 2 originally consisted of five items (4-9). Five items were attributed to this dimension by the participants (≥50%), four expected items, and one unexpected, created for dimension 3:

- **Item 4.** I had a positive impression of the robot's physical appearance. (100%) (D2)
- **Item 5.** I had a positive impression of the robot's dimensions i.e., high, width, length, weight. (80%) (D2)

- **Item 6.** I had a positive impression of the robot’s features e.g., form, material. (67%) (D2)
- **Item 7.** The level of anthropomorphism (machinelike or humanlike) of the robot was appropriate for the intended purpose. (63%) (D2)
- **Item 13.** I did not consider the robot’s appearance disturbing. (75%) (D3)

Items 4, 5 and 6 will be retained for the scale as they were attributed to dimension 2 the most by the participants and therefore appear to represent the dimension the best. Even though item 13 was assigned to dimension 2 by a higher percentage of participants compared to item 6 it will not be retained for the scale as it was originally created to measure dimension 3, “emotional appearance of the robot”. As the item includes the word appearance, that might explain most of the participants attributing it to the dimension called “robot physical appearance”. However, it does not properly measure the any of the aspects the dimension measures and will therefore not be retained for the scale.

Dimension 3 - Robot’s emotional appearance: Dimension 3 originally consisted of five items (10-14). Three items were attributed to this dimension by the participants (≥50%), all expected:

- **Item 10.** I believe that the robot’s emotional appearance was likable/attractive. (100%) (D3)
- **Item 11**. I believe that the robot's design does not cause emotional discomfort. (50%) (D3)
- **Item 14.** I felt that the robot displayed a sense of warmth during our collaboration e.g., it was social, friendly. (57%) (D3)

All three items will be retained for the scale as they were attributed by 50% or more of the participants. Item 10 represents the dimension extremely well as it was attributed by all of the participants. Items 11(50%) and 14 (57%) represent the dimension moderately well.

Dimension 4 - Robot's competence features*:* Dimension 4 originally consisted of three items (15-17). One item was attributed (≥50%), to this dimension, therefore, an exception was made to guarantee three items for this dimension:

- **Item 16.** I believe the robot is reliable and trustworthy in terms of competencies. (44%) (D4)
- **Item 17.** I found the robot to be responsive and transparent in terms of competencies. (86%) (D4)
- **Item 39.** After using the robot, I found myself accepting of its role in the collaboration. (43%) (D9)

The only item to meet the criteria, (≥50%), was item 17. To ensure there were at least three items from dimension 4 in the scale, the two items with the highest attribution percentage were chosen for the scale, item 16 (44%) and item 39 (43%) both items represent the dimension moderately. Item 39, originally belonged to dimension 9, however with a slight rephrasing it was adapted to measure dimension 4 better, "After using the robot, I found myself accepting of its competence during the collaboration."

Dimension 5 - Robot physical behavior: Dimension 5 originally consisted of four items (18-21). Two items were attributed to the dimension (≥50%):

- **Item 18.** I perceived the robot's movements to be smooth and flexible. (88%) (D5)
- **Item 20.** I believe that the robot's physical behavior (e.g., noise, movement, autonomy, interactivity) during the interaction was suitable. (67%) (D5)

Both items will be retained for the scale since they appear to represent the dimension well. To ensure three items were retained for the scale, the item with the highest attribution percentage under 50% was retained:

- **Item 21.** I believe that the robot's movements and behavior seemed lifelike and natural. (44%) (D5)

Item 21 will be retained for the scale, representing the dimension moderately.

Dimension 6 - Robot's social behavior: Dimension 6 originally consisted of five items (22-26). Four items were attributed to the dimension (≥50%):

- **Item 22.** I believe that the overall social behavior of the robot was appropriate. (88%) (D6)
- **Item 23.** I believe that the robot acted and communicated according to social norms. (75%) (D6)

  Item 24. I felt that the robot engaged in meaningful social interactions during our collaboration. (100%) (D6)
- **Item 25.** I think the robot gave me a sense of companionship during our collaboration. (50%) (D6)

The three items with the highest percentage of attribution to the dimension by participants will be retained for the scale, item 22, 23 and 24 as they represent the dimension the best.

Dimension 7 - Robot task performance: Dimension 7 originally consisted of three items (27-29). Two items were attributed to dimension 7 (≥50%):

- **Item 8.** The robot's perceived robustness (e.g., its ability to withstand physical stress, challenges etc.) met the specific requirements for the task and context of usage. (63%) (D2)
- **Item 28.** I believe that the robot was useful by enabling a correct (without error) performance. (63%) (D7)

Item 8 originally belonged to dimension 2, "robot physical appearance", it does however relate to the utility (one of the factors that the dimension aims to measure) and relates to task performance, therefore it will be retained for the scale under dimension 7 since it received a moderately high percentage of attribution from the participants (63%) and seems to represent the dimension well. It will however be rephrased to be more appropriate for measuring dimension 7, “The robot's perceived robustness (e.g., its ability to withstand physical stress, challenges etc.) met the specific requirements for the task”. Furthermore, Item 28 received a moderately high percentage of attribution from the participants (63%) and therefore seems to appropriately represent the dimension and will be retained for the scale.

To have three items retained for the scale, an exception was made to the original criteria, and the last item retained was item 40:

- **Item 40.** I was highly satisfied with the robot’s performance. (43%) (D9)

The item moderately represents this dimension and will be retained for the scale, the item will be slightly rephrased provide a better fit to this dimension,” I was highly satisfied with the robot’s task performance”.

Dimension 8 - Human judgment before the interaction with a cobot: Dimension 8 originally consisted of six items (30-35). All six of these items were attributed to this dimension (≥50%), as well as one unexpected item, originally created for dimension 10:

- **Item 30.** I expected the collaboration with the robot to be safe before using it. (57%) (D8)
- **Item 31.** I expected the robot to be reliable and quick prior to collaborate with it. (57%) (D8)

  Item 32. I did not experience any anxiety related to the robot prior to the collaboration with it. (57%) (D8)
- **Item 33.** My overall attitude towards using the robot was positive prior to our collaboration. (50%) (D8)
- **Item 34.** I accepted the idea of using the robot for the task prior to use. (60%) (D8)

- **Item 35.** Before interacting with the robot, I had the intention to use it for similar tasks or interactions in the future. (86%) (D8)
- **Items 48.** I do not see any ethical, personal, or social issues in collaborating with the robot for my job. (60%) (D10)

Item 35 received a high percentage of attribution (83%) to dimension 8, and therefore represents the dimension well and will be retained for the scale. Additionally, item 34 will be retained for the scale as it seems to represent this dimension well. Lastly, item 48 will be retained for the scale, although it was originally created for dimension 10, "Human-Factors personality based". However, the item reflects *attitude towards use* (one of the aspects the dimension aims to measure) and will therefore be retained for the scale under dimension 8. Items 30, 31, 32, and 33 will not be retained for the scale as they were attributed by a lower percentage of participants than items 34, 35 and 48, and do not represent the dimension as well.

Dimension 9 - Human judgement of the performance with a cobot: Dimension 9 originally contained 10 items (36-45). Four items were attributed to dimension 9 (≥50%), two expected, while the other two items originally belonged to other dimensions:

- **Item 15.** I perceived the robot as competent and smart in terms of behavior. (50%) (D4)
- **Item 29.** I believe that the collaboration with the robot was useful by enabling tasks to be completed in an efficient and effective manner. (67%) (D7)
- **Item 42**. I experienced no frustration while working with the robot. (50%) (D9)
- **Item 45.** I would like to use the robot again based on my experience during the task. (75%) (D9)

Item 45 received the highest percentage of attribution from the participants (75%) of these four items, therefore it represents the dimension well and will be retained for the scale. Even though it was not originally created for this dimension, item 29 will be retained for the scale as it was attributed to the dimension by 67% of the participants and it measures *usability* (one of the aspects this dimension aims to measure), thus it represents dimension 9 well enough to be retained for the scale. The third item that will be retained for the scale is item 42 (50%). It was attributed to dimension 9 an equal percentage to item 15, however item 42 was originally created for dimension 9 while item 15 was not, therefore item 42 represents this dimension better.

Dimension 10 - Human-Factors personality-based: Dimension 10 originally consisted of five items (46-50). Two of them were attributed to dimension 10 (≥50%):

- **Item 49.** I have a trusting personality. (86%) (D10)
- **Item 50.** I believe my personality traits have a significant influence on my collaboration with robots in general. (86%) (D10)

Both item 49 and 50 received a high percentage of attribution to dimension 10 by the participants and will be retained for the scale as they both represent dimension 10 well. To have three items for each dimension the criteria had to be broken, retaining item 43:

- **Item 43.** I felt calm (e.g., no stress) during the interaction with the robot. (43%) (D9)

The item originally belonged to dimension 9, but it does describe calmness and a lack of stress which can be tied to personality factors.

Dimension 11 - Human-Factors ability-based: Dimension 11 originally consisted of three items (51-53). Three items were attributed to dimension 11 (≥50%), two expected, and one originally made for dimension 10:

- **Item 46.** I feel confident in my ability to use the robot to achieve key tasks. (50%) (D10)

- **Item 51.** My level of expertise contributed to a successful interaction with the robot. (71%) (D11)
- **Item 52.** I believe that my general understanding of robotics contributed to a positive collaboration with the robot. (57%) (D11)

Even though item 46 was created for dimension 10, it does measure dimension 11 well, as it pertains to the user's *ability*, which this dimension measures. All three items (46, 51 and 52) will be retained for the scale, as they seem to represent the dimension well.

