

## Article

# Exploring the Influencing Factors on User Experience in Robot-Assisted Health Monitoring Systems Combining Subjective and Objective Health Data

Caterina Neef , Katharina Linden  and Anja Richert 

Cologne Cobots Lab, TH Köln—University of Applied Sciences, 50679 Köln, Germany

\* Correspondence: caterina.neef@th-koeln.de (C.N.); anja.richert@th-koeln.de (A.R.);

Tel.: +49-221-8275-2382 (C.N.); +49-221-8275-2756 (A.R.)

**Featured Application:** A health monitoring system that combines commercially available products, such as sensing devices and apps, with a social robot to conduct health assessments and collect subjective and objective health data. The study conducted in this work suggests that the system has a high rating for usability and user experience and that it has the potential to be used for self-managing users' health.

**Abstract:** As the population ages, the demand for care for older adults is increasing. To maintain their independence and autonomy, even with declining health, assistive technologies such as connected medical devices or social robots can be useful. In previous work, we introduced a novel health monitoring system that combines commercially available products with apps designed specifically for older adults. The system is intended for the long-term collection of subjective and objective health data. In this work, we present an exploratory user experience (UX) and usability study we conducted with older adults as the target group of the system and with younger expert users who tested our system. All participants interacted with a social robot conducting a health assessment and tested sensing devices and an app for data visualization. The UX and usability of the individual components of the system were rated highly in questionnaires in all sessions. All participants also said they would use such a system in their everyday lives, demonstrating the potential of these systems for self-managing users' health. Finally, we found factors such as previous experience with social robots and technological expertise to have an influence on the reported UX of the users.

**Keywords:** older adults; healthy ageing; health monitoring; health assessment; assistive technologies; social robotics; chatbots; user experience



**Citation:** Neef, C.; Linden, K.; Richert, A. Exploring the Influencing Factors on User Experience in Robot-Assisted Health Monitoring Systems Combining Subjective and Objective Health Data. *Appl. Sci.* **2023**, *13*, 3537. <https://doi.org/10.3390/app13063537>

