Transdisciplinarity and the Future of Engineering
B.R. Moser et al. (Eds.)

© 2022 The authors and IOS Press.
This article is published online with Open Access by IOS Press and distributed under the terms of the Creative Commons Attribution Non-Commercial License 4.0 (CC BY-NC 4.0).
doi:10.3233/ATDE220627

A Systematic Literature Review of User Experience Evaluation Scales for Human-Robot Collaboration

Elisa PRATI[a,1], Simone BORSCI[b], Margherita PERUZZINI[a], Marcello PELLICCIARI[c]
[a] *Department of Engineering "Enzo Ferrari", University of Modena and Reggio Emilia, Modena, Italy*
[b] *Faculty of Behavioural Management and Social sciences, Department of Learning, Data analysis, and Technology, group of Cognition, Data and Education, University of Twente, Enschede, Netherlands*
[c] *Department of Sciences and Methods for Engineering, University of Modena and Reggio Emilia, Reggio Emilia, Italy*

Abstract. In the last decade, the field of Human-Robot Collaboration (HRC) has received much attention from both research institutions and industries. Robot technologies are in fact deployed in many different areas (e.g., industrial processes, people assistance) to support an effective collaboration between humans and robots. In this transdisciplinary context, User eXperience (UX) has inevitably to be considered to achieve an effective HRC, namely to allow the robots to better respond to the users' needs and thus improve the interaction quality. The present paper reviews the evaluation scales used in HRC scenarios, focusing on the application context and evaluated aspects. In particular, a systematic review was conducted based on the following questions: (RQ1) *which evaluation scales are adopted within the HRI scenario with collaborative tasks?*, and (RQ2) *how the UX and user satisfaction are assessed?*. The records analysis highlighted that the UX aspects are not sufficiently examined in the current HRC design practice, particularly in the industrial field. This is most likely due to a lack of standardized scales. To respond to this recognized need, a set of dimensions to be considered in a new UX evaluation scale were proposed.

Keywords. Human-Robot Collaboration, User experience evaluation, User experience scale, Transdisciplinary engineering, Review

Introduction

The research in Human-Robot Interaction (HRI) and the integration of robots in the manufacturing industries to support human activities have seen exponential growth, especially in the last ten years. This evolution in HRI research and application mainly regards Human-Robot Collaboration (HRC) solutions, considered as a HRI subcategory. HRC is characterized by the presence of human and cobots sharing the same space at the same time while performing tasks to achieve a common goal.

The complexity of a HRC scenario requires a transdisciplinary team that involves experts in robotics (e.g., software engineers, mechatronic engineers) and human factors

[1] Corresponding Author, Mail: elisa.prati@unimore.it

(e.g. psychologist, cognitive ergonomist). The adoption of a transdisciplinary approach is important to overcome the limitations of traditional engineering, by involving social sciences and human factors in technical areas [1] that are more focused on human aspects. Indeed, HRI is traditionally focused on technology, but poorly on people. Merging human factors and robotics can leverage the applications to higher interaction quality and better UX.

For the advancement of the HRC field, it is necessary to evaluate the quality of the interaction between humans and robots. The scientific community has recognized the need of increasing the evaluation practices and, in particular, of introducing standardized tools (e.g., [2]). As stated in [3], "*if we are to make progress in the field of human-robot interaction then we shall have to develop standardized measurement tools"*.

Many methods (e.g. interview, questionnaire, expert evaluation, user observation) exist for the user experience and usability evaluation, particularly used in the Human-Machine Interaction (HMI) field. Compared to other evaluation methods, the questionnaire (or scale) allows to involve users and collect quantitative data in a structured way. Moreover, the evaluation scale is one of the most widespread techniques in the HRI field, and only few different techniques were applied (e.g. Heart rate variability and Electrocardiogram [4], Electroencephalogram [5], Virtual Reality [6], [7][8]). At the moment, the evaluation phase in HRI has been considered mainly by looking at the robotics components (e.g. performance criteria, technological aspects), especially in the industrial field. However, human aspects must be considered during all the design phases, i.e. from research to testing [9]. In fact, when there is a direct interaction between human and robot, and the robot work considerably impacts the human actions, also the user satisfaction must be evaluated. This is directly related to the concept of User eXperience (UX), defined as the perception and response that a person has consequently to the use of a product/system/service [10]. As pointed out in [11], the *"successful evaluation of user experience is important for managing a technology, offering design guidance and improving the way humans interact with the technology"*. Although the importance of the UX evaluation in HRC is already known (e.g. [3], [12]–[16]), there is a recognized gap in the literature of standardized tools for the UX evaluation [3], [4], [17], especially in the industrial field. In fact, there is a restricted number of studies that have investigated the user perception during the interaction with a cobot, proposing: frameworks of HRI classification (e.g. [18], [19]) or general metrics (e.g. [20]) that can guide the initial testing phase setup; new evaluation scales (e.g. [21], [22]) to collect the users' opinion about different aspects (e.g. trust, robot perception). The existing scales mainly regard social robots (e.g. assistant robots), which, however, need to consider different aspects than a HRC task in industrial contexts. Moreover, most of these researches are not recent and thus probably some considerations about the newest evolution of the HRI field are missed, requiring a review of the proposed framework or scales. The absence of standardized methods for the HRI evaluation from a human point of view (i.e. UX, user satisfaction, usability) slows down the improvement of robots performance to better meet human interaction needs and facilitate collaboration. To respond to this exigence, this paper aims to lay the foundation for the development of a standardized UX evaluation scale for HRC scenario in industrial context (UX-HRC), where the user directly interacts with the cobot (i.e. physical and haptic interfaces). The scenario that included the use of external interfaces (e.g. graphic interfaces) to mediate the interaction were excluded because in that case other specific scales are required (e.g. System Usability Scale).