Dimension 12 - Task performed: Dimension 12 originally consisted of five items (54-58). One expected item was attributed to this dimension (≥50%):

- **Item 55.** The task I performed with the robot did not require too much physical effort. (64%) (D12)

Item 55 represents the dimension well and will therefore be retained for the scale. An exception will be made for this dimension, as only one item met the set criteria, the two items with the highest attribution percentage will be retained:

- **Item 57:** I believe the robot was useful in accomplishing the task. (40%) (D12)
- **Item 58:** I felt that it was important to use the robot to perform this critical task. (33%) (D12)

Item 57 and 58 will be retained for the scale, as they seem to moderately represent this dimension.

Dimension 13 - The environment of interaction: Dimension 13 originally consisted of five items (59-63). Three of these items were attributed to this dimension (≥50%):

- **Item 59.** The environmental conditions (e.g., lighting, noise, dust) were disturbing the task. (100%) (D13)
- **Item 60.** The workstation layout facilitated a positive interaction with the robot. (83%) (D13)

- **Item 61:** The interaction took place in a workstation with a layout that facilitated the completion of the task. (88%) (D13)

Items 59, 60, 61 were all attributed to dimension 13 by a large majority of the participants, therefore all three items represent this dimension well and will be retained for the scale.

Dimension 14 - Team involved during the task performance: Dimension 14 originally consisted of two items (64-65), both items were attributed to this dimension (≥50%):

- **Item 64.** I believe that it is possible for multiple operators (a team) to collaborate proficiently to use the robot to achieve the task. (63%) (D14)
- **Item 65.** When multiple operators (a team) have to collaborate interacting with the robot all the operators can understand their roles. (50%) (D14)

As items 64 and 65 seem to represent this dimension well and will therefore both be retained for the scale.

Dimension 15 - Interaction aspects: Dimension 15 originally consisted of six items (66-71). None of these items received an agreement higher than 50%. The three items that had the highest percentage of attribution will be retained for the scale, one expected, and two originally belonging to dimension 15:

- **Item 41.** I found the robot's interface or interaction methods (e.g., touch panel, voice commands, haptic feedback) highly usable. (43%) (D9)
- **Item 66.** I was pleased with the overall interaction with the robot. (38%) (D15)
- **Item 71.** I found the type of interface used for interaction (e.g., physical-based, graphical-based, vocal-based, gesture-based) was appropriate and effective. (43%) (D15)

Items 41, 66 and 71 represent the dimensions moderately, they will be retained for the scale, thus dimension 15 can be kept for further analysis.

To attain a manageable number of items we decided to retain the three items with the best score per dimension, irrespectively if the agreement was below or above 50% on the item (apart from dimension 14, where only two items were available). Of the 71 items originally created, 44 of the most representative items were retained for the scale for further analysis. Table 5 shows an overview of each item that was retained, shown under the dimension it measures (for the scale in its entirety see appendix F)

Table 5. The table shows the 15 dimensions and all the items that were retained, categorised by dimension. The items in bold are items not previously considered for the scale but where an exception to the criterion.

D1	D2	D3	D4	D5	D6	D7	D8	D9	D10	D11	D12	D13	D14*	D15
i1	i4	i10	i16	i18	i22	i8	i34	i29	**i43**	i46	i55	i59	i64	**i41**
i2	i5	i11	**i17**	i20	i23	i28	i35	i42	i49	i51	**i57**	i60	i65	**i66**
i3	i6	i14	**i39**	**i21**	i24	**i40**	i48	i45	i50	i52	**i58**	i61		**i71**

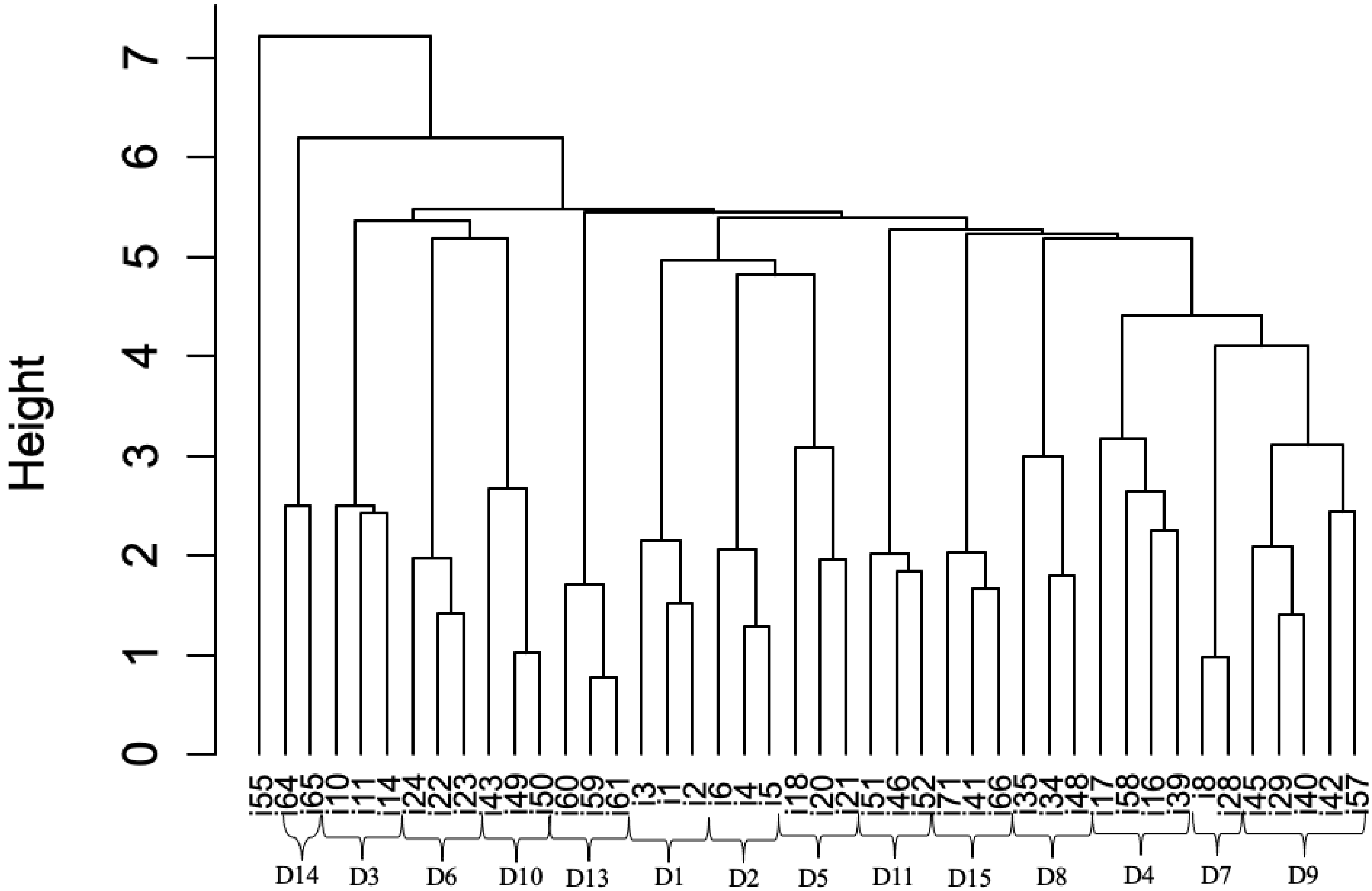
The table shows all the items retained per dimension. It is important to note that the items that are in bold, were originally excluded from further analysis, but were re-introduced to the analysis to have three items per dimension. Dimension 14 is marked with an asterisk since it was an exception to the three items per dimension rule, as it originally had two items.

Cluster analysis

A k-means cluster analysis was conducted on the 44 items that were retained for the scale from the card sorting analysis. According to the Elbow method analysis, the optimal number of clusters was 10. However according to the k-means cluster analysis 10 clusters only explained 62% of the variability of the items, while 15 clusters explained 88% of the variability of the items.

Additionally, a hierarchical cluster analysis was performed in the form of a dendrogram, see Figure 2.

Figure 2. The figure shows a dendrogram with the items from the scale. The dimensions that created clusters have been labelled with the number of the dimension.



The dendrogram shows a clustering pattern for 14 out of 15 dimensions. Dimension 12, task performed, was the exception, not showing any clustering pattern, however three other dimensions had slight deviations to the expected organizational structures, specifically:

- Dimension 4, robot’s competence features' items (i16, i17, i39) were clustered together with item 58 (I felt that it was important to use the robot to perform this critical task) from dimension 12, this item was one of the least representative items of the scale. However, this does not affect the general structure of this dimension, as the three items of dimension 4 are all clustered together.

- Dimension 7, robot task performance's items (i8, i28, i40) were not all clustered together, item 8 and item 28 were clustered together, item 40 was clustered with the items of dimension 9. This indicates that item 40 is not considered similar to the other two items of the dimension, as a result, it will be removed.
- Dimension 9, human judgement of the performance with a cobot, (i29, i42, i45) had all its items clustered together, along with items 40 (mentioned above), and item 57, belonging to dimension 12. As all three items belonging to dimension 9 were clustered together, we can assume they measure the dimension well.
- Dimension 12, task performed (i55, i57, i58), none of the items of this dimension were clustered together, indicating that the items are not considered similar and do not measure the dimension well. Consequently, this dimension will be removed from the scale.

