Extending the Evaluation of Social Assistive Robots with Accessibility Indicators. The AUSUS Evaluation Framework

Ana Iglesias, Javier García, Ángel García-Olaya, Raquel Fuentetaja, Fernando Fernández, Adrián Romero-Garcés, Rebeca Marfil, Antonio Bandera, Karine Lan Hing Ting, Dimitri Voilmy, Álvaro Dueñas and Cristina Suárez-Mejías

***Abstract*—The introduction of robots in the real world requires previous evaluation of the envisaged performance. Factors like usability, user experience, social acceptance or societal impact among others have been taken into account in evaluation frameworks defined during the last years. However, one of the most important factors that needs to be evaluated in any kind of interaction is whether all the users are able to work with the system with the same opportunities and easiness, and none of the current Human-Robot Interaction (HRI) evaluation frameworks include this factor yet. This paper proposes an extension of a popular HRI evaluation framework, including accessibility as a new evaluation factor. The proposed approach, named AUSUS, considers Accessibility, Usability, Social acceptance, User experience and Societal impact. This paper presents a use case of the framework, which is evaluated through a socially assistive robotic platform created to perform Comprehensive Geriatric Assessment: the CLARC - system. The details of the evaluation process in a hospital and a retirement home are reported and the main difficulties and recommendations of using AUSUS are discussed.**

I. Introduction

During the last two decades, the high rise of technology innovation in Information and Communication Technologies (ICTs) and Robotics, has resulted in products that have the potential to play an important role in assisting people with different necessities: older adults, people with temporal or permanent disability, etc. [1]. According to the 2002 United Nations (UN) survey [42], robotics can be grouped into three categories: industrial robotics, professional service robotics and personal service robotics. Assistive Robots are those designed to provide personal service or assistance to people. This paper is focused on Socially Assistive Robots (SAR), defined as robots whose main goal is to provide assistance to human users through social interaction [43].

Nowadays, SAR are being integrated in different healthcare services, having an impact on health and well-being of older adults [56].

A. Iglesias, J. García, A. García-Olaya, R. Fuentetaja and F. Fernández were with the Computer Science and Engineering Department, Universidad Carlos III de Madrid (Spain), e-mail: aiglesia@inf.uc3m.es

A. Romero-Garcés, R. Marfil and Antonio Bandera were with the Department of Electronic Technology, Universidad de Málaga, Spain

K. Lan and D. Voilmy were with University of Technology of Troyes, France

A. Dueñas and C. Suárez were with Hospital Universitario Virgen del Rocío, Seville, Spain

Manuscript received April 29, 2020; revised XXXX.

The integration of SAR into human healthcare environments is not a self-evident, and reveals to be a complex task. Users in human-robot interaction can face many problems and difficulties related to different factors like usability, social acceptance, user experience or societal impact [31]. All these factors have been studied in Human-Robot Interaction research field (HRI) as detailed in section II-C. One of the factors that most influences the interaction between humans and robots is accessibility. For instance, older people generally have limitations in interacting with software/robots or other technologies, mainly due to the alterations that accompany aging (vision, dexterity, hearing and cognitive impairment) and the ability of people to use the new technology. However, this essential factor has not been studied in depth yet in the HRI research field nor it is included in any of the existing evaluation frameworks, as we analyze in Section II.

For instance, USUS, a popular HRI evaluation framework, considers Usability, Social acceptance, User experience, and Societal impact factors [30]. But some people as older adults could face accessibility barriers, which directly impede the interaction between the human and the robot, even with systems that have received a positive HRI USUS evaluation.

The main contributions of this paper are firstly, the identification of accessibility as a necessary dimension to take into account in HRI evaluation (Section II-B). Secondly, the extension of USUS, one of the most complete evaluation frameworks for HRI, with the accessibility factor to create AUSUS (Accessibility, Usability, Social acceptance, User experience and Societal impact). AUSUS leverages tools and guidelines from Human-Computer Interaction (HCI) research field and provides information about the accessibility indicators, as well as recommendations of measures and methods to assess the accessibility of socially assistive robots (Section III). Finally, AUSUS is used in a case study to evaluate CLARC, a socially assistive robot to perform Comprehensive Geriatric Assessment (CGA) in retirement homes and hospitals. The paper details the evaluation setup, the recruitment process followed, the materials, the mechanisms used to evaluate the performance indicators, and finally reflexively discusses the main difficulties faced during the evaluation. The objective is to detail as much as possible the evaluation framework and the insights gained - both pragmatic and methodological - while presenting the specificities of our use case, in view of a possible generalization.

II. Literature Review

This section presents relevant previous work concerning three topics which are important for our research object. First, Socially Assistive Robots (SARs) are presented in resonance with a definition of ”robots”. Knowing the main characteristics of SARs allows a contextualized understanding of the most used HRI evaluation frameworks, which appear as essential, and is presented second. Third, this literature review presents accessibility as an important factor that influences the success of robotic systems before they are introduced in the real world.

A. Robots and Socially Assistive Robots

One of the difficulties in reaching a clear definition of what a *robot* is, is the wide range of types and uses – from industrial to service robots – and the fact that research has initially focused on the former. The European consortium Robolaw, in its Guidelines on Regulating Robotics [44], proposes, a taxonomy of robots. Aiming at inclusion rather than exclusion, and based on robots' main characteristics, takes into account the plurality of uses and applications: (i) Use or task, (ii) Environment, (iii) Nature, (iv) Human-robot interaction, and (v) Autonomy.

Social robots are characterized by understanding and communicating in a human-like way, allowing them to behave as social actors and be understood as such by their users [46]. Feil-Seifer and Mataric [47] propose a taxonomy of social robots, and identify 3 main categories: (i) Assistive Robotics (AR), which gives aid or support to a human user (rehabilitation, wheelchair and other mobility aides, companion, manipulator arms for the physically disabled, or educational robots). (ii) Socially Interactive Robotics (SIR), whose main goal is to develop close and effective interactions with the human for the sake of interaction itself. (iii) Socially Assistive Robotics (SAR), which also aims to create close and effective interaction with humans but in this case with the goal of giving assistance and achieving measurable progress in convalescence, rehabilitation, learning, etc.

This study is focused on Socially Assistive Robotics (SAR), because their performance does not only rely on the classical aspects of robot behaviour (mobility or speech capabilities) but also on social abilities, including more complex scenarios to be evaluated. Also it is more likely that people using them will face accessibility barriers during interaction.

B. Accessibility

Today, the importance of developing accessible and usable software is well known in the HCI field [4]. The concepts of software accessibility and usability are related, but they are not the same [22]. Accessibility addresses discriminatory aspects related to the equivalent user experience for all people, regardless of their cognitive and functional abilities and/or disabilities. The software usability addresses the aspects of ease of use and learning. In general, accessibility requirements improve usability for all.

Accessible software is, therefore, a software that all people (people with disabilities, older adults, people in special environments, etc.) can perceive, understand, navigate and interact with. Accessibility and design for all is currently being studied in depth in HCI. However, in HRI, just some research project are focused on specific disabilities. For example, [59] is focused in visually impaired needs when interacting with robots; [25] details a case study for accessible interface for telepresence robots; or [60] is focused on usability and accessibility issues for an autonomous humanoid robot living with elderly people. However, none of this studies introduce a methodological framework to evaluate the interfaces accessibility.

C. Evaluation frameworks for Human-Robot Interaction

Attending to the characteristics of SARs presented above, evaluating HRI is a key issue. This necessity of defining metrics to measure the success of robotic systems in the HRI research field made theoretical and methodological evaluation frameworks rise. The theoretical taxonomy of Yanco et al. [32] addressed multiple research areas like HCI or Computer Supported Cooperative Work (CSCW) and social sciences to offer a holistic picture of research aspects, setting therefore the first steps in generalizing the HRI research. The evaluation framework of Erfanian et al. [2] provides metrics to measure user satisfaction in collaborative Virtual Environments. The first attempt to define metrics was made by Steinfeld et al. [23], based on an HCI perspective, to measure navigation, perception, management, manipulation and social acceptance for mobile robots. Sung et al. [41], define the Domestic Robot Ecology (DRE) framework, evaluating the long-term acceptance of robotic technologies at home. Heerink et al. [40] assess the acceptance of assistive social agent technology by older adults, describing the Almere model, adapting and extending theoretically the Unified Theory of Acceptance and Use of Technology (UTAUT) model [26], which investigates which factors influence the intention to use the new technology.