The remaining of the paper is structured as follows: Section 1 reports the applied method used to conduct the research, Section 2 presents and discusses the resulting evaluation scales. Lastly, the concluding remarks and considerations regarding the future research directions are given in Section 3, based on the results of this survey.

1. Method

This systematic review was performed on journal articles examining the use of scales for the evaluation of HRI, in particular looking for whether and how the user experience is assessed. In particular, the survey sought to answer the following research questions:

RQ1: Which evaluation scales are adopted within the HRI scenario with collaborative tasks?
RQ2: How the UX and user satisfaction are assessed?

Records were retrieved from Scopus, Web of Science and Google scholar using the Boolean operators (AND/OR) to combine the following keywords: human-robot interaction (AND) user experience / user satisfaction / usability (AND) scale.

Given the specificity of the researched topic, temporal filters were not applied. The selected documents are written in english language and contain one or more of the following items:

- evaluation scale of framework in HRI;
- UX, user satisfaction or usability in different HRI application fields (e.g. industry, society).

On the other hand, were excluded:

- scale for the evaluation of not direct interaction (e.g., graphic interface, gestures);
- no-collaborative tasks, especially if not regarding industry fields.

2. Results and discussion

A total of 165 papers were identified from the research. After removing the duplicates, a scan of the remaining 95 records by title and abstract was performed. Consequently, the full text of the most pertinent papers was scanned, including other papers from references and citations of the resulted papers.

For an easier reading, the main results were summarized in a table (accessible from the Supplementary material section), which reports:

- the publication year, to highlight the content updating;
- if a general framework or a scale is proposed and which topic is discussed;
- the research field, to indicate if it regards industrial or social robots;
- the list of the considered factors, to make visible which are the most explored in the current literature;
- an indication (green background, and *) of the factors that can be included in the future UX-HRC scale.

Looking at the items included in the review, five researches proposed a general framework that can be useful to set up the HRI evaluation, while the others propose new assessment scales to assess specific aspects related to HRC. These scales were mainly developed considering social robotics scenarios, particularly for assistance application, underlining how this sector is more focused on the quality of human and robots collaboration than industrial ones.

2.1. General framework

The researches that propose a general framework or have identified factors for HRI evaluation can be considered as primary attempts toward a structured way to evaluate the HRC quality level taking into account the user satisfaction. From the HRI classifications proposed in [18] and in [19], several aspects at the base of an HRI evaluation can be extrapolated, like: robot typology (e.g. industrial robot, professional robot, personal service robot), interface type (e.g. graphic, speech, gesture, physical), human role and characteristics, context of use and task type (e.g. synchronous or not, collocated or not).

In [20], a set of metrics to assess human-robot tasks execution considering the performance of the system, operator and robot, are suggested. In particular, the operator performance is measured by the situation awareness, the workload and the accuracy of mental models of device operation. Although these parameters do not aim to investigate the user perception, they all impact the user satisfaction and must be considered in a proper UX evaluation tool for HRC. Moreover, it is worth noting that the HRC performance is affected by different factors, such as: operational factors (e.g. tactics, time on station), equipment factors (e.g., physical parameters, workspace layout), task factors (e.g. complexity,repetitiveness), personnel factors (e.g., training, motivation, stress), and external environmental factors (illumination, visibility).