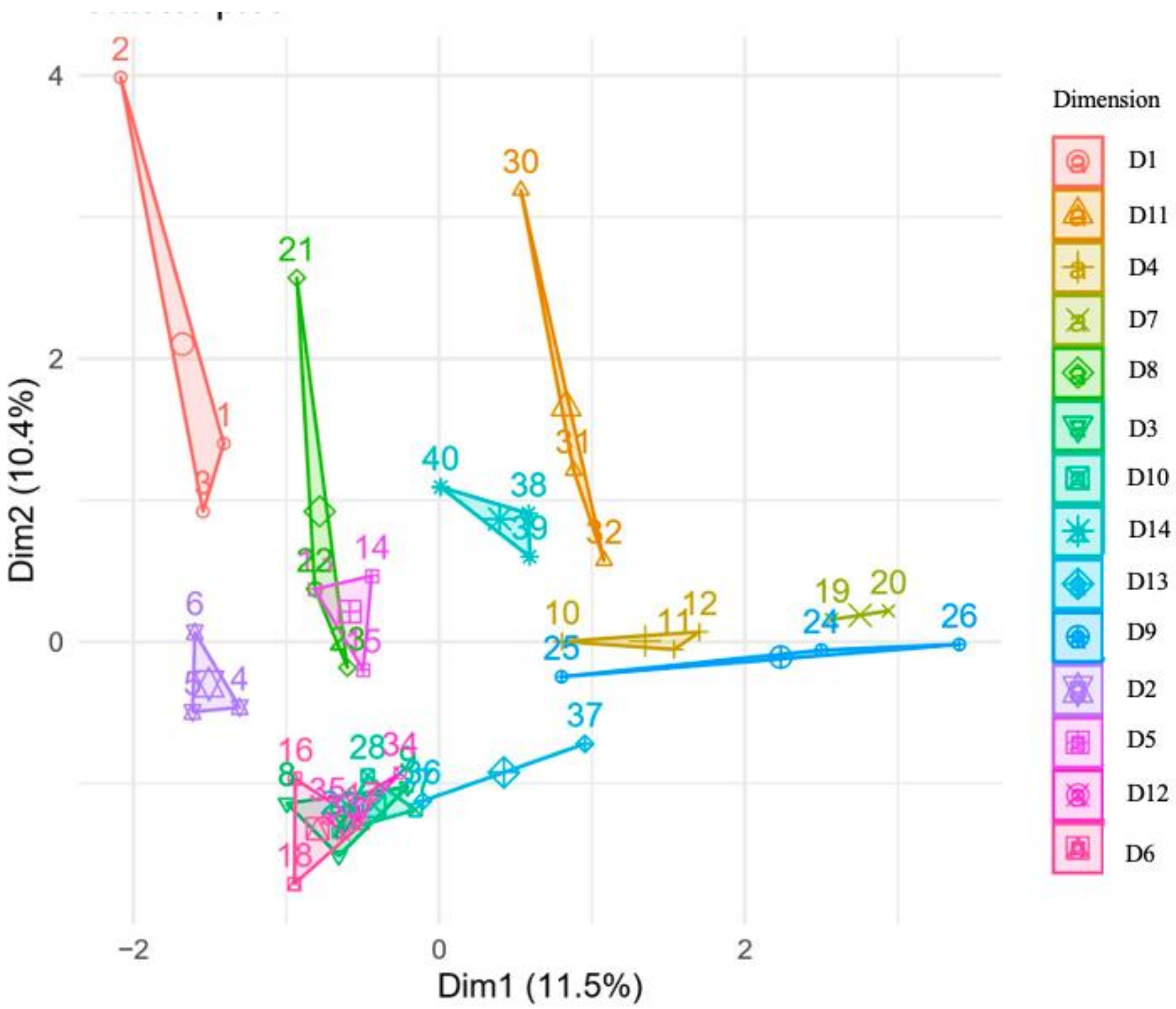
After the removal of dimension 12 and item 40 the dimensions and the items were given new numbers, see table 6.

Table 6. The dimension and their respective items with an updated numbering system.

D1	**D2**	**D3**	**D4**	**D5**	**D6**	**D7**	**D8**	**D9**	**D10**	**D11**	**D12**	**D13**	**D14**
i1	i4	i7	i10	i13	i16	i19	i21	i24	i27	i30	i33	i36	i38
i2	i5	i8	i11	i14	i17	i20	i22	i25	i28	i31	i34	i37	i39
i3	i6	i9	i12	i15	i18		i23	i26	i29	i32	i35		i40

To validate the 14-dimension model a k-means cluster analysis was performed a second time, this time using a visualisation, shown in figure 3.

Figure 3. A visualisation of the k-means cluster analysis using the 14-dimension model.



The visualisation shows support for the 14-dimension model as it is in line with the expected organizational structure of the dimensions. The k-means cluster analysis explains 85% of the variability in the data, these findings validate this model, which indicates that the items of the 14 dimensions measure the dimensions well.

Discussion of Phase 2

The research goal for phase 2 was to generate a list of items and subsequently conducting a card sorting study to establish face validity for the items created. The results of the card sorting study led to a large reduction in the item pool, which was expected. Most of the items that were retained seem to represent their respective dimensions well.

The cluster analysis led to the conclusion that the items of each dimension seem to be similar to each other, indicating that the items appropriately measure the dimensions, with the exception of dimension 12. Dimension 12 was not adequately measured by its items, which led to it being removed. Additionally, the cluster analysis led to the removal of item 40, as it was not clustered with the other items from dimension 7.

The research resulted in a 40-item scale, measuring 14 dimensions, see table 7.

Table 7. The scale in its final version, the dimensions have been given new numbers after the removal of dimension 12 and the items have been numbered from 1-40.

Dimension name	Items
1. Easiness of robot regulation.	**Item 1.** It was easy to physically regulate the robot. **Item 2**. I found it/ it appears to be easy to position the robot's components correctly. **Item 3.** I believe that from a physical point of view it appears to be easy to manipulate and put the robot into position.
2. Robot physical appearance.	**Item 4.** I had a positive impression of the robot's physical appearance. **Item 5.** I had a positive impression of the robot's dimensions i.e., high, width, length, weight. **Item 6.** I had a positive impression of the robot's features e.g., form, material.
3. Robot's emotional appearance.	**Item 7.** I believe that the robot's emotional appearance was likable/attractive. **Item 8**. I believe that the robot's design does not cause emotional discomfort. **Item 9.** I felt that the robot displayed a sense of warmth during our collaboration e.g., it was social, friendly.
4. Robot's competence features.	**Item 10.** I believe the robot is reliable and trustworthy in terms of competencies. **Item 11.** I found the robot to be responsive and transparent in terms of competencies. **Item 12.** After using the robot, I found myself accepting of its competence during the collaboration.
5. Robot's physical behavior.	**Item 13.** I perceived the robot's movements to be smooth and flexible. **Item 14.** I believe that the robot's physical behavior (e.g., noise, movement, autonomy, interactivity) during the interaction was suitable. **Item 15.** I believe that the robot's movements and behavior seemed lifelike and natural.

6. Robot's social behavior.	**Item 16.** I believe that the overall social behavior of the robot was appropriate. **Item 17.** I believe that the robot acted and communicated according to social norms. **Item 18.** I felt that the robot engaged in meaningful social interactions during our collaboration.
7. Robot task performance.	**Item 19.** The robot's perceived robustness (e.g., its ability to withstand physical stress, challenges etc.) met the specific requirements for the task. **Item 20.** I believe that the robot was useful by enabling a correct (without error) performance.
8. Human judgment before the interaction with a cobot.	**Item 21.** I accepted the idea of using the robot for the task prior to use. **Item 22.** Before interacting with the robot, I had the intention to use it for similar tasks or interactions in the future. **Items 23.** I do not see any ethical, personal, or social issues in collaborating with the robot for my job.
9. Human judgment of the performance with a cobot.	**Item 24.** I believe that the collaboration with the robot was useful by enabling tasks to be completed in an efficient and effective manner. **Item 25**. I experienced no frustration while working with the robot. **Item 26.** I would like to use the robot again based on my experience during the task.
10. Human-Factors personality-based.	**Item 27.** I felt calm (e.g., no stress) during the interaction with the robot. **Item 28.** I have a trusting personality. **Item 29.** I believe my personality traits have a significant influence on my collaboration with robots in general.
11. Human-Factors ability-based.	**Item 30.** I feel confident in my ability to use the robot to achieve key tasks. **Item 31.** My level of expertise contributed to a successful interaction with the robot. **Item 32.** I believe that my general understanding of robotics contributed to a positive collaboration with the robot.
12. The environment of interaction.	**Item 33.** The environmental conditions (e.g., lighting, noise, dust) were disturbing the task. **Item 34.** The workstation layout facilitated a positive interaction with the robot. **Item 35.** The interaction took place in a workstation with a layout that facilitated the completion of the task.
13. Team involved during the task performance.	**Item 36.** I believe that it is possible for multiple operators (a team) to collaborate proficiently to use the robot to achieve the task. **Item 37.** When multiple operators (a team) have to collaborate interacting with the robot all the operators can understand their roles.

14. Interaction aspects.	**Item 38.** I found the robot's interface or interaction methods (e.g., touch panel, voice commands, haptic feedback) highly usable. **Item 39.** I was pleased with the overall interaction with the robot. **Item 40.** I found the type of interface used for interaction (e.g., physical-based, graphical-based, vocal-based, gesture-based) was appropriate and effective.

The careful selection of items, coupled with the clustering analysis, underscores the face validity of the retained items in inspecting various dimensions of UX assessment of HRC. These results suggest the scale's ability to evaluate different aspects of HRC effectively, laying a solid groundwork for future research and real-world use.

General Discussion

This research provides a first edition of a scale that assesses the UX of interaction/collaboration with cobots. It includes 14 dimensions (five optional) measured with 40 items that assess different aspects of user experience related to cobots. This scale is the first of its kind, a comprehensive UX scale that assesses cobots. The current scale is solid starting point for future phases of the research, where the scale will be put to the test in a real-world scenario.