The USUS methodological framework [31] consists of a mix of methods derived from various research fields: HRI, HCI, Psychology and Sociology. It addresses usability, social acceptance, user experience and societal impact of humanoid robots used in collaborative tasks, giving a holistic view of robot evaluation.

However, none of the existing frameworks includes accessibility as an evaluation factor in HRI. The key point of this paper is that accessibility emerges as an indispensable factor when people with special needs interact with robots. The use of complete evaluation frameworks of HRI could help the designers and developers to improve the system, guaranteeing the success of the robot's development and integration in the society. This is the reason why the paper proposes to enrich the USUS HRI evaluation framework by including accessibility as a complementary factor.

III. Including Accessibility in an HRI Evaluation Framework: AUSUS

The USUS methodological framework [31] takes into account most of the difficulties and problems that users face when interacting with a robot, so it has been chosen as basis of our proposal. The evaluation factors included in USUS are:

- *Usability*: effectiveness, efficiency, learnability, flexibility, robustness and utility.
- *Social Acceptance*: performance expectancy, effort expectancy, attitude towards using technology, self efficacy, forms of grouping, attachment and reciprocity.
- *User Experience*: embodiment, emotion, human-oriented perception, feeling of security and co-experience.
- *Societal Impact*: quality of life, working condition and employment, education and cultural context.

The framework proposes different methods to assess these evaluation factors, such as expert evaluation, user studies, questionnaires, physiological measures, focus groups and interviews [31]. The framework has been successfully used and enhanced with User eXperience (UX) goals [45]. However, *Accessibility* is not part of this methodological framework and therefore, as explained above, our proposal includes this factor into the USUS evaluation framework to create AUSUS. The remaining of this section describes *Accessibility* as an HRI evaluation factor and specifies different methods to assess it.

A. Accessibility as Evaluation Factor

The term *accessibility* is defined in ISO 26800, ISO/TR 9241-100 and ISO/TR 224111 as the "extent to which products, systems, services, environments and facilities can be used by people from a population with the widest range of characteristics and capabilities to achieve a specified goal in a specified context of use". Using this term from an inclusive perspective in HRI, accessibility would benefit everyone interacting with robots, avoiding the presence of accessibility barriers and recognizing that no two people have exactly the same abilities/disabilities and interaction characteristics, which are influenced by their gender, age, health condition, body size, temporal or permanent impairments, training and experience.

As aforementioned, the position adopted in this paper is based on Petrie's research work, who claims that software accessibility and usability are related, but they are not the same [22]. On the one hand *accessibility* addresses discriminatory aspects related to the equivalent user experience for all people, regardless of their cognitive and functional abilities and/or disabilities. On the other hand, software *usability* addresses the aspects of ease of use, related to effectiveness, efficiency and satisfaction, according to its definition in ISO 9241-11:1998 standard. The indicator for the *accessibility* factor is based on the detection of accessibility barriers during the interaction, taking into account the requirements included in standards and guidelines related to hardware, software and safety of products and services. Therefore, the new factor of USUS (Accessibility factor) can be defined as:

> Detection of usage barriers that users could face when interacting with robots because they are not properly designed. This factor is related to hearing, visual, motor or cognitive disabilities, affecting the use of new technologies.

In order to evaluate the accessibility factor we propose the HCI indicators specified by the W3C Consortium at the WCAG guidelines [14], adapted to HRI evaluations by Qbilat [58]:

- Perception: the information presented by the robot must be perceivable by all the users and must be adapted to their needs, providing: multiple interaction modalities (alternatives to non-textual content, location of interaction devices, etc.); synchronized media equivalents for time-dependent content (captions for the robot's voice and any other relevant sound, audiodescription, sign language, etc.); adaptable content (meaningful sequence of contents, assurance of the user's orientation during the interaction, identification of the purpose of any interface, etc.); distinguishable content (adequate color contrast; proper use of color, audio and video control, content resizing)
- Operation: the interface components must be operable, by: providing an accessible control mechanism or access to assistive technology and enough time to interact (adjustable timing, pause/stop/hide components, etc.); avoiding negative physical reactions, like seizures produced by blinking and flashing contents, etc.; ensuring navigable contents (focus order, multiple navigable ways, use of proper headings, labels and titles, clear link purposes, etc.); and allowing multiple input modalities (pointer gestures and cancellation, target size, etc.)
- Understanding: the information must be understandable by providing: readable and audible content (known language, avoiding unusual words and phrases and abbreviations, proper reading level and pronunciation, etc.); predictable content (consistent navigation, components focus and identification, etc.); and input assistance (for error identification, prevention and suggestion, including labels/instructions and help, etc.)
- Robustness: the interface must be robust, by providing compatibility with current and future technology and particularly assistive technologies. Moreover, the interface should provide the option to communicate with a human to help the users if they need to.

Figure 1 shows the AUSUS framework evaluation factors and the methods proposed to assess them.

B. Methods to Assess Accessibility

In line with the core principles of the broadly termed user-centered design, i.e human-centred design [48], participatory design [49] and iterative design, users have to be involved in the accessibility evaluation of the robot's interface during the whole product's development, from the needs analysis to the pilot evaluation of a fully functional prototype [50].

Due to the very nature of accessibility, characterized by both requirements included in standards and guidelines, and barriers linked to specific individual impairments, the classical (quantitative) metrics of evaluation frameworks for HRI appeared insufficient to address the complexity of the issue. Instead, as explained in the Evaluation set-up (Section V below), we used what we would term "measures", which include both metrics and qualitative indicators. Indeed, if expertly conducted, as we believe is the case in this study, qualitative descriptions and judgments can be as valuable as, or more valuable than

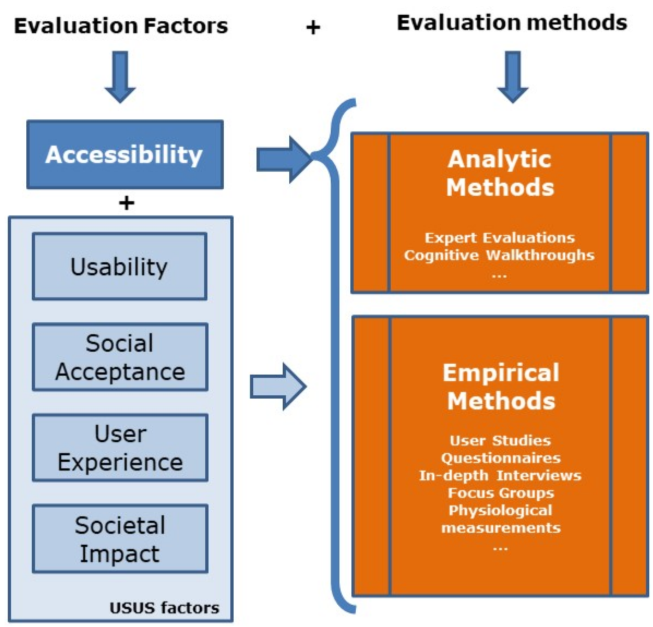

Fig. 1. AUSUS framework evaluation factors and methods.

measurements. In complex multifactorial circumstances, like evaluating accessibility, metrics can be misleading. Therefore, relevantly adopting this mixed methodological approach is central for assessing accessibility. The authors are convinced of the complementarity between qualitative and quantitative approaches, allowing a more meaningful, in-depth and contextualized understanding of the quantitative results [57]. The objective of using this mixed methodological framework is twofold: on the one hand, triangulation, which ensures the validity of quantitative analyses [51], and on the other hand, the complementarity of quantitative and qualitative methods, the latter allowing for a more refined and contextualized understanding of quantitative results [52]. Indeed, the underlying logic of the combination is that neither quantitative nor qualitative methods alone are sufficient to capture trends and details. Indeed, mixed methods studies in the health domain can access knowledge that would not emerge from a qualitative study, or from a quantitative study that would be conducted independently of each other [53]. This argument about complementarity will be demonstrated in the evaluation set-up and analyses (section V).