Human perception is more considered in [22], where the authors proposed a human-centered evaluation framework (i.e., USUS) for HRC with humanoid robots in a working environment. With the aim to better understand how to improve robot design, the USUS includes four factors that impact the interaction between humans and robots: usability, social acceptance, user experience, and societal impact. In particular, for the user experience factor the authors identified five dimensions to be evaluated: embodiment, emotion, human-oriented perception, feeling of security and co-experience with robots. Moreover, for each defined dimension are proposed methods for their evaluation, but without a precise explanation of how to apply them, limiting the framework use to no-expert.

2.2. Scales to assess specific aspects of HRC

In the specific domain of social robot interaction user satisfaction toward the service of the robot [3] is often considered the result of multiple aspects, specifically: acceptance, anthropomorphism, trust, but also users' personality and individual functioning.

In robot perception, acceptance is a fundamental topic in the design and evaluation of robots thought to socially interact with humans. For example, in [23] are investigated the variables that influence the acceptance of social robots by potential users, identifying six variables**:** usefulness, adaptability, enjoyment, sociability, companionship and perceived behavioral control. For sure, acceptability is an important aspect to consider

also in new industry 4.0 scenarios, where operators are asked to accept robots as work partners, impacting the work quality and operators' experience.

The robot acceptance is influenced by anthropomorphism (humanlike) aspects, increasing the robot's familiarity [24] and infecting the HRI [25]. Its influence in human perception is mainly investigated in social robots until now. Anthropomorphism is a social phenomenon that corresponds to people's tendency to attribute human-like characteristics (e.g. shape, facial expression, natural communication) to objects [24]. As pointed out in [24], several measurements (e.g. physiological benchmark, gaze cues) and methods (e.g. questionnaire) have been proposed for the anthropomorphism evaluation in social interaction. A relevant research about HRI evaluation is described in [3], where authors propose five questionnaires with the intent to develop a standardized tool for the HRI measurement. Being mostly related to social robots, the questionnaires (called Godspeed) consider five factors: anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots. Each factor is evaluated by multiple items using 5-points semantic differential scales. Thanks to its simplicity of use and its multilingual translation [26], the scale has a widespread use. Although this, the scale presents several structural problems [16], [27]. A further analysis of the Godspeed scale was conducted in [16], where authors have considered also a novel set of attributes focused on social judgment. From these analyses they developed the Robotic Social Attributes Scale (RoSAS) to measure social perception of robots, composed of three factors (warmth, competence, discomfort) and 18 items.

A more recent scale, reported in [2], evaluates how people perceive robots and attribute human characteristics to them. The Human-Robot Interaction Evaluation Scale (HRIES) consider four factors and their items: sociability (warm, likeable, trustworthy, friendly), animacy (alive, natural, real, human-like), agency (self-reliant, rational, intentional, intelligent), disturbance (creepy, scary, uncanny, weird). Using a 7-point Likert scale (1= not at all, 7= totally), it is asked to answer the question "*How closely are the words below associated with [the item]*?". The identified items can be a good starting point also for the anthropomorphism investigation in collaborative industrial robots. All these researches demonstrate that perception, analyzed through anthropomorphism and acceptance aspects, is an important point to be taken into account in HRI evaluation.

The trust plays a central role in teammate tasks [28], as could be a human-cobot scenario. As stated in [29], *"trust is necessary for humans to fully realize a robot's benefits to human-robot teams, but gaining trust is one of the most difficult challenges in design and implementation"*. The trust impact in the interaction outcomes is well known, and it is influenced by many factors. In [29], a human-robot trust model structured in three categories and dimensions is proposed: i) Robot characteristics *(Performance-Based: behavior, dependability, reliability of robot, predictability, level of automation, failure rates, false alarms, transparency; Attribute-Based: proximity/co-location, robot personality, adaptability, robot type, anthropomorphism)*, ii) Human characteristics *(Ability-Based: demographics, engagement, expertise, competency, operator workload, prior experiences, situation awareness; Personality-Based: personality traits, attitudes towards robots, comfort with robot, self-confidence, propensity to trust)* and, iii) Environment characteristics *(Team Collaboration: in-group membership, culture, communication, shared mental models; Tasking: task type, task complexity, multi-tasking requirement, physical environment)*.

The perceived trust is strongly correlated by anthropomorphism [30], impacting the user's understanding of how to interact with a robot [31], [32] and increasing the sense of trust, acceptance and enjoyment. Van Pinxteren et al. [30] conducted an experiment

about trust in humanoid robots for services marketing and have defined a questionnaire to collect the users' perception. The five measured aspects were based on other studies, and regard: perceived anthropomorphism, perceived interaction comfort, trust, perceived enjoyment and intention to use.

A specific scale to evaluate trust in industrial HRC was proposed in [21]. The authors developed a psychometric scale composed of 10 statements that the user has to rate after the interaction with an industrial collaborative robot.