In this research, firstly, a consensus study was conducted with five dimensions deemed problematic in previous research (Borsci et al., 2024), and whether they should be removed from the scale. The results showed a disagreement between experts on the importance of these dimensions. As a result, the five dimensions are offered as optional dimensions for the scale, as many experts suggested that their importance depended on context. This indicates there is need for nuance when assessing the UX of cobots, we need to look at the context of use in our assessment and varying it, for instance, what type of cobot is being used, or what situation is it being used in. Secondly, this research conducted a card sorting study with items generated by the research team for each of the 15 dimensions. This resulted in the exclusion of numerous items, yielding a selection of the most representative items. Through additional analysis, a dimension was eliminated due to its items ineffectively assessing the dimension. This left a strong representative scale, which is an important first step towards a very valuable measurement tool.

A gap seems to persist in literature and research, there is lack of appropriate assessment tools and evaluation methods, not only in the domain of HRC (Prati et al., 2022), but also in the domain of Artificial Intelligence (Schraagen, 2024) and in the domain of Human-Robot Interaction (Marvel et al., 2020). This scale is an important addition to the field of HRI/HRC as it is the starting point towards bridging this research gap and providing a validated assessment tool. Therefore, by creating this scale, we are providing a tool that will help the scientific community identify use problems with robots. By identifying problems with any system, we have the option to maximize the usability of the system. This is an important problem to solve in the field of HRC, as many cobots receive low usability scores (Frijns & Schmidbauer, 2021). With the increased identification of common usability problems, there's an opportunity to strengthen design guidelines for cobot by actively identifying and addressing these issues, thereby integrating user experience considerations into their design and development. Current research focuses highly on the technological aspects of design guidelines in HRC, often neglecting the human aspects (Prati et al., 2021) which is a problem that needs to be solved in future HRC research.

A few limitations were identified in this study. The initial phase had fewer participants than expected: approximately a quarter compared to the previous phase. This difference in participant numbers may somewhat limit the broad applicability of our findings. In the subsequent phase, uneven exposure of items during the card sorting task might have influenced our results. Furthermore, the limited visibility of dimension descriptions required participants to hover over dimension names for details, possibly affecting their attention and how they sorted items. Although having each participant sort all items would have been optimal, the extensive item list risked causing participant fatigue, leading us to opt for a more concise approach, where each participant sorted only a part of the items.

In future research phases, implementing the scale in real-world scenarios with cobots and evaluating their performance remains a priority. Additionally, conducting exploratory and confirmatory factor analysis would further enhance the scale's validation. Another potential step could be revisiting card sorting studies to refine the scale even more. As the scale now has fewer items, a card sorting study could be re-performed allowing each participant to sort all the items of the scale, not only part of it, as was done in this research.

In conclusion, this scale represents an important initial step in evaluating user experience with cobots. While acknowledging the study's limitations, this work sets the stage for further research in assessing collaboration with robots. By highlighting usability concerns and emphasizing a human-centered approach in Human-Robot Collaboration, this scale contributes to improving cobot design. This shift toward a more user-oriented focus holds promise for enhancing Human-Robot Interaction and Collaboration, and usability in the field.

References

Ahmed, S., & Hossain, M. (2019). The impact of robotics in the growth and economic development. *The Business and Management Review*, *10*(5). https://cberuk.com/cdn/conference_proceedings/2020-01-05-10-10-55-AM.pdf

Akalin, N., Kristoffersson, A., & Loutfi, A. (2022). Do you feel safe with your robot? Factors influencing perceived safety in human-robot interaction based on subjective and objective measures. *International Journal of Human-Computer Studies*, *158*, 102744. https://doi.org/10.1016/j.ijhcs.2021.102744

Allam, A., Razak, A., Hussin, C., & Dahlan, H. (2013). User Experience: Challenges and Opportunities. In *Journal of Information Systems Research and Innovation*. https://seminar.utmspace.edu.my/jisri/download/F1_FinalPublished/Pub4_UserExperienceChallenges.pdf

Bartneck, C., & Forlizzi, J. (2004). A design-centred framework for social human-robot interaction. *RO-MAN 2004. 13th IEEE International Workshop on Robot and Human Interactive Communication (IEEE Catalog No.04TH8759)*. https://doi.org/10.1109/roman.2004.1374827

Bartneck, C., Kulić, D., Croft, E., & Zoghbi, S. (2008). Measurement Instruments for the Anthropomorphism, Animacy, Likeability, Perceived Intelligence, and Perceived Safety of Robots. *International Journal of Social Robotics*, *1*(1), 71–81. https://doi.org/10.1007/s12369-008-0001-3

Beerlage-de Jong, N., Kip, H., & Kelders, S. M. (2020). Evaluation of the Perceived Persuasiveness Questionnaire: User-Centered Card-Sort Study. *Journal of Medical Internet Research*, *22*(10), e20404. https://doi.org/10.2196/20404

Berg, J., & Lu, S. (2020). Review of Interfaces for Industrial Human-Robot Interaction. *Current Robotics Reports*, *1*(2), 27–34. https://doi.org/10.1007/s43154-020-00005-6

Bevan, N., Kirakowski, J., & Maissel, J. (1991). *What is Usability?* https://usabilitynet.org/papers/whatis92.pdf

Borsci, S., Prati E., Prati, E., Landwehr, J., & Peruzzini, M., (unpublished) Quantifying the subjective experience in Human-Robot Collaboration: towards a validated construct.

Buchner, R., Wurhofer, D., Weiss, A., & Manfred Tscheligi. (2012). *User experience of industrial robots over time*. https://doi.org/10.1145/2157689.2157712

Bussolon, S., Russi, B., & Missier, F. D. (2006). Online card sorting. *Proceedings of the 13th Eurpoean Conference on Cognitive Ergonomics Trust and Control in Complex Socio-Technical Systems - ECCE '06*. https://doi.org/10.1145/1274892.1274912

Carpinella, C. M., Wyman, A. B., Perez, M. A., & Stroessner, S. J. (2017, March 1). *The Robotic Social Attributes Scale (RoSAS): Development and Validation*. IEEE Xplore. https://ieeexplore.ieee.org/document/8534914

Chacón, A., Ponsa, P., & Angulo, C. (2021). Usability Study through a Human-Robot Collaborative Workspace Experience. *Designs*, *5*(2), 35. https://doi.org/10.3390/designs5020035

Cohen, J., Margulieux, L. E., Renken, M., Smith, S. F., & Jones, W. M. (2018). Measuring Maker Mindset: Establishing Content Validity with Card Sorting. *Repository.isls.org*, *3*. https://repository.isls.org/handle/1/703

Dong, X., Wang, T., & Benbasat, I. (2016). Measuring IT affordance in OSC IT Affordances in Online Social Commerce: Conceptualization Validation and Scale Development Full papers. *In AMCIS*, *1*. https://core.ac.uk/download/pdf/301368775.pdf

Duarte, I. M., Pinto, A., Carvalho, C., Zornoza, A., & Santos, J. (2022). The Contribution of the User Experiences Goals for Designing Better Cobots: A Systematic Literature Review. *Applied System Innovation*, *5*(6), 119. https://doi.org/10.3390/asi5060119

El Zaatari, S., Marei, M., Li, W. & Usman, Z., (2019) Cobot programming for collaborative industrial tasks: *An overview,Robotics and Autonomous Systems,* vol. 116, pp. 162-180.

Frijns, H. A., & Schmidbauer, C. (2021). Design Guidelines for Collaborative Industrial Robot User Interfaces. In C. Ardito, R. Lanzilotti, A. Malizia, H. Petrie, A. Piccinno, G. Sesolda, & K. Inkpen (Eds.), *Human-Computer Interaction – INTERACT 2021, 18th IFIP TC 13 International Conference Bari, Italy, August 30 – September 3, 2021 Proceedings,* Part III. Springer.

George, D. A. S., & George, A. S. H. (2023). The Cobot Chronicles: Evaluating the Emergence, Evolution, and Impact of Collaborative Robots in Next-Generation Manufacturing. *Partners Universal International Research Journal*, *2*(2), 89–116. https://doi.org/10.5281/zenodo.8021406

Gualtieri, L., Fraboni, F., De Marchi, M., & Rauch, E. (2022). Development and evaluation of design guidelines for cognitive ergonomics in human-robot collaborative assembly systems. *Applied Ergonomics*, *104*, 103807. https://doi.org/10.1016/j.apergo.2022.103807

Hartson, R., & Pyla, P. S. (2012). The UX Book: Process and Guidelines for Ensuring a Quality User Experience. In *Google Books*. Elsevier. https://books.google.nl/books?hl=en&lr=&id=w4I3Y64SWLoC&oi=fnd&pg=PP1&dq=Hartson

Hassenzahl, M. (2014, January 1). *User Experience and Experience Design*. Interaction Design Foundation - IxDF. https://www.interaction-design.org/literature/book/the-encyclopedia-of-human-computer-interaction-2nd-ed/user-experience-and-experience-design

Hinderks, A., Winter, D., Schrepp, M., & Thomaschewski, J. (2019). Applicability of User Experience and Usability Questionnaires. *Idus.us.es*. https://doi.org/10.3217/jucs-025-13-1717

Hsu, C.-C., & Sandford, B. (2007). The Delphi Technique: Making Sense of Consensus. *Practical Assessment, Research, and Evaluation*, *12*(1), 10. https://doi.org/10.7275/pdz9-th90

ISO. (2010). Ergonomics of human-system interaction -- Part 210: Human-centred design for interactive systems. (ISO 9241-210). International Organization for Standardization.