Academic Editor: Oscar Reinoso García

Received: 16 December 2022

Revised: 4 March 2023

Accepted: 5 March 2023

Published: 10 March 2023

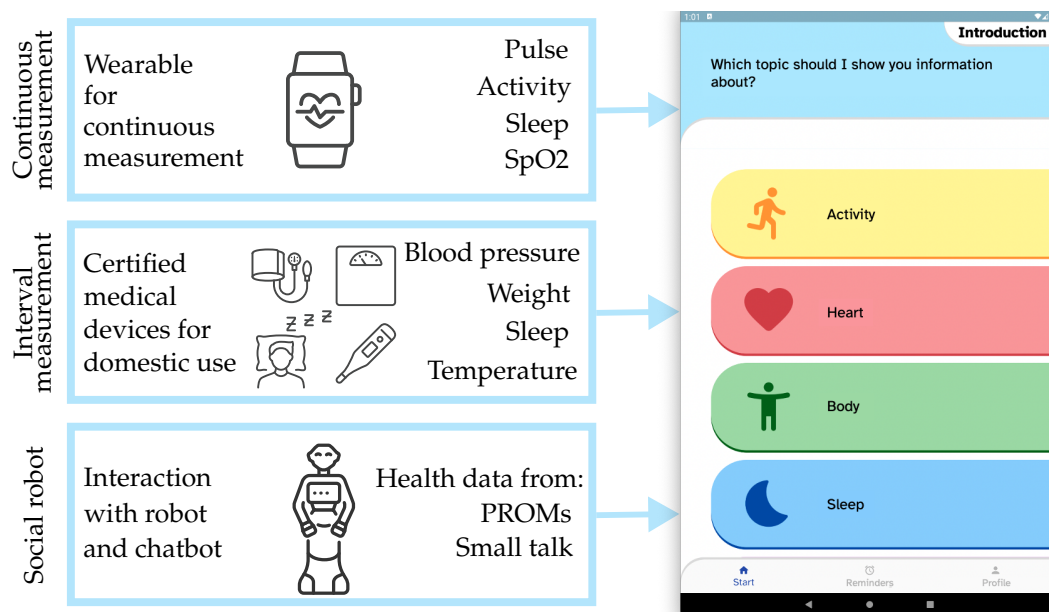


**Copyright:** © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction and Related Work

Developing and deploying health monitoring systems that older adults can use by themselves to self-manage their health while also increasing their health literacy remains a challenge relevant to many societies across the globe [1]. In this regard, it is vital to develop systems that this target group can and wants to use to ensure that they actually provide a benefit. Such systems present an opportunity to combat the challenges arising due to an increase in the number of older adults, with the simultaneously increasing shortage of healthcare workers [2]. Additionally, many older adults want to age in their own homes rather than move to long-term care facilities [3]. Assistive technologies can support and enable this, and many technologies, such as sensing devices or mobile applications, exist to this end [4]. However, they are often difficult to use for older adults, especially those in need of care, as they may have impairments that make it harder to interact with such systems, such as hearing, visual, motor, and cognitive impairments [5]. Therefore, it is important to design, develop and rigorously test assistive technologies for and with this intended target group.

In previous work [6], we introduced a system architecture for a health monitoring system comprised of a social robot, several connected health devices, and a mobile app for data visualization. An overview of the system is shown in Figure 1. By using a wearable for continuous measurement of health data and certified connected health devices for interval measurements the objective health data, e.g., pulse, activity, blood pressure, or body temperature, can be measured. Through interactions with a social robot, in the form of small talk and validated health assessments, such as PROMs, subjective health data can be acquired. This system was developed for everyday long-term autonomous use by both older adults in care facilities and community-dwelling older adults. In this work, we present a first exploratory UX evaluation of our health monitoring system to investigate factors influencing the reported UX and usability, and their implications for ensuring the system's successful long-term use.



**Figure 1.** An overview of our health monitoring system, described in detail in [6], is shown. Health data are acquired with a smartwatch, various connected devices for domestic use, as well as through interaction with a social robot and its underlying chatbot. The data can be enquired and visualized in a multi-platform mobile application, shown on the right.

### 1.1. Assistive Technologies for Health Assessment

A number of assistive technologies for health assessment and monitoring have been developed. In [7], Durán-Vega et al. present a remote health monitoring system with a wearable device and a mobile application. The system enables caregivers to make informed decisions about the health of their patients. The included mobile application is intended for relatives and caregivers. Dratsiou et al. [8] present a framework of various digital solutions that can support the well-being of older adults, including applications for physical and cognitive training. Many wearable technologies for different applications in the healthcare domain exist [9]. They can add value to a user's life when motivation, user experience, and device features suit the users' needs [10].

Another technology used in many different assistive capacities and environments is social robots. Among these is the robot iCat, a plastic cat with a face and a body, used for health self-management of older adults [11]. When behaving in a more social manner, such as expressing emotions and being attentive, this robot showed higher acceptance among users. Another example of a social robot used in an assistive context is the humanoid robot Pepper [12], which has been used to coach persons with dementia to motivate them to perform stimulating activities [13]. It has also been tested in domestic environments [14], showing high acceptability among elderly participants. The use of social robots has also

led to positive outcomes when used in psychosocial health interventions [15]. Additionally, they have shown high acceptance in terms of perceived enjoyment, ease of use, and trust with elderly users in their own homes [16], as well as high perceived usefulness and acceptance in supportive apartment living facilities [17]. Acceptance and interest in using social robots are higher among older adults than younger adults [18], as they tend to spend more time at home and have more difficulties with vision, hearing, or dexterity, areas in which robots can assist. While acceptance and willingness to use social robots for daily assistance are important factors, long-term motivation to actually use a robot is not guaranteed. The novelty effect can lead to users interacting with the robot less after it has been in the home for a certain time [19]. This is an important factor to be considered when developing robotic applications for long-term use.

Social robots have also been used to conduct health assessments using e.g., PROMs [20,21]. They have shown to be a promising approach in this context, as they lead to more positive interactions when delivering healthcare instructions, including asking health-related questions, compared to a computer tablet [22]. Boumans et al. [21] additionally found that geriatric patients find robot-assisted health assessments of equal value to non-robot-assisted assessments, further confirming the feasibility of using a robot for such health data acquisitions. Additionally, social robots have been tested for clinical screening interviews [23], including cognitive assessment, fall risk evaluation, and pain rating. They achieved satisfying results and high confidence and trust ratings from older adults. These findings indicate the suitability of social robots for conducting health assessments.