The number of research around the trust topic and the aspects that influenced it, underline the importance of trust consideration in a scale that aims to evaluate the whole user experience during a collaborative task.

2.3. Users' personality and individual functioning

In [33], a post-experimental questionnaire was developed for the evaluation of a human–humanoid collaborative physical interaction in an assembly task. The questions aim to acquire the user's impression about the task, the general interaction experience and the physical interaction with the robot. To the best of our knowledge, the questionnaire has not been validated before the use but can represent a significant starting point for the development of a new UX-HRC evaluation scale. The work of Ivaldi et al. suggests that the users' individual factors (i.e. social attitudes and personality traits) influence the engagement sensation and the interaction perception. The study shows that the user's behavior (e.g. exchange of gaze, the temporal dynamics of speech) observation is at the base of the most common metrics to evaluate the engagement during HRI tasks. However, the behavior depends also on individual factors (e.g. extroversion and negative attitude toward robots) and, for this reason, the researchers have used two scales: NEO-PIR (Revised Personality Inventory) [34] and NARS (Negative Attitude towards Robots Scale) [35]. The latter is composed of three subscales (*Negative attitude toward situations and interactions with robots, Negative attitudes toward social influence of robots, Negative attitudes toward emotions in interaction with robots*), adopting a 5-point scale (from 'Strongly disagree' to 'Strongly agree'). The NARS is used to measure the users' attitudes towards robots and how it changes over time as a result of prolonged interactions to better understand the user behavior [36]. From Syrdal et al. [37] research emerged that cultural differences may impact how participants evaluate their interaction with a robot as well as the robot's behavior. For this reason, a good option could be the combination of NARS with other scales for a better data interpretation based on the specific participant. For example, in [36] the NARS have been associated with the Robot Anxiety Scale (RAS) [38] for human behavior prediction in HRI. In particular, the RAS was developed to measure the anxiety level that may be evoked by robots. It is composed by three subscales (*Anxiety toward communication capacity of robots, Anxiety toward behavioral characteristics of robots, Anxiety toward discourse with robots*) and each item is scored using a 6-point scale (from 'I do not feel anxiety at all' to 'I feel very anxious').

Regarding the issue of how users judge their interaction with a robot, a Self-Efficacy in HRI scale (SE-HRI) was proposed in [14]. The scale was developed and validated in German and English versions, proposing a long version (18 items) and a short version (10 items), obtaining a good model fit and a satisfactory construct. Reasoning on the self-efficacy evaluation, it is possible to notice how it is connected to the user's expectations. The latter strongly affects the user's opinion about a product or an experience [39]: the more the result exceeds the expectations, the more the evaluation of the UX by the user

will be positive. On the contrary, if the expectations are disappointed, the UX is evaluated more negatively.

Another key human centered aspect considered relevant to assess the quality of HRC in literature is the one individual functioning, for instance when robots are collaborating with people with disabilities or when robots are used by elderly with and without health issues or comorbidities. In [40], three scales for the user satisfaction evaluation of a mobility assistant robot have been used for people with disabilities or under rehabilitation: the QUEST 2.0 (Quebec User Evaluation of Satisfaction with assistive Technology) [41], the ATD PA (Assistive Technology Device Predisposition Assessment) [42] and the PYTHEIA (A Scale for Assessing Rehabilitation and Assistive Robotics) [43].

In another study regarding the use of a humanoid robot for the monitoring and training of older adults under rehabilitation [44], the UX during the interaction was evaluated considering three factors based on existing scales: perceived usefulness [45], the control [46] and the enjoyment [47]. In the experiment, a graphic interface (i.e. tablet) and a humanoid robot (i.e. NAO) were compared, however the items used to investigate the HRI are quite general, but this set could represent a starting point for the formulation of the new UX-HRC scale.

A wide questionnaire was developed in [48] by measuring the perceived utility during the interaction with an assistant robot (i.e. Pepper). The dimensions and items were based on existing frameworks [49], [50] and study [51] about elderly people's expectation of the interaction with a robot. The questionnaire is composed for the most of a list of closed questions (yes/no) regarding: functionality, safety, appearance, mutual care, perception and affect, privacy, control and initiative, friendly relationship and complicity, social relationship, concerns about status. Some of these dimensions can be taken in consideration in the development of the new scale, while the used questions are very specific of elderly people and assistant robot scenarios.

3. Conclusion and future works

The evaluation of the interaction between humans and robots represent a fundamental phase for the HRC evolution in industry and especially for HRC scenarios it is important to consider the user point of view regarding the perceived quality of interaction. The present review shed some light on the factors that are considered by researchers as the most relevant to assess HRC, by also highlighting items and scales that are currently used in the UX and user satisfaction assessment in the HRC scenario. Moreover, the analysis indicated that the urgency of assessing in a systematic and comparable way the HRC is a priority in the social robotics community, while this seems less of a priority for the community around industrial robotics.