Of course, adopting this mixed methodological approach has consequences on the usual practices of HCI evaluations, including the number of participants. The choice of the number of participants for these preliminary user tests is both methodological and practical. Indeed, studies on usability tests [54] show no significant correlation between the number of users and the number of severe problems identified. Nielsen [55] recommends having tests with ideally 3 users, and a maximum of 5: "Elaborate usability tests are a waste of resources. The best results come from testing no more than 5 users and running as many small tests as you can afford ", he says. Therefore favoring in-depth iterative testing and following the guidelines for qualitative evaluations, few users could sufficient to form a representative sample of the population, in terms of IT use and functional and cognitive abilities.

Different methods can be used to assess accessibility in HRI, based on the HCI research field. Focusing on Web accessibility, the W3C/WAI model of accessibility assumes that to get *Universal Accessibility* in Web pages/sites, the key precondition is the conformance to WCAG (Web Content Accessibility Guideline) [14]; the tools used by the web developers must conform to ATAG (Authoring Tools Accessibility Guidelines) [12]; and the browsers and assistive technology used by the end users must be compliant with UAAG (User Agent Accessibility Guidelines) [13].

The World Wide Web Consortium (W3C) proposes a methodology to perform accessibility conformance evaluations of web pages/sites; the WCAG-EM [15], carried out by accessibility experts. The methodology involves several steps: defining the evaluation scope, exploring the target website, selecting a representative sample, auditing the selected sample and reporting the findings. The activities within the methodology steps are influenced by aspects like the main objective of the evaluation, the type of website, the website size, its complexity, or the technology used to create the website. Petrie and Bevan also propose a range of accessibility, usability and UX evaluation methods for interactive electronic systems (eSystems) [3].

Based on the methodologies, methods and techniques currently applied in HCI, a combination of analytic and user evaluation methods is recommended for AUSUS framework, to ensure that accessibility is provided considering a wide range of disabilities and situations:

- *Analytic methods* are based on inspections of the system carried out by experienced accessibility experts.
- *Empirical methods* are based on interaction experiences where accessibility experts and potential users (users with functional diversity and older adults) interact with the robot, following an inclusive user-centered design.

1) Analytic methods: The use of these methods helps to identify many accessibility issues that can be fixed before involving people with disabilities in the evaluation phase using empirical methods (which have a greater cost). Moreover, the analytic methods help to focus on special accessibility issues, therefore, decreasing the cost of the evaluation. On the one hand, the *Expert Evaluation* methods include *Heuristic Evaluations*, also known as *Conformance evaluations* [15], where experts analyze how well the robot's interface meets accessibility standards and guidelines, resulting in a list of accessibility problems ranked according to their severity. A recommendation is to use automatic tools for checking the conformance with standards and guidelines as long as it is possible. For instance, if the robot includes displays, its accessibility can be checked with some automatic tools, as the *508checker*[1] tool, which checks web-pages for 508 Rehabilitation Act compliance. Another example of automatic tools are those which check the accessibility of the colours used in the interfaces, as the *Accessible Colour Evaluator (ACE)*[2], which helps web developers and designers to balance aesthetic

[1]http://www.508checker.com [last accessed August 2021

[2]http://daprlab.com/ace/ [last accessed August 2021

and accessibility constraints when choosing a website's colour scheme. The automated checking and expert evaluations can be performed when initial prototypes or full implementations are available. This will ensure that the software meets the accessibility guidelines and standards. Expert evaluations are needed; none of the existing automatic tools detect all the accessibility barriers that users could face.

The analytic evaluation must take into account relevant laws and regulations, at national or international levels, protecting the rights of people with disabilities or trying to avoid age discrimination. In addition, there are several standards addressing the basic accessibility conditions of the products, environments or services, which experts should take into account. For example, the European UNE 170001:2007 standard, which includes the criteria to facilitate Universal Accessibility Management of the places in which one can enjoy goods or services; the UNE-EN 301549:2015 standard that regulates the accessibility requirements suitable for public procurement of ICT products and services in Europe; the ISO 26800:2011, which presents the general ergonomics approach, its principles and concepts; ISO/TR 9241-100:2010, which identifies ergonomics standards particularly relevant to software development; the ISO/TR 22411:2008, which is focused on ergonomics data and guidelines for older persons and persons with disabilities; or the ISO/IEC Guide 71:2014, which provides guidance to standards' developers on addressing accessibility requirements and recommendations in standards that focus on systems used by people, among others.

In HCI, there are numerous recommendations, guidelines and standards that summarize the main problems that people with disabilities can face when interacting with software products, such as those provided by the W3C Consortium [14], accessibility research groups such as Funka Nu or companies such as IBM [21]. Some of them are becoming international standards in which, in some cases, the laws of different countries rely to preserve the rights of people with disabilities, such as ISO/IEC 40500:2012 and UNE 139803:12, which include accessibility guidelines for web content, based on the previously introduced W3C guidelines. These recommendations and standards serve as a guide for designers and developers of systems to build accessible systems for all. However, there is a lack of specific guidelines for HRI. Indeed, after a thorough review of the literature, only two guides related to accessibility in HRI have been found, one for robot's displays [58] and the other for hardware, software and safety issues [36]. According to robot interface's characteristics, the selection of the proper guidelines and standards is necessary to carry out the heuristic evaluation. As aforementioned, AUSUS relies on the HRI guidelines of Qbilat [58] to focus on the specific evaluation of human-robot interfaces.

On the other hand, accessibility experts can also use methods as *cognitive walkthrough*, imagining how users would solve some tasks [10], and employ models or technology to simulate the functional diversity of the potential users. For instance, the use of screen-reader technology can simulate the experience of a person with visual disabilities.

2) Empirical methods: User studies involve persons with functional diversity in the accessibility evaluation of the robot's interfaces, first in a laboratory-based context, when potential disturbances to the user interaction are reduced to the minimum, and later in field-based context [31]. The user studies are particularly important in accessibility evaluations because they help to identify accessibility issues that are not discovered by experts alone: people with disabilities use products differently.

Disabilities are sometimes grouped into four high-level categories: visual, hearing, physical and cognitive, but there is a vast variability within each category. For instance, older adults whose sight is deteriorating but who can interact without assistive technology, persons with low vision, who can use screen magnification software, or blind persons, who need screen readers or similar software to interact with the robot. Due to this variability within a category, is it important to identify a realistic range of participants, taking into account their abilities/disabilities and their interaction mode. Petrie and Bevan summarize the main issues to take into account when designing user evaluations in an inclusive way [3].

Finally, both objective, like interaction issues (time to complete tasks, systems' errors, accessibility barriers, etc.), and subjective, like satisfaction surveys, data are recommended to be collected.

Analytical methods should be used when user testing is not practical or viable, for instance, when the time to implement the software is critical and/or for economic or safety reasons, evaluations using models and simulations could be performed. Using this kind of evaluations, we could predict accessibility barriers as the time or cognitive difficulty to complete a task. On the other hand, evaluations with users are always recommendable (at all the stages of development, if possible) and at least when the full system is developed. This kind of evaluations provides evidence of the system accessibility and the real use by the target audience.

IV. Use Case of AUSUS: CGA and the CLARC robot

CLARC [27] is a robot aimed to perform Comprehensive Geriatric Assessment (CGA) in hospitals and retirement homes. CGA is a medical evaluation process to assess the capabilities of older adults [16] by means of a series of tests of different nature. Performing such tests requires to endow the robot with many capabilities, from classical low level skills, like reactive obstacle avoidance, to high level ones like speech recognition or user adaptation. Figure 2 shows the external aspect of the CLARC robot and a patient interacting with it at a retirement home during the evaluation process.