Jan Maarten Schraagen. (2023). Responsible use of AI in military systems: Prospects and challenges. *Ergonomics*, 66(11), 1719–1729. https://doi.org/10.1080/00140139.2023.2278394

Jordan, P. W. (1998). *An introduction to usability*. Taylor & Francis. https://www.taylorfrancis.com/books/mono/10.1201/9781003062769/introduction-usability-patrick-jordan

Kolbeinsson, A., Lagerstedt, E., & Lindblom, J. (2019). Foundation for a classification of collaboration levels for human-robot cooperation in manufacturing. *Production & Manufacturing Research*, *7*(1), 448–471. https://doi.org/10.1080/21693277.2019.1645628

Koumpouros, Y., Papageorgiou, E., Karavasili, A., & Koureta, F. (2016). PYTHEIA: A Scale for Assessing Rehabilitation and Assistive Robotics. *International Journal of Mechanical and Materials Engineering*, *10*(11), 522–526. https://publications.waset.org/10005588/pytheia-a-scale-for-assessing-rehabilitation-and-assistive-robotics

Lantz, E., Keeley, J. W., Roberts, M. C., Medina-Mora, M. E., Sharan, P., & Reed, G. M. (2019). Card Sorting Data Collection Methodology: How Many Participants Is Most

Efficient? *Journal of Classification*, *36*(3), 649–658. https://doi.org/10.1007/s00357-018-9292-8

Lindblom, J., & Andreasson, R. (2016). Current Challenges for UX Evaluation of Human-Robot Interaction. *Advances in Ergonomics of Manufacturing: Managing the Enterprise of the Future*, 267–277. https://doi.org/10.1007/978-3-319-41697-7_24

Loo, R. (2002), "The Delphi method: a powerful tool for strategic management", *Policing: An International Journal*, Vol. 25 No. 4, pp. 762-769. https://doi.org/10.1108/13639510210450677

Marvel, J. A., Bagchi, S., Zimmerman, M., & Antonishek, B. (2020). Towards Effective Interface Designs for Collaborative HRI in Manufacturing. *ACM Transactions on Human-Robot Interaction*, 9(4), 1–55. https://doi.org/10.1145/3385009

Matheson, E., Minto, R., Zampieri, E. G. G., Faccio, M., & Rosati, G. (2019). Human–Robot Collaboration in Manufacturing Applications: A Review. *Robotics*, *8*(4), 100. https://doi.org/10.3390/robotics8040100

Müller, R., Vette, M., & Geenen, A. (2017). Skill-based Dynamic Task Allocation in Human-Robot-Cooperation with the Example of Welding Application. *Procedia Manufacturing*, *11*, 13–21. https://doi.org/10.1016/j.promfg.2017.07.113

Murphy, R. R., Nomura, T., Billard, A., & Burke, J.L., (2010) "Human–Robot Interaction," in *IEEE Robotics & Automation Magazine, 17(2)*, pp. 85-89, https://doi.org/10.1109/MRA.2010.936953.

Pervez, A., & Ryu, J. (2008). Safe physical human robot interaction-past, present and future. *Journal of Mechanical Science and Technology*, *22*(3), 469–483. https://doi.org/10.1007/s12206-007-1109-3

Prati, E., Pellicciari, M., Borsci, S., & Peruzzini, M., (2022) A Systematic Literature Review of User Experience Evaluation Scales for Human-Robot Collaboration. In: Moser B, editor. Transdisciplinarity and the Future of Engineering: Proceedings of the 29th International Society of Transdisciplinary Engineering (ISTE) Global Conference, July 5–July 8, 2022, Cambridge, MA, USA. 28: IOS Press. p. 13.

Prati, E., Peruzzini, M., Pellicciari, M., & Raffaeli, R. (2021). How to include User eXperience in the design of Human-Robot Interaction. *Robotics and Computer-Integrated Manufacturing*, *68*, 102072. https://doi.org/10.1016/j.rcim.2020.102072

Rai, A. K., M., J., Kumar, N., Katiyar, D., Singh, O., Bhojendra, Sreekumar, G., & Verma, P. (2023). Unlocking Productivity Potential: The Promising Role of Agricultural Robots in Enhancing Farming Efficiency. *International Journal of Plant & Soil Science*, *35*(18), 624–633. https://doi.org/10.9734/ijpss/2023/v35i183327

Roth, R. E., Finch, B. G., Blanford, J. I., Klippel, A., Robinson, A. C., & MacEachren, A. M. (2011). *Card Sorting For Cartographic Research and Practice. Cartography and Geographic Information Science, 38(2), 89–99.* doi:10.1559/1523040638289

Rubagotti, M., Tusseyeva, I., Baltabayeva, S., Summers, D., & Sandygulova, A. (2022). Perceived safety in physical human–robot interaction—A survey. *Robotics and Autonomous Systems*, *151*, 104047. https://doi.org/10.1016/j.robot.2022.104047

Sanders, T., Oleson, K. E., Billings, D. R., Chen, J. Y. C., & Hancock, P. A. (2011). A Model of Human-Robot Trust: Theoretical Model Development. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *55*(1), 1432–1436. https://doi.org/10.1177/1071181311551298

Schmettow, M., & Sommer, J. (2016). Linking card sorting to browsing performance – are congruent municipal websites more efficient to use? *Behaviour & Information Technology*, *35*(6), 452–470. https://doi.org/10.1080/0144929x.2016.1157207

Sherwani, F., Asad, M. M., & Ibrahim, B. S. K. K. (2020). Collaborative Robots and Industrial Revolution 4.0 (IR 4.0). In Editor(s) of the proceedings (Ed.), 2020

International Conference on Emerging Trends in Smart Technologies (ICETST), Karachi, Pakistan (pp. 1-5). doi:10.1109/ICETST49965.2020.9080724.

Simões, A. C., Pinto, A., Santos, J., Pinheiro, S., & Romero, D. (2022). Designing human-robot collaboration (HRC) workspaces in industrial settings: A systematic literature review. *Journal of Manufacturing Systems*, *62*, 28–43. https://doi.org/10.1016/j.jmsy.2021.11.007

Singh, J., Rhoads, G. K., & Detelina Marinova. (2016). Frontline Store Manager's Entrepreneurial Orientation for Merchandising and Service Role Performance: Scale Development and Validation. *Developments in Marketing Science: Proceedings of the Academy of Marketing Science*, 93–98. https://doi.org/10.1007/978-3-319-19428-8_24

Spencer, D., & Warfel, T. (2004). Card sorting: a definitive guide. *Boxes and arrows*, *2*(2004), 1-23.

Sriviboon, Y., & Jiamsanguanwong, A. (2022). *Usability Evaluation and User Acceptance of Cobot: Case Study of Universal Robots CB Series*. https://ieomsociety.org/proceedings/2022istanbul/362.pdf

Vicentini, F. (2020). Collaborative Robotics: A Survey. *Journal of Mechanical Design*, *143*(4). https://doi.org/10.1115/1.4046238

Weiss, A., Bernhaupt, R., Lankes, M., & Tscheligi, M. (2009). *The USUS evaluation framework for human-robot interaction*. https://www.researchgate.net/publication/313559458_The_USUS_evaluation_framework_for_human-robot_interaction

Appendix A

UX-HRC Delphi 2

Q8 **Participants' information sheet**

Before you decide to take part in this study it is important for you to understand why the research is being done and what it will involve. Please take a couple of minutes to read the following information carefully. A member of the team can be contacted (see below) if there is anything that is not clear or if you would like more information. Take time to decide whether or not you wish to take part in this consultation.

Purpose of the research

This research aims to contribute to the development of a new instrument to assess the experience of the users in the context of Human-Robot interaction and or Collaboration (HRI/HRC).

Results from the previous research phase

In the first consultation, we asked a group of 81 international experts on HRI/HRC about 15 dimensions that can be considered relevant in order to evaluate the User Experience in Human-Robot Interaction/Collaboration (7 robot-related aspects, 4 human-related aspects, and 4 context-related aspects). The prior survey led to a common consensus regarding 10 of the factors. However, it also revealed disagreement on 5 dimensions (2 robot-related aspects, 2 human-related aspects; 1 context-related aspect) which need further research.

In the following survey, we will ask you once again for your expert opinion on these 5 factors as well as ask some follow-up questions.

To have a look at the 15 dimensions and their descriptions, please open the description of the 15 Dimensions in a new tab.

What we are asking you to do

In this second consultation, we would like you to look at the 5 dimensions that resulted in a disagreement to gain more insights about the usefulness of these aspects in assessing user experience with robots. Specifically, for each dimension, we will ask you to perform the following three actions:

- Mandatory: Rate how much from 1 (strongly disagree) to 9 (strongly agree) you agree with the result of the previous consultation i.e., we should not consider this dimension among the main dimensions for assessing the UX after the interaction/collaboration with a robot?
- Optional: Why do you think there is disagreement with a certain dimension?
- Optional: Is there a way the dimension can be improved?

Expected time for the survey

To perform the mandatory actions, we do not expect you to invest more than **15-20 minutes** of your time. Of course, if you would like to provide us with additional insights and suggestions by filling in the optional fields this might increase the time of your consultation.

How will we use your data?

Your participation to the present study is voluntary and you can decide to quit at any time. Your personal data are going to be anonymised and used in the form of aggregated statistics for scientific purposes e.g., journal publications, conference presentations, etc. Only the researcher team will have access to your data and the data will be stored in a secure server in line with GDPR. This research project has been reviewed and approved by the Ethics Committee of the Faculty of Behavioural, Management and Social Sciences at the University of Twente. The project is a collaboration between University of Twente and University of Modena and Rggio Emilia. For questions regarding this study please contact the research team: Ásthildur Stefánsdóttir (a.l.stefansdottir@student.utwente.nl), Rufaro M. Hoto (r.m.hoto@student.utwente.nl), Dr Simone Borsci (s.borsci@utwente.nl) and Elisa Prati (elisa.prati@unimore.it).