Currently the UX aspects are not sufficiently considered during the testing phase of a HRC design process. In particular, it emerged that:

- there is a general absence and need of standardized tools (e.g. scales) for the HRC evaluation;
- UX and user satisfaction were considered in few studies, but mainly in the social robot sector;
- several proposed frameworks and scales were developed many years ago and would need to be adapted to the more recent technologies;

These research outputs make it evidence that there are not appropriate and ready-to-use scales for the UX evaluation of an industrial HRC scenario, and the existent ones require a significant revision. In particular, it would be necessary to better consider the typology of the task and the characteristics of an industrial environment that can impact on the UX. On the other hand, all the existing research around the evaluation scale topic represents a valid starting point for the development of a new UX-HRC scale. In particular, some of the factors emerged from the survey may result useful to investigate the UX within the industrial field, e.g.: Robot's behaviour, Self-efficacy (intended as user preparation), Acceptance, Perception about the performance, Physical interaction, Safety, Control, Comfort, Trust, Task factors, Environment factors. Another important factor to be considered is the User expectation, especially for evaluations conducted after the implementation of a new HRC scenario. For a comprehensive analysis and data interpretation, the future UX-HRC scale should be used together with the already existing NARS scale.

For the development of a UX-HRC standard scale for the industrial sector, several future research activities will be necessary. For example, HRC experts and end users must be involved in interviews and focus group sessions, and a literature review of the UX evaluation in different fields is to be performed. Next, application testing phases must be conducted for a proper validation. The development of a proper UX-HRC scale is a strongly interdispliplinary process which requires the involvement of experts in different disciplines from both the human sciences and the engineering areas.

Supplementary material

A comparative table of the reviewed papers is available at:
https://drive.google.com/drive/folders/1j7vh50ogqmN_lCXhuLRMdx_QY1cuopVL?usp=sharing.

References

[1] N. Wognum, C. Bil, F. Elgh, M. Peruzzini, J. Stjepandić, and W. J. C. Verhagen, Transdisciplinary systems engineering: Implications, challenges and research agenda, *International Journal of Agile Systems and Management*, 2019, vol. 12, no. 1, pp. 58–89.

[2] N. Spatola, B. Kühnlenz, and G. Cheng, Perception and Evaluation in Human–Robot Interaction: The Human–Robot Interaction Evaluation Scale (HRIES)—A Multicomponent Approach of Anthropomorphism, *International Journal of Social Robotics,* 2021, vol. 13, no. 7, pp. 1517–1539.

[3] C. Bartneck *et al.*, Measurement Instruments for the Anthropomorphism, Animacy, Likeability, Perceived Intelligence, and Perceived Safety of Robots, *Int J Soc Robot*, 2009, vol. 1, pp. 71–81.

[4] S. Shao, T. Wang, Y. Li, C. Song, Y. Jiang, and C. Yao, Comparison Analysis of Different Time-Scale Heart Rate Variability Signals for Mental Workload Assessment in Human-Robot Interaction, *Wireless Communications and Mobile Computing*, vol. 2021, 8371637, https://doi.org/10.1155/2021/8371637.

[5] L.D.J. Fiederer, M. Völker, R.T. Schirrmeister, W. Burgard, J. Boedecker, and T. Ball, Hybrid brain-computer-interfacing for human-compliant robots: Inferring continuous subjective ratings with deep regression, *Frontiers in Neurorobotics*, 2019, vol. 13, 76.

[6] E. Prati, V. Villani, M. Peruzzini, and L. Sabattini, An Approach Based on VR to Design Industrial Human-Robot Collaborative Workstations, *Applied Sciences*, Dec. 2021, vol. 11, no. 24, 11773.

[7] V. Villani, B. Capelli, and L. Sabattini, Use of Virtual Reality for the Evaluation of Human-Robot Interaction Systems in Complex Scenarios, *RO-MAN 2018 - 27th IEEE International Symposium on Robot and Human Interactive Communication*, Nov. 2018, pp. 422–427.

[8] F. Grandi, M. Peruzzini, S. Cavallaro, E. Prati, and M. Pellicciari, Creation of a UX index to design human tasks and workstations, *International Journal of Computer Integrated Manufacturing*, 2022, vol. 35, no. 1, pp. 4–20.

[9] E. Prati, M. Peruzzini, M. Pellicciari, and R. Raffaeli, How to include User eXperience in the design of Human-Robot Interaction, *Robotics and Computer-Integrated Manufacturing*, Apr. 2021, vol. 68, 102072.