The robot is equipped with a Microsoft Kinect V2 sensor, a shotgun microphone, a touch screen, speakers, and a web-cam for recording the sessions. The robot is also fully connected to an external device: a tablet equipped with large buttons to facilitate the interaction of older adults. The voice of CLARC is generated from text using the Text-To-Speech (TTS) software provided by the Microsoft Speech Platform SDK. This software is also used for voice recognition, with the help of specific grammars that are loaded for each question, in order to maximize recognition rates.

Several of the software modules within the architecture are in charge of providing a multi-modal interface for HRI. Users

Fig. 2. Patient performing a test interacting with CLARC robot in the retirement home.

can interact with the robot using voice, the touchscreen or the external device, which allows the patient to hand-write or draw, as required in some tests.

For autonomous behavior the robot relies on the CORTEX architecture, which maintains a central inner representation for internalizing the outer world. This representation is collaboratively constructed by all the components on the architecture, whether they are reactive, like those detecting and tracking people in the scene or recognizing the speech, or deliberative, like the one that uses Automated Planning to determine a course of actions or the one that takes care of path planning.

V. Evaluation Set-up

It is important to note that the goal of this evaluation is not to assess the CLARC platform itself. Instead, from a reflexive and methodological perspective, the objective is to check whether the AUSUS evaluation framework is accurate for its purpose, and whether the user test protocol is adapted to older adults and is efficient in gathering the necessary data (especially qualitative). With this mixed methods framework, the aim of the qualitative approach is to take into consideration the specificities of the context of use. Indeed, the objective of the proposed AUSUS framework is to achieve a holistic and situated evaluation of HRI of socially assistive robots, considering the needs of specific users, older adults, in a given task, CGA, in a concrete organisational context, the hospital or retirement home. Through evaluating the viability of our method, and the conclusions obtained, we expect to make new contributions to the evaluation process of HRI of CLARC, and eventually to the evaluation methodology of other interactive technologies designed to fit older adults specific needs. Generalization should be possible, considering at the same time the situated and specific characteristics of this case study.

A. Evaluation Environment

The evaluation was carried out in Seville, Spain, in two different scenarios: a retirement home and at the *Virgen del Rocío University Hospital*. In the retirement home, the focus of the evaluation was to determine how the CLARC robot performs a CGA test with older adults. The CGA test chosen for the evaluation was the Index test (BI) [6][19] (Barthel test). It is a clinical application instrument to evaluate older adults' daily living activities through 10 multiple choice questions. Barthel test questions have between two and four options of different complexity each. In the hospital, different types of clinicians could interact with CLARC robot and observe how other clinicians or patients completed the Barthel test.

One of the authors, a clinician specialized in geriatric evaluation at the *Virgen del Rocío University Hospital*, prepared a case study as close as possible to a real procedure of a Barthel test in his daily practice. Characteristics and peculiarities of his own behavior during the medical consultation had been translated and implemented into interaction design functions in the robot. The robot was able to: introduce the test to the user, ask the clinical test questions (by voice and showing the question and possible answers in a display), and receive the answer from the user (by voice, touchscreen or remote control).

For the evaluation in the retirement home, two alongside rooms were used; one for the interviews with the patients, and the other to perform the Barthel test with the robot. The rooms were spacious and familiar for all the patients. Figure 2 shows a photo of a patient interacting with CLARC robot while performing the Barthel test at the retirement home.

For the project presentation and evaluation at the Hospital, a large conference room was used. Patients and voluntary clinicians could perform the Barthel test with the robot, while their colleagues could watch the procedure.

B. Recruitment of Participants

Two categories of end-users were involved in the study: patients and clinicians. Patients were volunteers, over 55 years old and with a Minimental State Examination (MMSE) score greater than 20 [17]. Clinicians were volunteers too, with a professional category of doctors or physiotherapists who, in their daily clinical practice, work with geriatric patients.

Eight patients were involved in the evaluation. They had different age-related impairments: three of them with light hearing impairments and one with medium hearing impairments; three of them with light visual impairments and one with medium visual impairments; and three of them with light cognitive impairments. They formed a representative sample of the retirement home's population, in terms of IT use and functional and cognitive abilities.

Table I describes the main characteristics of the patients who participated in the evaluation. The mean age was 81.37 ± 12.07 years. There were seven women (87.5%) and one man (12.5%).

Seventeen clinicians participated in the evaluation, of whom six were primary care doctors (35.3%), four were rehabilitation specialists (23.5%) and seven were physiotherapists (41.2%). The clinicians presented a mean age of 32.35 ± 7.13 years. Out of the 17 clinicians, twelve were women (70.6%) and five were men (29.4%).

TABLE I
USER'S CHARACTERISTICS AT THE RETIREMENT HOME EVALUATION

User ID	Gender	Age	Motor Impairments	Hearing Impair.	Visual Impair.	Cognitive Impair.	New technologies skills
user1	Woman	86	In a wheelchair from a recent hip fracture	no	no	no	Cell phone and tablet (phone calls and photos).
user2	Woman	75	Uses a walker to move around due to senile impairment	light	no	no	Cell phone (phone calls and photos) and computer (Internet navigation)
user3	Woman	84	Uses a walker to move around due to senile impairment	light	medium	light	Cell phone (phone calls and photos)
user4	Woman	55	In a wheelchair due to lower limb impairments	no	light	light	Cell phone and tablet. Advanced user (continuously).
user5	Man	93	Light impairments related to age	medium	light	no	Cell phone (phone calls and photos)
user6	Woman	84	Uses a walking stick to move around due to senile impairment	no	light	light	Cell phone (phone calls and photos) and computer (Internet navigation)
user7	Woman	82	Light walking impairments but without technical assistance	no	no	no	Cell phone (phone calls and photos) and computer (Internet navigation)
user8	Woman	92	Uses a walking stick to move around due to senile impairment	light	no	no	Cell phone (phone calls and photos)

C. Materials

Quantitative and qualitative methods were complementary combined to answer the research questions, examining both the patients' and the clinicians' perspectives:

- *Heuristic Evaluation*: accessibility evaluation of the robot interface made by experts according to accessibility guidelines, recommendations and standards. Automatic tools were also used.
- *Test and structured interviews*: user tests and interviews were performed with the patients in the retirement home. Simple tests and a structured interview to the clinicians who participated in the observations collected clinicians' point of view and requirements. Interviews were used before and after interacting with the robot. Before interacting, sociodemographic variables and technology use and skills (cell phone and computer skills among others) were surveyed. After the interaction with the robot, questionnaires and structured interviews were conducted to measure subjective usability criteria and future intention to use the robot. The test, questionnaire and structured interviews contained questions on a Five-point Likert scale [5] (from 1 = do not agree to 5 = fully agree), examining users' perception of usability, social acceptance, user experience, impact of the CLARC robot (see details on the description of the results, Section VI), as well as questions related to accessibility barriers found during the interaction.
- *Objective data*: During the user tests, the robot recorded objective accessibility and usability criteria about the interaction, in terms of success in achieving the planned tasks: percentages, mean average time per test / question, standard deviation, total number of answered questions, etc. Also, as a useful assistant performing the evaluation autonomously, the robot also saved the score for each answered question. These quantitative data were complemented with knowledge about how the tests were actually achieved. E.g: *Log* analysis revealed that *user2* failed in answering questions 1 and 3.
- *Observations*: observation of the interaction with the robot during the test (*in situ* or videotaped) allowed an identification of the exact difficulty and accessibility barrier found: interaction with the interface, hearing problems, not knowing what option to answer, Automatic Speech Recognition (ASR) or touchscreen not considering the answer given by the user, etc.