### 1.2. UX of Health Monitoring Systems

When developing applications for socially interactive robots, it is important to ensure a positive UX for potential target groups of the application [24]. There are several categories of barriers that can impact the usability and UX of mobile health applications in older adults, namely cognition, motivation, physical ability, and perception [5]. Different recommendations, therefore, apply to applications for older adults, such as ensuring a reduction of complexity, clear structure, consistency, continuous feedback, user support, as well as optimizing the interface in terms of, e.g., font sizes and making use of multi-modal interaction modes [25].

The most common methods to assess user experience are questionnaires and interviews, and it is crucial to perform such assessments early in the development process [26]. Questionnaires used include the UEQ [27], which has been used, e.g., to assess the perceived attractiveness and usability of a robot for older adults with memory impairments [28]. The UEQ also exists in a modular version, called the UEQ+ [29], which includes a variety of short questionnaires that can be selected as needed. This includes questions for assessing, e.g., perspicuity, dependability, usefulness, or intuitive use, among others. A questionnaire suitable specifically to assess the usability of technologies is the SUS [30]. Several sources [31,32] concluded that the SUS is a valuable, reliable tool for usability evaluations. The SUS has also been used to assess the usability of robotic systems for older adults [33] and cognitive testing with social robots [34].

In the following, we present the usability and UX study of our developed health monitoring system, including all used system components, a detailed description of the study design, the results and a discussion of the resulting implications for the future. We took the learnings from the presented works into account for the development and evaluation of our health monitoring system. While many different health monitoring systems exist, ours is, to the best of our knowledge, the first to combine subjective and objective health data, as this provides a more complete overview of a person's overall health status. This is described in more detail in [6].

## 2. Materials and Methods

The tested health monitoring system comprises a health assessment administered by a social robot, the use of connected health devices, as well as a mobile app for data visualization, all of which are described in the following.

### 2.1. Social Robot

The robotic platform chosen for this project was the social robot Pepper, introduced in [12]. The humanoid robot is manufactured by Aldebaran (Paris, France), is 1.21 m tall, and weighs 28 kg. While the focus on interactions with Pepper lies in the verbal interactions with users, the 10" Android tablet on the robot's chest serves as a UI, used to display supporting text or images. Pepper's childlike, androgynous appearance was designed to avoid bias and unrealistic expectations [12]. It is also already widely used and receives a generally high acceptance, specifically also among older users [13,14,35]. We, therefore, chose to employ Pepper in this study, with the expectation that this appearance would help ensure that users are not intimidated or frightened by the robot.

We used Pepper's NAOqi SDK, and applications for the robot were developed as an Android app. The programming language used was Kotlin. We developed our robotic application for the project in accordance with the interaction principles described in DIN EN ISO 9241-110 [36]: In an interactive introduction, the robot itself provides instructions for how to interact with it, thus adhering to the principles of learnability and self-descriptiveness of a system. The introduction was scripted in QiChat, a dialogue framework developed by Aldebaran for the robot Pepper. After the introduction is complete, the robot begins the actual health assessment.