[10] N.N., ISO 9241-210:2010 Ergonomics of human-system inter- action – part 210: human-centred design for interactive systems. Brussels (Belgium): CEN; 2010.

[11] N. McNamara and J. Kirakowski, Functionality, usability, and user experience, *Interactions*, Nov. 2006, vol. 13, no. 6, pp. 26–28.

[12] M. M. Scheunemann, R. H. Cuijpers, and C. Salge, Warmth and Competence to Predict Human Preference of Robot Behavior in Physical Human-Robot Interaction, *29th IEEE International Conference on Robot and Human Interactive Communication, RO-MAN 2020*, Aug. 2020, pp. 1340–1347.

[13] A. Weiss and C. Bartneck, Meta analysis of the usage of the Godspeed Questionnaire Series, *Proceedings - IEEE International Workshop on Robot and Human Interactive Communication*, vol. 2015-November, pp. 381–388.

[14] A. R. von der Pütten and N. Bock, Development and Validation of the Self-Efficacy in Human-Robot-Interaction Scale (SE-HRI), *ACM Transactions on Human-Robot Interaction (THRI)* , Dec. 2018, vol. 7, no. 3, pp. 1-30.

[15] R. E. Yagoda, Development of the human robot interaction workload measurement tool (HRI-WM), *Proceedings of the human factors and Ergonomics society annual meeting*, Sep. 2010, vol. 54, no. 4, pp. 304–308.

[16] C. M. Carpinella, A. B. Wyman, M. A. Perez, and S. J. Stroessner, The Robotic Social Attributes Scale (RoSAS): Development and Validation, *ACM/IEEE International Conference on Human-Robot Interaction*, Mar. 2017, vol. Part F127194, pp. 254–262.

[17] M. Paliga and A. Pollak, Development and validation of the fluency in human-robot interaction scale. A two-wave study on three perspectives of fluency, *International Journal of Human-Computer Studies*, Nov. 2021, vol. 155, p. 102698.

[18] S. Thrun, Toward a Framework for Human-Robot Interaction, *Human–Computer Interaction*, 2004, vol. 19, no. 1–2, pp. 9–24.

[19] H.A. Yanco and J. Drury, Classifying human-robot interaction: An updated taxonomy, *Conference Proceedings - IEEE International Conference on Systems, Man and Cybernetics*, 2004, vol. 3, pp. 2841–2846.

[20] A. Steinfeld *et al.*, Common metrics for human-robot interaction, *HRI 2006: Proceedings of the 2006 ACM Conference on Human-Robot Interaction*, vol. 2006, pp. 33–40.

[21] G. Charalambous, S. Fletcher, and P. Webb, The Development of a Scale to Evaluate Trust in Industrial Human-robot Collaboration, *International Journal of Social Robotics*, 2016, vol. 8, pp. 193–209.

[22] A. Weiss, R. Bernhaupt, M. Lankes, and M. Tscheligi, The USUS Evaluation Framework for Human-Robot Interaction, *AISB2009: proceedings of the symposium on new frontiers in human-robot interaction*, 2009, vol. 4, pp. 11–26.

[23] M. M. A. de Graaf and S. ben Allouch, Exploring influencing variables for the acceptance of social robots, *Robotics and Autonomous Systems*, Dec. 2013, vol. 61, no. 12, pp. 1476–1486.

[24] J. Fink, Anthropomorphism and Human Likeness in the Design of Robots and Human-Robot Interaction, *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 7621, 2012, pp. 199–208.

[25] E. P. Bernier and B. Scassellati, The similarity-attraction effect in human-robot interaction, *2010 IEEE 9th International Conference on Development and Learning, ICDL-2010 - Conference Program*, 2010, pp. 286–290.

[26] N.N., The Godspeed Questionnaire Series - Christoph Bartneck, Ph.D., https://www.bartneck.de/2008/03/11/the-godspeed-questionnaire-series/ (accessed Apr. 14, 2022).

[27] C. C. Ho and K. F. MacDorman, Revisiting the uncanny valley theory: Developing and validating an alternative to the Godspeed indices, *Computers in Human Behavior*, Nov. 2010, vol. 26, no. 6, pp. 1508–1518.

[28] J. D. Lee and K. A. See, Trust in automation: Designing for appropriate reliance, *Human Factors*, Aug. 2004, vol. 46, no. 1, pp. 50–80.

[29] T. Sanders, K. E. Oleson, D. R. Billings, J. Y. C. Chen, and P. A. Hancock, A Model of Human-Robot Trust: Theoretical Model Development, *Proceedings of the Human Factors and Ergonomics Society Annual Meeting.*, Sep. 2011, pp. 1432–1436.