Consent **Consent form**
I have read and understood the participant information sheet (Version October 2023) above. I voluntarily consent to be a participant in this study and understand that I can refuse to answer questions, and I can withdraw from the study at any time, without having to give a reason. I understand that personal information collected about me will not be shared beyond the research team. *

○ I understand and agree to participate voluntarily (1)

○ No, I would like to end this session (2)

Skip To: End of Survey If Consent form I have read and understood the participant information sheet (Version October 2023)... = No, I would like to end this session

End of Block: Introduction

Start of Block: Question introduction

Q13 In the next page we will show you individually each one of the 5 dimensions and their descriptions.

For each dimension we would like you to answer this question:
How much do you agree with the result of the previous consultation i.e., we should not consider this dimension among the main dimensions for assessing the UX after the interaction/collaboration with a robot?

Please, answer considering the dimensions and their descriptions and rate how much you agree with the previous expert opinion for each dimension on a 9-point scale ranging from 1 (Strongly disagree) to 9 (Strongly agree).

End of Block: Question introduction

Start of Block: Robot emotional appearance

Q8 **Robot's emotional appearance**
This aspect refers to how the robot's physical and behavioral characteristics, that delineate the "robot's emotional profile", can affect the user's judgment. In particular, it considers the following sub-factors: *Robot's Likeability* (e.g., happy, kind), *Warmth* (e.g., social, friendly), *Disturbance* (e.g., creepy, scary), *Discomfort* (e.g., awkward, dangerous), *Attractiveness*.

Results of previous consultation
Experts in the previous consultation **moderately disagree** about the relevance/importance of this dimension, suggesting removing or not considering this Dimension in the assessment of UX in HRI/HRC context.

Question:
How much do you agree with the result of the previous consultation i.e., we should not consider the <<Robot's emotional appearance>> among the main dimensions for assessing the UX after the interaction/collaboration with a robot? *

	1 (1)	2 (2)	3 (3)	4 (4)	5 (5)	6 (6)	7 (7)	8 (8)	9 (9)	
Strongly disagree	○	○	○	○	○	○	○	○	○	Strongly agree

Q27 (Optional) Why do you think there is disagreement with the importance of the dimension: **Robot's emotional appearance**? Please explain below.

Q25 (Optional) Do you think there is a way that the description of the dimension **Robot's emotional appearance** could be improved? If yes please explain below.

Q5 **Robot's social behavior.**
This dimension is described as: the user's judgment of the robot's social behavior considering parameters such as e.g., *Companionship*, *Initiative* (e.g., not giving orders, not being intrusive), *Social relationship* (e.g., telling its story, having a real exchange of opinion), *Social norms* (e.g., no knowledge, full knowledge), *Communication*.

Results of previous consultation: Experts in the previous consultation **strongly disagree** about the relevance/importance of this dimension, suggesting removing or not considering this dimension in the assessment of UX in HRI/HRC context.

Question: How much do you agree with the result of the previous consultation i.e., we

should not consider the <<Robot's social behavior>> among the main dimensions for assessing the UX after the interaction/collaboration with a robot? *

	1 (1)	2 (2)	3 (3)	4 (4)	5 (5)	6 (6)	7 (7)	8 (8)	9 (9)	
Strongly disagree	○	○	○	○	○	○	○	○	○	Strongly agree

Q28 (Optional) Why do you think there is disagreement with the importance of this dimension: **Robot's social behavior**? Please explain below.

Q26 (Optional) Do you think there is a way that the factor **Robot's social behavior** could be improved? If yes please explain below.

Q10 **Human-Factors personality-based.**
This dimension is described as: the user's self- description regarding their own personality characteristics, like e.g., *ethics* (e.g., social impact, social acceptance), *Personality traits*, *Self-confidence*, and *Personality to trust.*

Results of previous consultation: Experts in the previous consultation **strongly disagree** about the relevance/importance of this dimension, suggesting removing or not considering this dimension in the assessment of UX in HRI/HRC context.

Question: How much do you agree with the result of the previous consultation i.e., we should not consider the <<Human-Factors personality-based>> among the main dimensions for assessing the UX after the interaction/collaboration with a robot? *

	1 (1)	2 (2)	3 (3)	4 (4)	5 (5)	6 (6)	7 (7)	8 (8)	9 (9)	
Strongly disagree	○	○	○	○	○	○	○	○	○	Strongly agree

Q29 (Optional) Why do you think there is disagreement with the importance of the factor: **Human-Factors personality-based**? Please explain below.

Q28 (Optional) Do you think there is a way that the factor **Human-Factors personality-based** could be improved? If yes please explain below.

Q7 **Team involved during the task performance.**
This dimension is described as: the members involved in the specific task performed, considering e.g., *the number of humans and robots, Members' roles*.

Results of previous consultation: Experts in the previous consultation **moderately disagree**

about the relevance/importance of this dimension, suggesting removing or not considering this dimension in the assessment of UX in HRI/HRC context.
Question: How much do you agree with the result of the previous consultation i.e., we should not consider the <<Team involved during the task performance>> among the main dimensions for assessing the UX after the interaction/collaboration with a robot? *

	1 (1)	2 (2)	3 (3)	4 (4)	5 (5)	6 (6)	7 (7)	8 (8)	9 (9)	
Strongly disagree	○	○	○	○	○	○	○	○	○	Strongly agree

Q31 (Optional) Why do you think there is disagreement with the importance of the factor: **Team involved during the task performance**? Please explain below.

Q30 (Optional) Do you think there is a way that the factor **Team involved during the task performance** could be improved? If yes please explain below.

Q9 **Human judgment before the interaction with a cobot.**
This dimension is described as: the user's perception of the robot before the interaction, based on. *Perception and effect*, *anxiety* (e.g., toward communication capability, toward behavioral characteristics), *Attitudes toward use*, *Expectation* (e.g., performance expectancy, effort expectancy), *Acceptance*, *Perceived safety* (e.g., speed), *Trust* (e.g., Reliability), *Intention to use*.

Results of previous consultation: Experts in the previous consultation **moderately disagree** about the relevance/importance of this dimension, suggesting removing or not considering this dimension in the assessment of UX in HRI/HRC context.

Question: How much do you agree with the result of the previous consultation i.e., we should not consider the <<Human judgment before the interaction with a cobot>> among the main dimensions for assessing the UX after the interaction/collaboration with a robot? *

	1 (1)	2 (2)	3 (3)	4 (4)	5 (5)	6 (6)	7 (7)	8 (8)	9 (9)	
Strongly disagree	○	○	○	○	○	○	○	○	○	Strongly agree

Q30 (Optional) Why do you think there is disagreement with the importance of the factor: **Human judgment before the interaction with a cobot**? Please explain below.

Q27 (Optional) Do you think there is a way that the factor **Human judgment before the interaction with a cobot** could be improved? If yes please explain below.

TEXT PERSONAL INFO In order to better categorise your answers, please provide the following information:

Q28 Did you participate in the first phase of the research project? *

◯ Yes (1)

◯ No (2)

◯ Prefer not to say (3)

CONTRY In which country do you currently reside? *

▼ Afghanistan (1) ... Zimbabwe (1357)

SEX What is your sex (as assigned at birth)? *

◯ Female (1)

◯ Male (2)

◯ Prefer not to answer (3)

ROLE How do you describe your job in relation to HRC? Multiple answers are possible. *

- ☐ Robotics engineer (who design & built robots) (1)
- ☐ Workstation layout designer (e.g, selection of layout based on the production requirements, selection of hardware) (2)
- ☐ Software developer (e.g., robot programming, controller programming & development) (3)
- ☐ Hardware designer (e.g., design of new components, integration of multi-brand instrumentation) (4)
- ☐ Robot assembly worker (e.g., assembly of robot's mechanical components) (5)
- ☐ Human-factors specialist (e.g., user interface designer, ergonomist, phycologist) (6)
- ☐ Researcher (please specify the research area) (7) ____
- ☐ Robot user (i.e., if you use the robot for its final scope, e.g., assembly, physical support) (8)
- ☐ Other (please indicate) (9) ____

EXPERTISE How many years of experience do you have in HRC? *

- ○ Less than 1 year (1)
- ○ From more than 1 to 5 years (2)
- ○ From more than 5 to 10 years (3)
- ○ More than 10 years (4)

DOMAIN In which HRC application domain(s) is your experience? Multiple answers are possible. *

- ☐ Cobot for Industry (1)

- [ ] Cobot for Warehouse (2)
- [ ] Cobot for Healthcare (3)
- [ ] Cobot for Domestic (4)
- [ ] Cobot for Entertainment (5)
- [ ] Cobot for Military and police (6)
- [ ] Cobot for Space expedition (7)
- [ ] Cobot for Surgery (8)
- [ ] Cobot for Social (e.g., waitress, information support) (9)
- [ ] Cobot for Education (10)
- [ ] Cobot for Agriculture (11)
- [ ] Other (please indicate): (12) ____

ROBOT TYPE What type of robot(s) do you work on? If possible, please specify the robot model. Multiple answers are possible. *

- ☐ Collaborative robotic arm (1) ________________
- ☐ Humanoid robot (2) ________________
- ☐ Robot pet (3) ________________
- ☐ Autonomous Mobile Robot (4) ________________
- ☐ Automated Guided Vehicle (5) ________________
- ☐ Unmanned aerial vehicles (6) ________________
- ☐ Unmanned ground vehicles (7) ________________
- ☐ Unmanned underwater vehicle (8) ________________
- ☐ Toy (9) ________________
- ☐ Other (please indicate) (10) ________________

TYPE OF TASK Please provide an example of a task (e.g., assembly, physiotherapy) that the robot you are working with can perform:

DESIGN FLOW In order to improve our research, can you briefly describe what activities you and your team carry out during a HRC design project?