#### 2.1.1. Health Assessment

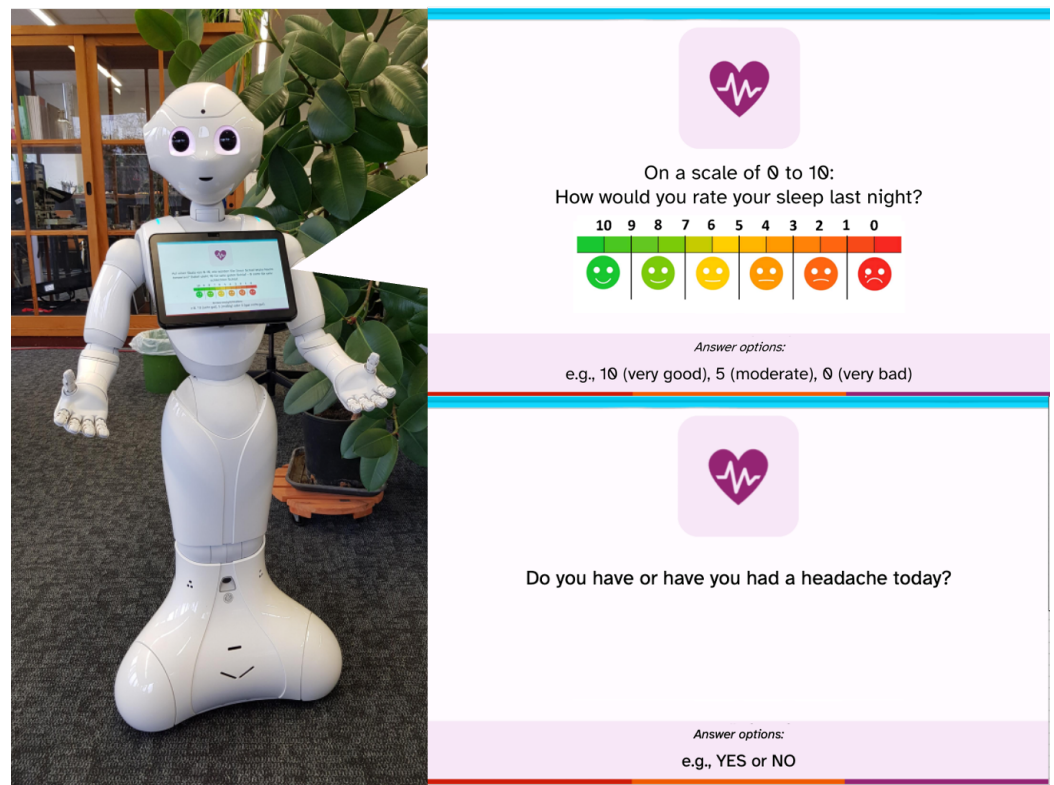
The health assessment used in the study includes a total of 31 standard questions (items), along with two conditional items, and is listed in Appendix A. We created the questionnaire ourselves and designed it specifically for daily use. In an interview with a healthcare professional, we first identified parameters with both medical relevance and variability within short periods of time, for which a daily assessment would prove beneficial. Based on these parameters, we created individual questions, that can be divided into five sub-categories: (1) general condition, (2) balanced nutrition and hydration level, (3) sleep quality, (4) cognitive issues, and (5) common symptoms for specific health impairments that are prevalent in older adults. To create these items, we considered a specific disease pattern, e.g., hypertension, and the respective symptoms, such as headaches or vertigo. An example of a symptom-derived question is, as shown in Figure 2: "Do you have or have you had a headache, today?"

Each item of our questionnaire has predetermined answer options, which are displayed on the tablet to assist users in answering the robot's questions (see Figure 2). For two of the items, users are prompted to rate their condition based on a scale from 0 to 10. To make these rating scales more accessible to persons with cognitive impairments [37], the SAS is displayed in addition. Additionally, the font Atkinson Hyperlegible, designed by the Braille Institute (Los Angeles, CA, USA) for low-vision readers, was used for all interfaces of the health assessment app.

#### 2.1.2. Dialogue Management

As a dialogue manager and underlying chatbot, we use the conversational AI framework Rasa (San Francisco, CA, USA). The health assessment is structured using stories, which include the user intent, as well as appropriate responses from the bot, which vary depending on the user's answers. Additionally, parts of the conversation are structured using custom actions, which can be defined in Python scripts, as well as forms, which are provided by Rasa for rule-based information gathering [38]. The default configuration for the components and policies of the machine learning model was used. The application

running on the social robot connects to the Rasa dialogue manager via a REST API, further details are described in [6].



**Figure 2.** Two screenshots of the health assessment app are shown: One question enquiring about the user’s sleep, the other about possible headaches. To ensure users with visual or hearing impairments correctly understand Pepper’s questions, the robot not only asks the questions verbally but they are also printed on its tablet. The screenshots were translated into English, the app was developed in German.