[30] M. M. E. van Pinxteren, R. W. H. Wetzels, J. Rüger, M. Pluymaekers, and M. Wetzels, Trust in humanoid robots: implications for services marketing, *Journal of Services Marketing*, Sep. 2019, vol. 33, no. 4, pp. 507–518.

[31] T. Fong, I. Nourbakhsh, and K. Dautenhahn, *A Survey of Socially Interactive Robots: Concepts, Design, and Applications*, The Robotics Institute, Carnegie Mellon University, 2002.

[32] B. R. Duffy, Anthropomorphism and the social robot, *Robotics and Autonomous Systems*, Mar. 2003, vol. 42, no. 3–4, pp. 177–190.

[33] S. Ivaldi, S. Lefort, J. Peters, M. Chetouani, J. Provasi, and E. Zibetti, Towards Engagement Models that Consider Individual Factors in HRI: On the Relation of Extroversion and Negative Attitude Towards Robots to Gaze and Speech During a Human-Robot Assembly Task Experiments with the iCub humanoid, *International Journal of Social Robotics*, 2017, vol. 9, pp. 63–86.

[34] P. T. Costa and R. R. McCrae, The revised NEO personality inventory (NEO-PI-R), *The SAGE Handbook of Personality Theory and Assessment: Volume 2 - Personality Measurement and Testing*, pp. 179–198, Jan. 2008.

[35] T. Nomura, T. Suzuki, T. Kanda, and K. Kato, Measurement of negative attitudes toward robots, *Interaction Studies. Social Behaviour and Communication in Biological and Artificial Systems*, Nov. 2006, vol. 7, no. 3, pp. 437–454.

[36] T. Nomura, T. Kanda, T. Suzuki, and K. Kato, Prediction of human behavior in human - Robot interaction using psychological scales for anxiety and negative attitudes toward robots, *IEEE Transactions on Robotics*, Apr. 2008, vol. 24, no. 2, pp. 442–451.

[37] D. Syrdal, K. Dautenhahn, K. L. Koay, and M. L. Walters, The negative attitudes towards robots scale and reactions to robot behaviour in a live human-robot interaction study, *Adaptive and emergent behaviour and complex systems - Proceedings of the 23rd Convention of the Society for the Study of Artificial Intelligence and Simulation of Behaviour, AISB 2009*, 2009.

[38] T. Nomura, T. Suzuki, T. Kanda, and K. Kato, Measurement of anxiety toward robots, *Proceedings - IEEE International Workshop on Robot and Human Interactive Communication*, pp. 372–377, 2006.

[39] X. Wang, R. Zhou, and R. Zhang, The impact of expectation and disconfirmation on user experience and behavior intention, *Lecture Notes in Computer Science*, 2020, vol. 12200, pp. 464–475.

[40] Y. Koumpouros, A. Karavasili, E. Efthimiou, S.-E. Fotinea, T. Goulas, and A. Vacalopoulou, "User Evaluation of the MOBOT Rollator Type Robotic Mobility Assistive Device," *Technologies*, Nov. 2017, vol. 5, no. 4, p. 73.

[41] L. Demers, R. Weiss-Lambrou, and B. Ska, The Quebec User Evaluation of Satisfaction with Assistive Technology (QUEST 2.0): An overview and recent progress, *Technology and Disability*, Jan. 2002, vol. 14, no. 3, pp. 101–105.

[42] M. J. Scherer and G. Craddock, Matching Person & Technology (MPT) assessment process, *Technology and Disability*, Jan. 2002, vol. 14, no. 3, pp. 125–131.

[43] Y. Koumpouros, E. Papageorgiou, A. Karavasili, and F. Koureta, PYTHEIA: A scale for assessing rehabilitation and assistive robotics, *International Journal of Medical, Health, Biomedical, Bioengineering and Pharmaceutical Engineering*, 2016, vol. 10, no. 11, pp. 456–460.

[44] R. A. C. M. Olde Keizer *et al.*, Using socially assistive robots for monitoring and preventing frailty among older adults: a study on usability and user experience challenges, *Health and Technology*, Aug. 2019, vol. 9, no. 4, pp. 595–605.

[45] A. H. Segars and V. Grover, Re-examining perceived ease of use and usefulness: A confirmatory factor analysis, *MIS Quarterly: Management Information Systems*, 1993, vol. 17, no. 4, pp. 517–525.

[46] H. van der Heijden, User acceptance of hedonic information systems, *MIS Quarterly: Management Information Systems*, 2004, vol. 28, no. 4, pp. 695–704.