Q31 In case we will need to ask you additional questions we would like to have your email contact. Do you agree to be contacted in the future? *

- ○ Yes (please write here your email) (1) ____

- ○ No (2)

Appendix B

Rate how much you agree or disagree with the following statements regarding your collaboration with the [System/collaborative robot name].

1. **Easiness of robot regulation**
   Item 1. It was easy to physically regulate the [system/collaborative robot name].
   Item 2. I found it/ it appears to be easy to position the [System/collaborative robot name] components correctly.
   Item 3. I believe that from a physical point of view it appears to be easy to manipulate and put the [System/collaborative robot name] into position.

2. **Robot physical appearance.**
   Item 4. I had a positive impression of the [System/collaborative robot name]'s physical appearance.
   Item 5. I had a positive impression of the [System/collaborative robot name]'s dimensions i.e., high, width, length, weight.
   Item 6. I had a positive impression of the [System/collaborative robot name]'s features e.g., form, material.
   Item 7. The level of anthropomorphism (machinelike or humanlike) of the [System/collaborative robot name] was appropriate for the intended purpose.
   Item 8. The [System/collaborative robot name]'s perceived robustness (e.g., its ability to withstand physical stress, challenges etc.) met the specific requirements for the task and context of usage.
   Item 9. The type of robot (e.g., Robotic Arm, Humanoid Robot) seems appropriate for the task and context of usage.

3. **Robot's emotional appearance.**
   Item 10. I believe that the [System/collaborative robot name]'s emotional appearance was likable/attractive.
   Item 11. I believe that the [System/collaborative robot name]'s design does not cause emotional discomfort.
   Item 12. I believe that the [System/collaborative robot name]'s behavior does not cause emotional discomfort.
   Item 13. I did not consider the [System/collaborative robot name]'s appearance disturbing.
   Item 14. I felt that the [System/collaborative robot name] displayed a sense of warmth during our collaboration e.g., it was social, friendly.

4. **Robot's competence features.**
   Item 15. I perceived the [System/collaborative robot name] as competent and smart in terms of behavior.
   Item 16. I believe the [System/collaborative robot name] is reliable and trustworthy in terms of competencies.

Item 17. I found the [System/collaborative robot name] to be responsive and transparent in terms of competencies.

5. **Robot's physical behavior.**
   Item 18. I perceived the [System/collaborative robot name] movements to be smooth and flexible.
   Item 19. I believe that [System/collaborative robot name] is (physically) adaptable and autonomous.
   Item 20. I believe that the [system/collaborative robot name]'s physical behavior (e.g., noise, movement, autonomy, interactivity) during the interaction was suitable.
   Item 21. I believe that the [System/collaborative robot name] movements and behavior seemed lifelike and natural.

6. **Robot's social behavior.**
   Item 22. I believe that the overall social behavior of the [System/collaborative robot name] was appropriate.
   Item 23. I believe that the [System/collaborative robot name] acted and communicated according to social norms.
   Item 24. I felt that the [System/collaborative robot name] engaged in meaningful social interactions during our collaboration.
   Item 25. I think the [System/collaborative robot name] gave me a sense of companionship during our collaboration.
   Item 26. I perceived the [System/collaborative robot name] to be intrusive during our collaboration.

7. **Robot task performance.**
   Item 27. I believe that the [System/collaborative robot name] was useful by enabling a timely efficient performance.
   Item 28. I believe that the [System/collaborative robot name] was useful by enabling a correct (without error) performance.
   Item 29. I believe that the collaboration with the [System/collaborative robot name] was useful by enabling tasks to be completed in an efficient and effective manner.

8. **Human judgment before the interaction with a cobot (collaborative robot).**
   Item 30. I expected the collaboration with the [System/collaborative robot name] to be safe before using it.
   Item 31. I expected the [system/collaborative robot name] to be reliable and quick prior to collaborate with it.

Item 32. I did not experience any anxiety related to the [System/collaborative robot name] prior to the collaboration with it.
Item 33. My overall attitude towards using the [System/collaborative robot name] was positive prior to our collaboration.
Item 34. I accepted the idea of using the [System/collaborative robot name] for the task prior to use.
Item 35. Before interacting with the [System/collaborative robot name], I had the intention to use it for similar tasks or interactions in the future.

9. **Human judgment of the performance with a cobot** (collaborative robot)
Item 36. I realized while collaborating with the [System/collaborative robot name] that it is safe and trustworthy in use.
Item 37. I realized while collaborating with the [System/collaborative robot name] that it was pleasing to use and easy to control during the task.
Item 38. I felt comfortable during my collaboration with the [System/collaborative robot name].
Item 39. After using the [System/collaborative robot name], I found myself accepting of its role in the collaboration.
Item 40. I was highly satisfied with the [System/collaborative robot name]'s performance.
Item 41. I found the [System/collaborative robot name]'s interface or interaction methods (e.g., touch panel, voice commands, haptic feedback) highly usable.
Item 42. I experienced no frustration while working with the [System/collaborative robot name].
Item 43. I felt calm (e.g., no stress) during the interaction with the [System/collaborative robot name].
Item 44. I perceived an appropriate amount of cognitive workload during the collaboration with the [System/collaborative robot name].
Item 45. I would like to use the [System/collaborative robot name] again based on my experience during the task.

10. **Human-Factors personality-based.**
Item 46. I feel confident in my ability to use the [System/collaborative robot name] to achieve key tasks.
Item 47. I feel a sort of natural tendency to align well with the [System/collaborative robot name] during the collaboration to achieve certain goals.
Items 48. I do not see any ethical, personal, or social issues in collaborating with the [System/collaborative robot name] for my job.
Item 49. I have a trusting personality.
Item 50. I believe my personality traits have a significant influence on my collaboration with robots in general.

11. **Human-Factors ability-based.**
   Item 51. My level of expertise contributed to a successful interaction with the [System/collaborative robot name].
   Item 52. I believe that my general understanding of robotics contributed to a positive collaboration with the [System/collaborative robot name].
   Item 53. I believe that I have enough competence using collaborative robots to be able to properly handle the [System/collaborative robot name].

12. **Task performed.**
   Item 54. I believe that I can do this task more efficiently without the [System/collaborative robot name].
   Item 55. The task I performed with the [System/collaborative robot name] did not require too much physical effort.
   Item 56: I felt that the task I performed with the [System/collaborative robot name] did not require too much mental effort.
   Item 57: I believe the [System/collaborative robot name] was useful in accomplishing the task.
   Item 58: I felt that it was important to use the [System/collaborative robot name] to perform this critical task.

13. **The environment of interaction.**
   Item 59. The environmental conditions (e.g., lighting, noise, dust) were disturbing the task.
   Item 60. The workstation layout facilitated a positive interaction with the [System/collaborative robot name].
   Item 61: The interaction took place in a workstation with a layout that facilitated the completion of the task.
   Item 62. The elements present in the workstation during the interaction were suitable to achieve the task.
   Item 63. The [System/collaborative robot name] is well-suited for the demands of the applicational context.

14. **Team involved during the task performance.**
   Item 64. I believe that it is possible for multiple operators (a team) to collaborate proficiently to use the [System/collaborative robot name] to achieve the task.
   Item 65. When multiple operators (a team) have to collaborate interacting with the [System/collaborative robot name] all the operators can understand their roles.

15. **Interaction aspects.**
   Item 66. I was pleased with the overall interaction with the [system/collaborative robot name].

Item 67. I had good knowledge of the [System/collaborative robot name] status during the task performance.
Item 68. I found the [System/collaborative robot name]'s interface and interaction modality easy to use.
Item 69. I found the [System/collaborative robot name]'s interface and interaction modality easy to learn.
Item 70. I found the [System/collaborative robot name]'s interface and interactions modality easy to remember.
Item 71. I found the type of interface used for interaction (e.g., physical-based, graphical-based, vocal-based, gesture-based) was appropriate and effective.

Appendix C

Introduction:
Welcome to this Card Sorting study, and thank you for agreeing to participate!

Task Overview:
First you will be presented with a practise task, where you can get a feel for how the sorting in the experiment will work.

Then you will be presented with 10 items, and your goal is to categorize each of them into one of the 15 groups or, if you believe an item doesn't belong in any of the provided categories, you can place it into a special group labeled "None of the above."

The activity shouldn't take longer than 15 to 20 minutes to complete.

Your participation is anonymous and confidential. Your responses will be used solely for research purposes and will not be linked to your personal information.

Practise instructions:
THIS IS A PRACTISE ROUND AND WILL NOT COUNT.

Task Description:
- On the lower side of the screen, you will find 15 boxes of dimensions/groups that are relevant to the assessment of the user eXperience (UX) when operating collaborative robots.
- On the top of the screen, you will see two items related to this assessment, and your task is to sort the items into the dimension that best aligns with the provided statement.

Sorting Guidelines:
- Drag and drop each item into the dimension that you believe is the most appropriate based on the statement provided. More than one item can go into the same group.
- If you feel an item doesn't fit into any of the 15 dimensions, you can move it to a special group labeled "None of the above."
- You can make adjustments if you realize you initially placed an item in the wrong dimension.

Notes:
- There are no right or wrong sorting choices; trust your judgment and go with what feels correct to you.

Task instructions:
Please drag and drop these 10 items into the category you think is most appropriate.
More than one item can go into the same category.