## 2.2. Connected Health Devices

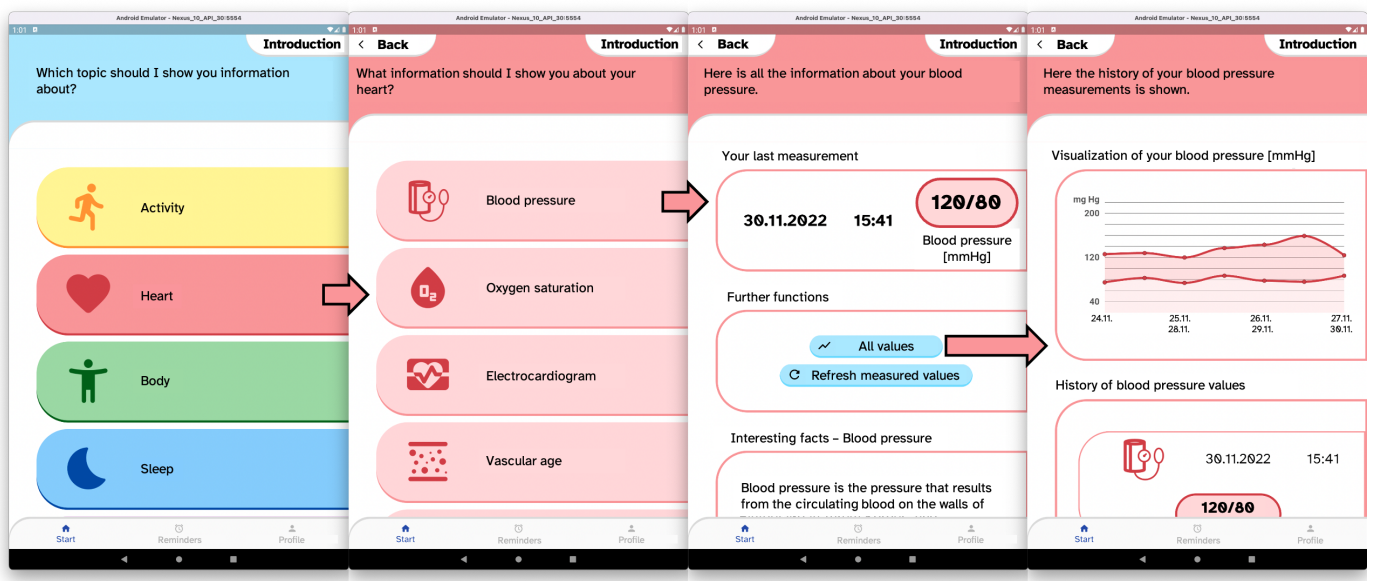
To complement the subjective data from the assessment, we used various connected health devices to collect objective health data, such as vital parameters and activity data. We chose devices by manufacturer Withings France SA (Issy-les-Moulineaux, France), who offer medical-grade sensors that have been used and validated in various studies, e.g., [39–41]. Moreover, Withings offers an open API that allows the integration of the devices into third-party applications [42]. Within the course of our study, the participants tested a wearable (ScanWatch), a smart thermometer (Thermo), as well as a blood pressure monitor (BPM core).

The ScanWatch is a hybrid smartwatch with a range of features and functions. In addition to the continuous monitoring of vital parameters and activity detection, it also serves as an oximeter and performs a medical-grade EKG [40,43]. The smart thermometer Thermo is a no-contact device for measuring core body temperature from the temporal artery. The result of a measurement is equivalent to the rectal temperature with a clinical accuracy of 0.2 °C [43]. The blood pressure monitor BPM core measures systolic and diastolic blood pressure with an accuracy of 2% of the reading. It can also perform a full EKG and record heart sounds. A detailed overview of the system architecture and the explanations for the chosen devices and their measurable parameters can be found in [6].

## 2.3. Mobile Application

All collected health data can be visualized in a multi-platform mobile app we developed using the framework Flutter, created by Google (Mountain View, CA, USA). The app

can be used on mobile, desktop and web clients. For this study, we chose an Android app on an 11" tablet. The health data from the connected devices are stored in MySQL databases. Subsequently, they are imported via REST API into our mobile app for visualization. A detailed description of the system architecture including all interfaces can be found in [6]. The app's UI were designed specifically considering the needs of older users. Thus, the structure of the layouts was kept consistent, we used clearly recognizable clickable items and icons, and made a choice of colors taking into account good visibility and high-contrast gradation. Again, the font Atkinson Hyperlegible was used for all interfaces of the app. Furthermore, to have the app adapt to the user's needs, the font size of the text items on the screen automatically adapts based on the system's default font size setting. Screenshots of the app are shown in Figure 3.