D. Defining the Evaluation Procedure

As stated, eight older adults interacted with the CLARC robot to perform the Barthel test one after each other, and were interviewed by the researchers before and after the interaction. Each test with the robot lasted (in total) about 30 minutes.

The pre-test interview focused on three aspects: 1) the patient was explained the study protocol in detail as well as his/her rights, and signed the informed consent document; 2) sociodemographic variables were collected: users' age, IT tools use and proficiency; 3) user's opinion was sought about his/her perception of robots. The objective was to acquire knowledge about a possible link between familiarity with technological interfaces and a better performance in interacting with the robot; and whether an a priori positive/negative attitude towards robots influenced performance. This pre-test interview took place in the first room.

Then, the patient was taken to the second room, where the robot was introduced. The patient was invited to sit comfortably, and the robot's main features were explained. The patient performed the Barthel test with the robot. As explained above, one of the authors, the physiotherapist having designed the interaction with the robot for the Barthel test, was present in the room and proceeded with his own Barthel Test evaluation, while observing the interaction of patients and robot. He coded his usual grid based on his observation and the patient's answers to the robot. However, with the aim of reflecting as objectively as possible the patient's current functional state, after the sessions, the clinician compared and completed his scores with a previous Barthel Test made a few days before in the retirement home, as part of the medical care facilities of the institution.

Once the test was finished, patients were accompanied back to the first room. The post-test interview and satisfaction questionnaire was built according to the USUS framework and the new accessibility factor (see aspects examined and results below).

On the other hand, clinicians who participated in the evaluation were summoned in a room of the Virgen del Rocío University Hospital in Seville. After reading and signing the informed consent, a clinical session was carried out in which participants could visualize a Barthel test performed by the CLARC robot on a patient. After the Barthel test visualization, an AUSUS framework-based questionnaire specific to clinicians' needs and practices was distributed. All the clinicians filled the questionnaire and submitted it before leaving.

VI. Evaluation of the Performance Indicators

This section summarizes the evaluation process and methods used to asses the performance indicators. The main difficulties faced by the users are also summarized.

A. Evaluating Usability, Social Acceptance, User Experience and Societal Impact

A in-depth analysis of the robot's performance factors according to USUS was carried out: usability, social acceptance, user experience and societal impact. For the usability factor, the following performance indicators were considered:

- Effectiveness: CLARC robot is effective if the robot is able to successfully perform the Barthel test, properly processing patient's answers and providing an adequate evaluation and recommendation to the doctor. Therefore, objective data related to the number of successful questions answered by the users were evaluated to asses this performance indicator. Moreover, one clinician supervised evaluation and recommendation given by the robot after the Barthel test for each user. Finally a clinician completed the Barthel test with the same users a few days after the interaction with the robot, in order to compare the results of CLARC robot and the results obtained by the clinician.
- Efficiency: a comparison between the time spent by the robot to complete a Barthel test and the time spent by a clinician was done. The robot is efficient if the users do not spend more time interacting with the robot than necessary.
- Learnability: to evaluate this indicator, firstly, the users were asked about their skills related to the use of new technologies (smart phones, computers, tablets, robots, etc.) and if they were familiar with this kind of robots and/or the Barthel test before the interaction. Secondly, the interaction sessions were observed by a clinician and an engineer. Finally, after the interaction session and through a structured interview (questions q1–q4 in Table II, the subjective opinion of the patients related to the system's learnability was surveyed in a 5-point Likert scale from 1 (strongly disagree) to 5 (strongly agree). Finally, correlations were studied.
- Flexibility: it evaluates the different ways the patients can use to communicate with the system: voice, a touchscreen tablet and a remote control with physical buttons. Each patient could choose the interaction modality that better corresponds to his/her abilities and capabilities. Firstly, an expert evaluation was carried out to evaluate the multimodality of the system. Secondly, observations were done to check which ones were preferred by users. Finally, question 5 in the structured interview revealed if they found the robot to be flexible.
- Robustness: it evaluates if CLARC is able to correct and prevent novel user's errors. Firstly observations of the interaction sessions were analyzed. Secondly, the number and type of errors in each interaction session was stored and analyzed objectively taking into account the functional diversity of the patients. Finally, the patients were asked about their subjective opinion after the interaction through an open-question (question 6 in Table II).
- Utility: to asses this indicator and taking into account the results of the effectiveness indicator, the patients were asked if they think the robot is useful for CGA assessments (question 7 in Table II).

Due to space limitations the resulting data for the remaining USUS factors are not detailed. To study if CLARC robot is accepted by patients and clinicians (social acceptance), the attitude towards using technology, the self efficacy and reciprocity was analyzed through pre-test and post-test, personal interviews and observations. The *performance expectancy* and *effort expectancy* indicators were not evaluated, because the patients had no idea about the functionality of the robot until it was introduced during the interaction session. In the same way, the indicator related to *forms of grouping* was not evaluated because the interaction should be individual for the Barthel Test in the CGA assessment. Finally, the *attachment* was not assessed because users could not interact with the system enough time to evaluate this indicator.

To study if patients and clinicians have good experiences interacting with CLARC robot (user experience), the embodiment, user's emotions, human-oriented perception and feeling of security were assessed through session observations and structured interview analysis, similar to previous assessments. The only indicator that was not evaluated in this study is the co-experience with robots, because it is related to how individuals develop their personal experience based on social interaction with others.

Finally, the societal impact of CLARC robot was evaluated through *quality of life*, *working conditions* and *employment indicators* in a similar way. *Education* and *cultural context* were not assessed in this study, because these indicators were out of the main aims of the study.

B. Accessibility

This indicator is related to the accessibility barriers found during the interaction process. During the design phase, heuristic evaluations of the prototypes were performed by

TABLE II
PATIENT'S STRUCTURED INTERVIEW DEALING WITH USABILITY

Question ID	Description	Mean	SD
q1	I could hear and understand the robot clearly	4.14	1.21
q2	Learning to operate the robot was easy for me	4.28	1.11
q3	I found the robot easy to use	5.00	0.00
q4	The robot's explanations of what I had to do were clear and understandable	5.00	0.00
q5	The robot is flexible for me to interact with	4.57	1.13
q6	I found errors during the interaction	2.00	1.57
q7	I think the robot is useful for CGA assessments	4.71	0.48

an accessibility expert, taking into account specific recommendations based on accessibility guidelines for HRI [35], [36], [58], guidelines for HCI [14], [21] and standards ISO 9241-171, 2008 and ISO/IEC 13066-1, 2011 dealing with accessible interfaces. Details related to the heuristic evaluation are provided as supplementary documents of the paper.

Automatic tools were used by the experts to help them to evaluate the accessibility heuristics of the robot's display program, such as *Contrast Checker*[3] tool, which checks the compliance with the heuristics related to the color combinations, according the W3C content guidelines [14], or the *Readability Grader1.0*[4] which is a tool that allows people to check whether their content is easy-to-read. Errors found by the accessibility expert were fixed before the beginning of the user evaluation, ensuring an accessible interaction from the experts and guidelines point of view.

Two formal expert evaluations were performed during the design process, where the accessibility indicators were analyzed. During the interaction sessions in the pilot studies, objective data like the number of successful questions answered by the users, or if an accessibility error or incidence occurred, were collected. Moreover, the interaction sessions were observed by a clinician and an accessibility expert, who detected problems dealing with accessibility. A structured interview (questions q1–q4 at Table III confirmed and extended the information obtained during the observations. The subjective opinion of the patients related to the system's accessibility was surveyed in a 5-point Likert scale from 1 (strongly disagree) to 5 (strongly agree). Finally, correlations were studied.

Some accessibility barriers found by the analytic methods during the design process and solved before user studies were:

- Operation: A button to pause the interaction in the remote control was needed. On the display this button was available at every moment.
- Understanding: an echo produced by the robot case made the voice difficult to understand. It was recommended to open new sound exits in certain parts of the robot's case.
- Understanding: the buttons' look was different on the display and the remote control, so the correspondence was not easy for users with cognitive disabilities. It was recommended to use similar colors and shapes in both devices and to highlight the option buttons to be pressed in each moment in the control device (those which are shown at the display at that moment).