If you hover your mouse over the title of each dimension you will see a description of the dimension. This will be very helpful during the sorting (it can take a few seconds to show up). If you would rather see a full list of the dimensions and their descriptions, click here.

If you feel that the dimensions or the items are not clear in any way, please take note of it. You will be able to express your opinions if you feel there is a lack of clarity on the next page of the survey.

Appendix D

Q9 What is your age?

____

Q10 What is your sex (as assigned at birth)?

- ○ Female (1)
- ○ Male (2)

Q11 Do you have any previous experience with collaborative robotics?

- ○ Yes (1)
- ○ No (2)
- ○ Im not sure (3)
- ○ Prefer not to say (4)

Q12 (Optional) Were the items and dimensions clear to you? If not, please explain why.

____

Appendix E

Item 1. It was easy to physically regulate the [system/collaborative robot name]. 83% (D1)
Item 2. I found it/ it appears to be easy to position the [System/collaborative robot name] components correctly. 60% (D1)
Item 3. I believe that from a physical point of view it appears to be easy to manipulate and put the [System/collaborative robot name] into position. 57% (D1)
Item 4. I had a positive impression of the [System/collaborative robot name]'s physical appearance. 100% (D2)
Item 5. I had a positive impression of the [System/collaborative robot name]'s dimensions i.e., high, width, length, weight. 80% (D2)
Item 6. I had a positive impression of the [System/collaborative robot name]'s features e.g., form, material. 67% (D2)
Item 7. The level of anthropomorphism (machinelike or humanlike) of the [System/collaborative robot name] was appropriate for the intended purpose. 63% (D2)
Item 8. The [System/collaborative robot name]'s perceived robustness (e.g., its ability to withstand physical stress, challenges etc.) met the specific requirements for the task. 63% (D7)
Item 10. I believe that the [System/collaborative robot name]'s emotional appearance was likable/attractive. 100% (D3)
Item 11. I believe that the [System/collaborative robot name]'s design does not cause emotional discomfort. 50% (D3)
Item 13. I did not consider the [System/collaborative robot name]'s appearance disturbing. 75% (D2)
Item 14. I felt that the [System/collaborative robot name] displayed a sense of warmth during our collaboration e.g., it was social, friendly. 57% (D3)
Item 15. I perceived the [System/collaborative robot name] as competent and smart in terms of behavior. 50% (D9)
Item 17. I found the [System/collaborative robot name] to be responsive and transparent in terms of competencies. 86% (D4)
Item 18. I perceived the [System/collaborative robot name] movements to be smooth and flexible. 88% (D5)
Item 20. I believe that the [system/collaborative robot name]'s physical behavior (e.g., noise, movement, autonomy, interactivity) during the interaction was suitable. 67% (D5)
Item 22. I believe that the overall social behavior of the [System/collaborative robot name] was appropriate. 88% (D6)
Item 23. I believe that the [System/collaborative robot name] acted and communicated according to social norms. 75% (D6)
Item 24. I felt that the [System/collaborative robot name] engaged in meaningful social interactions during our collaboration. 100% (D6)
Item 25. I think the [System/collaborative robot name] gave me a sense of companionship during our collaboration. 50% (D6)

Item 28. I believe that the [System/collaborative robot name] was useful by enabling a correct (without error) performance. 63% (D7)
Item 29. I believe that the collaboration with the [System/collaborative robot name] was useful by enabling tasks to be completed in an efficient and effective manner. 67% (D9)
Item 30. I expected the collaboration with the [System/collaborative robot name] to be safe before using it. 57% (D8)
Item 31. I expected the [system/collaborative robot name] to be reliable and quick prior to collaborate with it. 57% (D8)
Item 32. I did not experience any anxiety related to the [System/collaborative robot name] prior to the collaboration with it. 57% (D8)
Item 33. My overall attitude towards using the [System/collaborative robot name] was positive prior to our collaboration. 50% (D8)
Item 34. I accepted the idea of using the [System/collaborative robot name] for the task prior to use. 60% (D8)
Item 35. Before interacting with the [System/collaborative robot name], I had the intention to use it for similar tasks or interactions in the future. 86% (D8)
Item 42. I experienced no frustration while working with the [System/collaborative robot name]. 50% (D9)
Item 45. I would like to use the [System/collaborative robot name] again based on my experience during the task. 75% (D9)
Item 46. I feel confident in my ability to use the [System/collaborative robot name] to achieve key tasks. 50% (D11)
Items 48. I do not see any ethical, personal, or social issues in collaborating with the [System/collaborative robot name] for my job. 60% (D8)
Item 49. I have a trusting personality. 86% (D10)
Item 50. I believe my personality traits have a significant influence on my collaboration with robots in general. 86% (D10)
Item 51. My level of expertise contributed to a successful interaction with the [System/collaborative robot name]. 71% (D11)
Item 52. I believe that my general understanding of robotics contributed to a positive collaboration with the [System/collaborative robot name]. 57% (D11)
Item 55. The task I performed with the [System/collaborative robot name] did not require too much physical effort. 64% (D12)
Item 59. The environmental conditions (e.g., lighting, noise, dust) were disturbing the task. 100% (D13)
Item 60. The workstation layout facilitated a positive interaction with the [System/collaborative robot name]. 83% (D13)
Item 61. The interaction took place in a workstation with a layout that facilitated the completion of the task. 88% (D13)
Item 64. I believe that it is possible for multiple operators (a team) to collaborate proficiently to use the [System/collaborative robot name] to achieve the task. 63% (D14)

Item 65. When multiple operators (a team) have to collaborate interacting with the [System/collaborative robot name] all the operators can understand their roles. 50% (D14)

Item 69. I found the [System/collaborative robot name]'s interface and interaction modality easy to learn.50% (D1)

Appendix F

Rate how much you agree or disagree with the following statements regarding your collaboration with the [System/collaborative robot name] .

1. **Easiness of robot regulation**
   Item 1. It was easy to physically regulate the robot.
   Item 2. I found it/ it appears to be easy to position the robot's components correctly.
   Item 3. I believe that from a physical point of view it appears to be easy to manipulate and put the robot into position.

2. **Robot physical appearance.**
   Item 4. I had a positive impression of the robot's physical appearance.
   Item 5. I had a positive impression of the robot's dimensions i.e., high, width, length, weight.
   Item 6. I had a positive impression of the robot's features e.g., form, material.

3. **Robot's emotional appearance.**
   Item 10. I believe that the robot's emotional appearance was likable/attractive.
   Item 11. I believe that the robot's design does not cause emotional discomfort.
   Item 14. I felt that the robot displayed a sense of warmth during our collaboration e.g., it was social, friendly.

4. **Robot's competence features.**
   Item 16. I believe the robot is reliable and trustworthy in terms of competencies.
   Item 17. I found the robot to be responsive and transparent in terms of competencies.
   Item 39. After using the robot, I found myself accepting of its competence during the collaboration.

5. **Robot's physical behavior.**
   Item 18. I perceived the robot's movements to be smooth and flexible.
   Item 20. I believe that the robot's physical behavior (e.g., noise, movement, autonomy, interactivity) during the interaction was suitable.
   Item 21. I believe that the robot's movements and behavior seemed lifelike and natural.

6. **Robot's social behavior.**
   Item 22. I believe that the overall social behavior of the robot was appropriate.
   Item 23. I believe that the robot acted and communicated according to social norms.
   Item 24. I felt that the robot engaged in meaningful social interactions during our collaboration.

7. **Robot task performance.**

Item 8. The robot's perceived robustness (e.g., its ability to withstand physical stress, challenges etc.) met the specific requirements for the task.
Item 28. I believe that the robot was useful by enabling a correct (without error) performance.

8. **Human judgment before the interaction with a cobot (collaborative robot).**
   Item 34. I accepted the idea of using the robot for the task prior to use.
   Item 35. Before interacting with the robot, I had the intention to use it for similar tasks or interactions in the future.
   Items 48. I do not see any ethical, personal, or social issues in collaborating with the robot for my job.

9. **Human judgment of the performance with a cobot** (collaborative robot)
   Item 29. I believe that the collaboration with the robot was useful by enabling tasks to be completed in an efficient and effective manner.
   Item 42. I experienced no frustration while working with the robot.
   Item 45. I would like to use the robot again based on my experience during the task.

10. **Human-Factors personality-based.**
    Item 43. I felt calm (e.g., no stress) during the interaction with the robot.
    Item 49. I have a trusting personality.
    Item 50. I believe my personality traits have a significant influence on my collaboration with robots in general.

11. **Human-Factors ability-based.**
    Item 46. I feel confident in my ability to use the robot to achieve key tasks.
    Item 51. My level of expertise contributed to a successful interaction with the robot.
    Item 52. I believe that my general understanding of robotics contributed to a positive collaboration with the robot.

13. The environment of interaction.
Item 59. The environmental conditions (e.g., lighting, noise, dust) were disturbing the task.
Item 60. The workstation layout facilitated a positive interaction with the robot.
Item 61. The interaction took place in a workstation with a layout that facilitated the completion of the task.

14. **Team involved during the task performance.**
    Item 64. I believe that it is possible for multiple operators (a team) to collaborate proficiently to use the robot to achieve the task.
    Item 65. When multiple operators (a team) have to collaborate interacting with the robot all the operators can understand their roles.

15. **Interaction aspects.**

Item 41. I found the robot's interface or interaction methods (e.g., touch panel, voice commands, haptic feedback) highly usable.
Item 66. I was pleased with the overall interaction with the robot.
Item 71. I found the type of interface used for interaction (e.g., physical-based, graphical-based, vocal-based, gesture-based) was appropriate and effective.