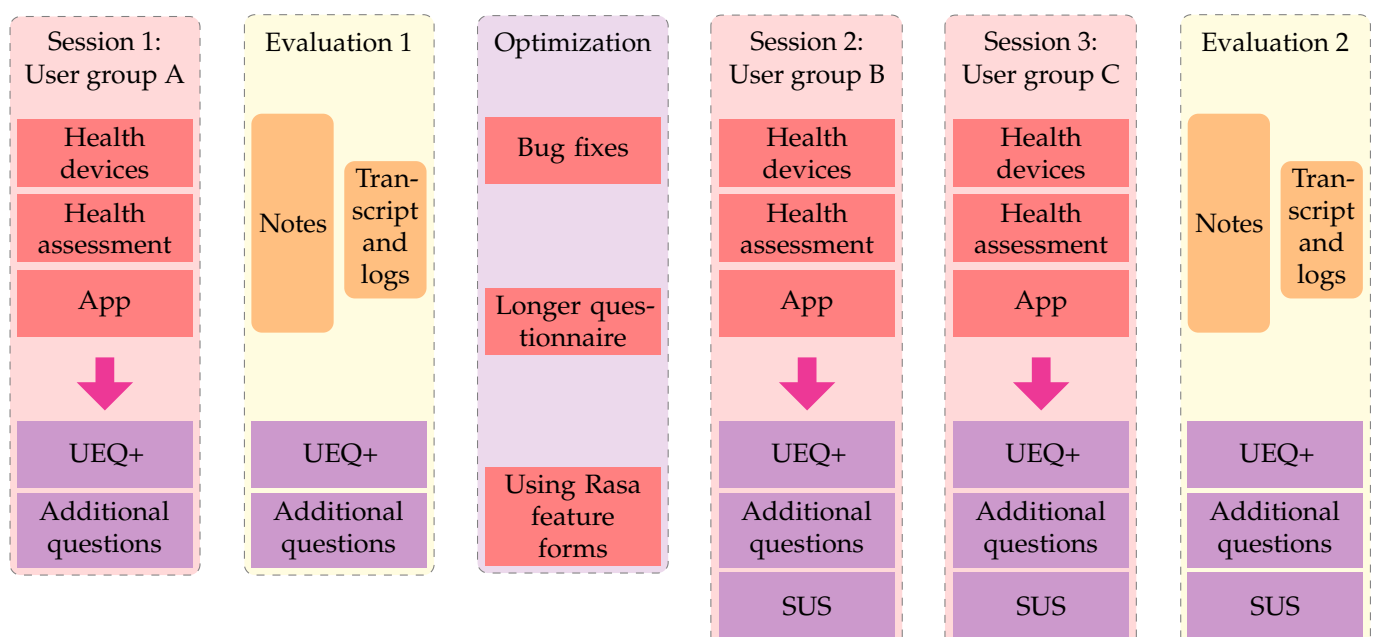
**Figure 3.** Screenshots of the mobile app are shown, including the category heart and its sub-category blood pressure, where both the current blood pressure value and the history of the measurements are displayed. The screenshots were translated into English, the app was developed in German.

From the app's main menu, users can select the data they wish to view from the following main categories: activity, heart, body, sleep, and health assessment. Selecting one of the main categories routes the user to another UI, where corresponding sub-categories can be selected for viewing. With "heart" being an example of one of the main categories, examples of the corresponding sub-categories are "blood pressure" and "oxygen saturation". Comprehensive specific reports and visualizations of the individual health parameter progressions are available in each of the sub-categories. Furthermore, a summary report is available that combines the objective data from the devices with the subjective results from the health assessment.

#### 2.4. Study Design and Participants

We designed our study as a UX and usability test with a total of three sessions, as shown in Figure 4. We followed the recommendation for UX testing according to Nielsen [44], which recommends testing repeatedly with smaller groups of users. Over the course of the study, the system was tested with three different user groups. There was no predefined time limit for the tests. In all user groups, the tests lasted between 30 and 60 min with each participant. As the study was conducted in the German language, the participants in all groups were required to have native-level language proficiency. The study was reviewed and approved by the Ethics Research Committee of the TH Köln—University of Applied Sciences, and informed consent was obtained from all participants.

User group A consisted of eight participants from the target group, whose ages ranged from approximately 50 to 90 years. Five of the eight participants were between the ages of 60 and 79, two were between 80 and 89, and one participant was aged between 50 and 59. They all lived in the same care facility in Germany and were recruited by facility staff, who were instructed to select participants interested in testing assistive technologies in the context of health monitoring. The use of our system might prove beneficial for this user group, as it may increase their independence from care professionals and support them in retaining autonomy over their health care. Five of the eight participants took part in one of our previous studies with Pepper and therefore had some experience with the robot. Due to various physical and/or cognitive impairments, these participants are seen as a cross-section of adults living in residential care. Additionally, to gain insight into the user experience of the system from the point of view of care personnel, a caregiver of the older adults also tested and evaluated the system in session 1. They also had previous experience with the robot.