[3]http://contrastchecker.com/ [last accessed August 2021

[4]https://jellymetrics.com/readability-grader/ [last accessed August 2021

As a result of the analytical and empirical studies, the final version of the robot's interface was:

- Perceivable: It provided multiple modalities to interact (voice, touch screen, remote control) in order to adapt to user's needs. Moreover, alternative to non-textual content was always provided. Robot voice was captioned, with some exceptions. The robot's interface followed a meaningful sequence of contents, and headings, labels, and other components were used to ensure the orientation of the users during the interaction, identifying the purpose of any interface. The content was distinguishable, by providing a proper use of color and contrast, and resizing it was also allowed.
- Operable: It provided multimodality to operate with the robot and the robot allowed time enough to answer the questions (providing two opportunities to answer before skipping each question/task). Moreover, blinking and flashing contents were not present in the interface. Furthermore, the touch screen interface provided navigable contents: proper focus order, multiple navigable ways, proper headings, labels and titles, etc. and multiple input modalities were provided.
- Understandable: The interface used readable text and audible content using common language and avoiding unusual words, complex sentences or abbreviations. Moreover, the content was predictable and input assistance was provided (highlighting potential physical buttons to press in the remote control, disabling buttons and content which is not available for specific tasks, etc.). Moreover, users were able to ask the robot to repeat phrases and instructions.Error prevention and suggestions were always taken into account.
- Robust: the interface presented a button to communicate with a clinician at any time and no interaction errors were found.

However, some limitations, which must be addressed before deploying the robot in daily routine, were found:

- Perception (adaptation): The location of some hardware components is not configurable and cannot be adapted to the users' needs. Just the remote control can be located wherever the user wants.
- Perception (synchronized media equivalents): Some captions were missing. Moreover, sign language or audio-description were not provided.
- Perception (distinguishable): there was no way for the user to pause/stop the robot's speech when a new task started.

- Operation (assistive technology): the robot does not provide any way to plug other input devices or assistive technology.
- Operation (time): the timing was not adjustable and the users were not able to pause/stop the task.
- Operation (input modalities): other input modalities should be considered if the robot would allow connection of other devices.
- Understanding (pronunciation): The robot voice was not very natural (robotic voice) and the pronunciation was not the best for some specific sentences. The use of a better and more human-like text-to-speech model is recommended.
- Robustness: the interface was implemented in C# language, which is a standard language, but new implementation and software updates should be done to be compatible with future and assistive technology.

The insights gained from both the expert analysis and the observational analysis confirm the usefulness of considering accessibility in our use case of HRI for older adults.

C. Difficulties Met

One of the main difficulties found during the evaluation process was the high cost of organizing and doing the evaluation itself, due to the mixed methodological approach. Next, the main difficulties are summarized:

- *Users recruitment*. One of the main difficulties in case studies involving users with functional diversity is to recruit enough users of each type of ability/disability for the study to get statistically significant results. The difficulty is not only to involve the users in the research, by explaining the test simply and honestly, but also to make sure the users can come to the place where the evaluation is carried out (usually involving caregivers or relatives) and prepare everything according to their needs.
- *Duration of structured interviews*. Taking into account that the CLARC robot's users are older adults, they could get tired and lose concentration if the evaluation session is too long. To recall, the interaction sessions include: the user test itself (interaction with the robot for the Barthel test), a pre-test interview and a post-test AUSUS questionnaire-based interview. When the authors designed the evaluation process, they realized the situation and tried to reduce to the minimum the number of questions of the structured interview. But, even taking this into account, the authors realized that the sessions with the patients became too long. If the authors would evaluate each of the factors thoroughly, including more questions in the questionnaire, it would be very likely that some users could not finish properly the evaluation sessions.
- *Methods' cost*. The methods used during the evaluation process require the presence of at least one expert (clinician, engineer, field researcher, etc.). Also, the time spent during the evaluation process and the *a posteriori* analysis of the results is considerable. Apart from being time consuming, qualitative methods imply the engagement of the researcher. Contrary to quantitative methods, doing user studies, or more generally fieldwork, is not just "data collection". The researcher is required to actively interact with the users, from welcoming them and presenting the tests, to competent interviewing. The validity of the data collected depends on these skills, which require that computer scientists are familiar, favorable or *a minima* trained to these methods.
- *Number of guidelines/heuristics/recommendations*. The large number of accessibility, usability and UX guidelines, heuristics, recommendations, etc. requires a high effort to learn and apply them appropriately. Moreover, the heuristic evaluations are very time consuming, so a selection of the main interaction tasks types is recommended.

However, despite the cost, the authors are convinced of the usefulness of this mixed methodological approach to gather relevant insights about socially assistive robots' use. Indeed, through our approach, and in particular through the proposal of the AUSUS framework, the aim is an in-depth understanding of situations involving robots and older adults. This includes the HRI dimension for sure, but considers design of HRI as part of a broader research aim. Thus, our proposal of the AUSUS framework - that makes the users' specific needs become the central concern - is part of an effort to go a step further in adopting a human-centred approach [38], an approach to the design of technologies that is focused on older adults' *abilities* (rather than dis-abilities), which corresponds to the latter's needs and values, and aims at empowering them.

VII. Recommendations for Using AUSUS

AUSUS is a methodological framework which takes into account a considerable number of evaluation factors, each one with different qualitative and quantitative indicators to be considered. This makes AUSUS a very complete evaluation framework for HRI. However, the relative high cost of applying this framework, as compared to mainstream quantitative methods, requires a strategy to be followed during the robot's development. The main recommendation is to apply different evaluation methods during the robot's iterative design process: start with qualitative methods which allow to collect and analyze the main user needs, use this in-depth understanding of users' needs and values to inform design, and continue with methods that allow to refine and extend the user requirements. Despite of the techniques used, the approach to design - of HRI or any other technology intended for older adults - is *human-centred*, giving primacy to older adults needs, values and abilities, at all times during development and evaluation. From a procedural perspective, research could move forward starting with the lower cost methods such as heuristic evaluations, to the higher cost methods, such as complete user studies.

In order to reduce the cost of user studies for accessibility evaluations, for example, it is recommended to carry out *informal evaluations* of specific accessibility issues during the robot's interface development, by testing even small improvements and iterating often, following Nielsen's recommendations for user testing [39]. For instance, ask someone

TABLE III
PATIENT'S STRUCTURED INTERVIEW DEALING WITH ACCESSIBILITY

Question ID	Description	Mean	SD
q1	I could hear and understand the robot clearly	4.14	1.21
q2	I could see everything in the robot's tablet (text, buttons, images, etc.)	4.28	1.11
q3	I could easily interact with the robot by speaking	2.00	1.57
q4	I could easily interact with the robot by using the touch screen	4.28	1.11
q5	I could easily interact with the robot by using the remote control	4.85	0.37
q6	I could understand the robot easily	4.14	1.21
q7	I found errors during the interaction	2.00	1.57

to execute a specific task, observe the interaction and discuss the issues with him/her. Then, a *formal evaluation* of accessibility issues can be carried out once the system seems to be accessible (after conformance evaluations and informal user studies), gathering both quantitative and qualitative data from representative users performing specific tasks. These evaluations would then involve a fairly large number of test subjects, and the accessibility indicators would be measured carefully in several different ways, so that the results can be generalized to other situations. This can be also applied to evaluate other factors of AUSUS framework.

Considering the wide range of functional diversity when working with users with special needs or presenting any kind of disability, as older users, the recommendation is to focus the evaluations on the target users. If they include a higher percentage of people with disabilities, focus on those relevant characteristics. For example, if the target users are seniors and the robot's main aim is to interact with persons with diabetes, even if most of them would have age-related disabilities, the most important disability to take into account is the visual disability directly related to diabetes.