[47] L. van Velsen, T. van der Geest, L. van de Wijngaert, S. van den Berg, and M. Steehouder, Personalization has a Price, Controllability is the Currency: Predictors for the Intention to use Personalized eGovernment Websites, *Journal of Organizational Computing and Electronic Commerce*, Jan. 2015, vol. 25, no. 1, pp. 76–97.

[48] C. Fattal *et al.*, Perspectives on usability and accessibility of an autonomous humanoid robot living with elderly people, *Disability and Rehabilitation: Assistive Technology*, 2020, pp. 1–13.

[49] S. Frennert, H. Eftring, and B. Oštlund, Older People's Involvement in the Development of a Social Assistive Robot, *Lecture Notes in Computer Science*, vol. 8239, 2013, pp. 8–18.

[50] S. Frennert, H. Eftring, and B. Östlund, What Older People Expect of Robots: A Mixed Methods Approach, *Lecture Notes in Computer Science*, vol. 8239, 2013, pp. 19–29.

[51] I. Ocnarescu and I. Cossin, Rethinking the Why of Socially Assistive Robotics Through Design, *Lecture Notes in Computer Science*, vol. 10652, 2017, pp. 383–393.

Appendix

Comparison table of the main research results.

		FRAMEWORK		SCALE	FIELD																			
Reference	Year	general	evaluation		Industry	social robot	Robot's features	Robot's behaviour *	Robot's aspect	Human factors	Social impact	Self-efficacy *	Attitudes toward robots	Perception about the robot	Acceptance *	Emotions / Feelings	Perception about the performance *	Physical interaction *	Safety *	Control *	Comfort *	Trust *	Task factors *	Environment factors *
[18]	2004	HRI classification																						
[19]	2004	HRI classification																						
[20]	2006		X				X (e.g. physical parameters)					X (e.g. training, motivation, stress)											X (e.g. complexity, repetitiveness)	X (e.g. illumination, visibility)
[22]	2009		X USUS		X						- Quality of life - Working conditions - Education - Cultural context	x			**Social acceptance:** - performance expectancy - effort expectancy - attitude toword using technology - self efficacy - forms of grouping - attachment - reciprocity	x	**Usability:** - effectiveness - efficiency - learnability - flexibility - robustness - utility **User experience** - embodiment - emotion - human-oriented perception - co-experience		X					X
[23]	2013	X				X		x							**Acceptance:** - usefulness - adaptability - sociability - companionship - perceived behavioral control	x	**Usability**			x				
[3]	2009			Godspeed		X		**Animacy:** - dead-alive - stagnant-lively - mechanical-organic - artificial-lifelike - inert-interactive - apathetic-responsive **Perceived intelligence:** - incompetent-competent - ignorant-knowledgeble - irresponsible-responsible - unintellingent-intelligent - foolish-sensible"	**Anthropomorphism:** - fake-natural - machinelike-humanlike - unconscious-conscious - artificial-lifelike - moving rigidly-moving elegantly					**Likeability:** - dislike-like - unfriendly-friendly - unkind-kind - unpleasant-pleasant - awful-nice					**Perceived safety** - anzious-relaxed - agitate-calm - quiescent-surprised					
[16]	2017			ROSAS		X								**Social perception:** - warmth - competence - discomfort										
[2]	2020			HRIES		X		**Animacy:** - alive - natural - real - human-like **Agency:** - self-reliant - rational - intentional - intelligent						**Sociability:** - Warm - Likeable - Trustworthy - Friendly **Disturbance:** - creepy - scary - uncanny - weird										
[33]	2017			X		X							X					X						
[35]	2006			NARS		X							X											
[38]				RAS		X										**Anxiety:** - Anxiety toward communication capacity of robots - Anxiety toward behavioral characteristics of robots - Anxiety toward discourse with robots								
[14]	2018			SE-HRI		X						X												
[44]	2019			X		assistive										**Enjoyment**	**Perceived usefulness**			X				
[41]	2002			QUEST 2.0		assistive											**Satisfaction**	X						
[43]	2016			PYTHEIA		assistive											**Satisfaction**							
[48]	2020			X		assistive	**Robot's funtionalities**		**Appearance** **Anthropomorphism:** - apparence - familiarity		x			**Friendly relationship and complicity** **Social relationship** **Perception and affects**		**Affection**			X	**Privacy, control, Initiative**				
[29]	2011			X		X	x	x		x		x										**Robot characteristics** - Performance-Based - Attribute-Based) **Human characteristics** - Ability-Based - Personality-Based) **Environment characteristics** - Team Collaboration - Tasking	x	x
[30]	2019			X		X							**Intention to use**			**Enjoyment**					X	X		
[21]	2016			X	X																	X		