Finally, the last recommendation is to reduce the time spent by each user in the evaluation. As explained, one of the difficulties met is the risk of fatigue and loss of concentration. From a pragmatic perspective, this could result in incomplete/biased data. From an ethical perspective, tests that are too long could tire older adults uselessly and put them in a failure situation. Therefore, to avoid these risks, it could be helpful to plan the experiments according to issues under investigation at a given moment. The research could be organized to evaluate one of the five factors of the AUSUS framework, in different evaluation tests during the whole robot's design process, taking into account that designs go through several iterations of interface testing. Then, the nearly final version of the HRI design (which is likely to receive less "to-be-improved" feedback, shortening the test length) could be evaluated in a formal way, based on the whole AUSUS framework. Combined by the quantitative aspect (adequate number of representative test subjects / participants), the rigour and generalizability of the results produced by the AUSUS framework will be guaranteed.

VIII. Conclusions

The paper discusses and proposes to improve classical Human-Robot Interaction (HRI) evaluation frameworks. The HRI *accessibility* factor is identified as a necessary evaluation item to take into account in systems where users with functional diversity are interacting. Therefore, a classical HRI evaluation framework has been relevantly chosen to be complemented with the *accessibility* factor. The framework chosen, USUS, is one of the most complete evaluation frameworks for HRI, combining methods derived from various research fields: HRI, HCI, psychology and sociology, and therefore, providing a holistic view of the evaluation of robots. The extension of USUS has been named AUSUS. A use case of the application of AUSUS in real scenarios is presented, evaluating the performance indicators of CLARC robot, a socially assistive robot interacting with seniors in a retirement home and in a hospital. The paper presents in detail the evaluation setup, the recruitment process followed, the materials and the mechanisms used to evaluate the performance indicators. The main difficulties faced during the evaluation are summarized and recommendations about the use of AUSUS are discussed.

Acknowledgement

This work has been partially funded by the EU ECHORD++ (FP7-ICT-601116) and PT13/006/001 projects, by FEDER/Ministerio de Ciencia, Innovación y Universidades - Agencia Estatal de Investigación RTI2018-099522-B-C43, TIN2017-88476-C2-2-R and CSO2017-86747-R projects and by AT17-5509-UMA and UMA18-FEDERJA-074 regional projects. It has been also supported by the Madrid Government (Comunidad de Madrid-Spain) under the Multiannual Agreement with UC3M in the line of Excellence of University Professors (EPUC3M17), and in the context of the V PRICIT (Regional Programme of Research and Technological Innovation). Javier García is partially supported by the Comunidad de Madrid (Spain) funds under the project 2016-T2/TIC-1712; Karine Lan Hing Ting by Berger Levrault as part of the company's robotics research delegation to the ActivAgeing Living Lab. The authors warmly thank the members of the "Amis du Living Lab" community and the patients and clinicians of Hospital Virgen del Rocío of Seville for their participation in this research.

References

[1] Zheng, Zhi and Zhao, Huan and Swanson, Amy R and Weitlauf, Amy S and Warren, Zachary E and Sarkar, Nilanjan Design, dev°elopment, and evaluation of a noninvasive autonomous robot-mediated joint attention intervention system for young children with ASD. IEEE transactions on human-machine systems, 48(2):125–135, 2017.

[2] Erfanian, Aida and Hu, Yaoping and Zeng, Tao Framework of multiuser satisfaction for assessing interaction models within collaborative virtual environments IEEE Transactions on Human-Machine Systems, 47(6):1052–1065, 2017.

[3] Petrie, Helen and Bevan, Nigel The Evaluation of Accessibility, Usability, and User Experience The universal access handbook, 1: 1–16, 2009.

[4] Julio Abascal and Colette Nicolle. Moving towards inclusive design guidelines for socially and ethically aware hci. *Interacting with computers*, 17(5):484–505, 2005.

[5] I Elaine Allen and Christopher A Seaman. Likert scales and data analyses. *Quality progress*, 40(7):64, 2007.

[6] JJ Baztán, J Pérez del Molino, T Alarcón, E San Cristóbal, G Izquierdo, and J Manzarbeitia. Índice de barthel: instrumento válido para la valoración functional de pacientes con enfermedad cerebrovascular. *Revista Española de Geriatría y Gerontología*, 28(1):32–40, 1993.

[7] BBC. Accessibility standards and guidelines. 2014.

[8] BBC_Recommendations. Accessibility standards and guidelines, 2014.

[9] Marco Billi, Laura Burzagli, Tiziana Catarci, Giuseppe Santucci, Enrico Bertini, Francesco Gabbanini, and Enrico Palchetti. A unified methodology for the evaluation of accessibility and usability of mobile applications. *Universal Access in the Information Society*, 9(4):337–356, 2010.

[10] Giorgio Brajnik. Beyond conformance: the role of accessibility evaluation methods. In *International Conference on Web Information Systems Engineering*, pages 63–80. Springer, 2008.

[11] Joost Broekens, Marcel Heerink, Henk Rosendal, et al. Assistive social robots in elderly care: a review. *Gerontechnology*, 8(2):94–103, 2009.

[12] World Wide Web Consortium. Web Accessibility Initiative. Authoring Tool Accessibility Guidelines (ATAG) 2.0. https://www.w3.org/TR/ATAG20/ 2015.

[13] World Wide Web Consortium. Web Accessibility Initiative. User Agent Accessibility Guidelines (UAAG) 2.0. https://www.w3.org/WAI/standards-guidelines/uaag/ 2015.

[14] World Wide Web Consortium. Web Accessibility Initiative. Web content accessibility guidelines (wcag) 2.0. https://www.w3.org/TR/WCAG20/ 2008.

[15] World Wide Web Consortium. Web Accessibility Initiative. Website Accessibility Conformance Evaluation Methodology (WCAG-EM) 1.0. https://www.w3.org/TR/WCAG-EM/ 2014.

[16] Graham Ellis and Peter Langhorne. Comprehensive geriatric assessment for older hospital patients. *British Medical Bulletin*, 71(1):45–59, 2005.

[17] MV Escribano-Aparicio, M Pérez-Dively, FJ García-García, A Pérez-Martín, L Romero, G Ferrer, E Martin-Correa, and MI Sanchez-Ayala. Validación del mmse de folstein en una población española de bajo nivel educativo. *Rev Esp Geriatr Gerontol*, 34(6):319–326, 1999.

[18] Susanne Iwarsson and Agnetha Ståhl. Accessibility, usability and universal design—positioning and definition of concepts describing person-environment relationships. *Disability and rehabilitation*, 25(2):57–66, 2003.

[19] RI Mahoney. Barthel index (bi). *Surya Shah, PhD, OTD, MEd, OTR, FAOTA, Professor Occupational Therapy and Neurology, Visiting Professor Neurorehabilitation, University of Tennessee Health Sciences Center*, 930:1, 1965.

[20] Nielsen_Norman_Group_Report. Beyond ALT Text: Making the Web Easy to Use for Users with Disabilities, 2001.

[21] Funka Nu. Mobile navigation guideline. 2014.

[22] Helen Petrie and Omar Kheir. The relationship between accessibility and usability of websites. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 397–406. ACM, 2007.

[23] Aaron Steinfeld, Terrence Fong, David Kaber, Michael Lewis, Jean Scholtz, Alan Schultz, and Michael Goodrich. Common metrics for human-robot interaction. In *Proceedings of the 1st ACM SIGCHI/SIGART Conference on Human-robot Interaction*, HRI '06, pages 33–40, New York, NY, USA, 2006. ACM.

[24] Sebastian Thrun. Toward a framework for human-robot interaction. *Human-Computer Interaction*, 19(1):9–24, 2004.

[25] Katherine M. Tsui, James M. Dalphond, Daniel J. Brooks, Mikhail S. Medvedev, Eric McCann, Jordan Allspaw, David Kontak, and Holly A. Yanco. Accessible human-robot interaction for telepresence robots: A case study. *Paladyn*, 6(1), 2015.

[26] Viswanath Venkatesh, Michael G Morris, Gordon B Davis, and Fred D Davis. User acceptance of information technology: Toward a unified view. *MIS quarterly*, pages 425–478, 2003.

[27] D. Voilmy, C. Suarez, A. Romero-Garcés, C. Reuther, J.C. Pulido, R. Marfil, L.J. Manso, K. Lan Hing Ting, A. Iglesias, J.C. González, J. García, A. García Olaya, R. Fuentetaja, F. Fernández, A. Dueñas, L.V. Calderita, P. Bustos, T. Barile, J.P. Bandera, and A. Bandera. CLARC: A cognitive robot for helping geriatric doctors in real scenarios. In *ROBOT (1)*, volume 693 of *Advances in Intelligent Systems and Computing*, pages 403–414. Springer, 2017.

[28] Romero Garcés, A., Marfil, R., Martinez-Cruz, J., Bandera-Rubio, J. P., Bandera-Rubio, A. J., Garcés, R. Towards a Robust Robotic Assistant for Comprehensive Geriatric Assessment Procedures: the CLARC system. In *27th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN'18)*, In press. 2018.

[29] Karine Lan Hing Ting, Dimitri Voilmy, Ana Iglesias and Rebeca Marfil Data Collection in CGA Evaluation: Analyzing Practices to Inform Design. In *Data-work in Healthcare Workshop at the 21st ACM Conference on Computer-Supported Cooperative Work and Social Computing (CSCW'18)*, In press. 2018.

[30] Astrid Weiss, Regina Bernhaupt, Michael Lankes, and Manfred Tscheligi. The usus evaluation framework for human-robot interaction. In *AISB2009: proceedings of the symposium on new frontiers in human-robot interaction*, volume 4, pages 11–26, 2009.

[31] Astrid Weiss. Validation of an evaluation framework for human-robot interaction: the impact of usability, social acceptance, user experience, and societal impact on collaboration with humanoid robots. PhD thesis. University of Salzburg, 2010.

[32] Holly A Yanco and Jill Drury. Classifying human-robot interaction: an updated taxonomy. In *systems, man and cybernetics, 2004 IEEE International Conference on*, volume 3, pages 2841–2846. IEEE, 2004.

[33] Fillenbaum, Gerda G. Multidimensional functional assessment of older adults: The Duke Older Americans Resources and Services procedures. Psychology Press, 2013.

[34] Kelly, B., Sloan, D., Brown, S., Seale, J., Petrie, H., Lauke, P., and Ball, S. Accessibility 2.0: people, policies and processes. In *Proceedings of the 2007 international cross-disciplinary conference on Web accessibility (W4A)*, pp. 138-147, ACM, 2013.

[35] Qbilat, M. and Iglesias, A. Accessibility Guidelines for Tactile Displays in Human-Robot Interaction. A Comparative Study and Proposal. In *Proceedings of international conference on Computers Helping People with Special Needs*, pp. 217-220, Springer, 2018.

[36] Qbilat, M. and Iglesias, A. Accessibility Guidelines for Hardware and Software Assistive Technology in Human-Robot Interaction. In *Workshop of User-Centered Methods in Human-Robot Interaction. 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems, In press, 2018.*

[37] Brajnik, G. Comparing accessibility evaluation tools: a method for tool effectiveness. Universal access in the information society. Vol 3, Pp. 252-263. 2004.

[38] Bannon, L. Reimagining HCI: toward a more human-centered perspective. Interactions, 18(4), 50-57.

[39] Nielsen, Jakob Usability engineering Morgan Kaufmann Publishers Inc. San Francisco, CA; 1994.

[40] Heerink, M., Kröse, B., Evers, V., Wielinga, B. Assessing acceptance of assistive social agent technology by older adults: the almere model International journal of social robotics, 2(4), 361-375; 2010

[41] Sung, J., Grinter, R. E., Christensen, H. I. Domestic robot ecology International Journal of Social Robotics, 2(4), 417-429; 2010

[42] United Nations United Nations and the International Federation of Robotics. Proceedings of the World Robotics 2002, New York.

[43] Feil-Seifer, David and Mataric, Maja J. Defining Socially Assistive Robotics Proceedings of the 9th International Conference on Rehabilitation Robotics (ICORR), 465-468, 2005, IEEE.

[44] Palmerini et al. Palmerini, Erica and Azzarri, Federico and Battaglia, Fiorella and Bertolini, Andrea and Carnevale, Antonio and Carpaneto, Jacopo and Cavallo, Filippo and Di Carlo, Angela and Cempini, Marco and Controzzi, Marco and others Guidelines on regulating robotics 2014

[45] Wallström, Josefine and Lindblom, Jessica Design and Development of the USUS Goals Evaluation Framework Human-Robot Interaction, 177-201,2020.

[46] De Graaf, Maartje MA and Allouch, Somaya Ben Exploring influencing variables for the acceptance of social robots Robotics and autonomous systems, 61(12), 1476-1486, 2013

[47] Feil-Seifer, David and Mataric, Maja J. Defining socially assistive robotics 9th International Conference on Rehabilitation Robotics (ICORR'05), 465-468, 2005.

[48] Bannon, Liam Reimagining HCI: toward a more human-centered perspective Interactions, 18(4), 50-57, 2011 ACM New York, NY, USA

[49] Bratteteig, Tone and Wagner, Ina What is a participatory design result? Proceedings of the 14th Participatory Design Conference: Full papers-Volume 1, 141-150, 2016.

[50] Karine Lan Hing Ting, Dimitri Voilmy, Ana Iglesias, Rebeca Marfil and Quitterie De Roll Fieldwork and field trials in hospitals: Co-designing a robotic solution to support data collection in geriatric assessment Applied Sciences, 2021 (In press).

[51] Östlund, Ulrika and Kidd, Lisa and Wengström, Yvonne and Rowa-Dewar, Neneh Combining qualitative and quantitative research within mixed method research designs: a methodological review International journal of nursing studies, 48(3), 369-383, 2011. Elsevier.

[52] Mays, Nicholas and Pope, Catherine Assessing quality in qualitative research Bmj, 320(7226), 50-52, 2000. British Medical Journal Publishing Group.

[53] O'Cathain, Alicia and Murphy, Elizabeth and Nicholl, Jon Integration and publications as indicators of" yield" from mixed methods studies Journal of mixed methods research, 1(2), 147-163, 2007. Sage Publications Sage CA: Los Angeles, CA

[54] Lindgaard, Gitte and Chattratichart, Jarinee Usability testing: what have we overlooked? Proceedings of the SIGCHI conference on Human factors in computing systems, 1415–1424, 2007.

[55] Nielsen, J. Why You Only Need to Test with Five users (Online) http://www.useit.com/alertbox/20000319.html (last accessed on March, 2021)

[56] Okamura, Allison M and Matarić, Maja J and Christensen, Henrik I Medical and health-care robotics IEEE Robotics & Automation Magazine 17(3), 26-37, 2010. IEEE

[57] Creswell, John W and Fetters, Michael D and Ivankova, Nataliya V Designing a mixed methods study in primary care The Annals of Family Medicine 2(1), 7-12, 2004

[58] Al-Qbilat, M., Iglesias, A. and Belpaeme, T. A Proposal of Accessibility Guidelines for Human-Robot Interaction Electronics 2021, 10(5), 561-575, 2021 https://doi.org/10.3390/electronics10050561

[59] Kulyukin, Vladimir and Gharpure, Chaitanya and De Graw, Nathan Human-robot interaction in a robotic guide for the visually impaired AAAI Spring Symposium, pp. 158–164, 2004

[60] Fattal, Charles and Cossin, Isabelle and Pain, Frédérique and Haize, Emilie and Marissael, Charline and Schmutz, Sophie and Ocnarescu, Ioana Perspectives on usability and accessibility of an autonomous humanoid robot living with elderly people Disability and Rehabilitation: Assistive Technology, pp.1–13, 2020