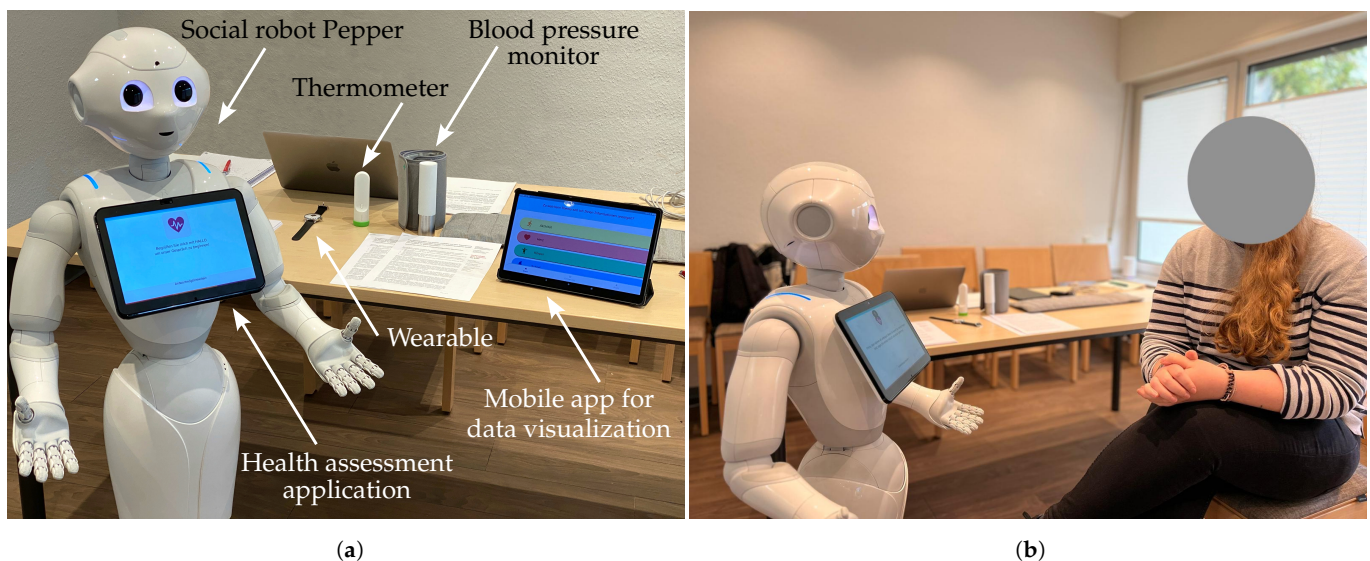


**Figure 4.** The design of our study, including three sessions with three different user groups, is visualized. Group A included older adults in a care facility, some with prior Pepper experience. Group B consisted of expert users with extensive chatbot or robot experience, while group C included older adults in assisted living with extensive Pepper experience. In each session, the health devices, the robot-assisted health assessment, and the visualization app were tested. We observed the system test and took notes. Additionally, the robot interaction was audio-recorded. Afterward, the questionnaire was filled out by participants. Minor system optimizations were performed after session 1.

To ensure that bugs and errors did not have a negative effect on the reported UX of the users, after session 1, we optimized the system based on the participants' feedback and our own findings. The optimizations include bug fixes, using a longer questionnaire for the health assessment, as well as using the Rasa feature forms, to enable more successful interactions with the robot. With the optimized system, we conducted test session 2 in our lab with user group B. We recruited participants for this group by asking for interested volunteers within our lab. These five participants were all men between the ages of 25 and 35 with a technical occupational orientation, and all worked with chatbots or robots. While they are not part of the intended target group of the system, this group was included to provide the perspective of expert users. The goal of session 2 was to further improve the optimized system, based on the UX and the opinions of these expert participants, as well as making sure everything worked error-free.

Finally, in session 3, we conducted a test of the optimized system with user group C. Unfortunately, due to the effects of the pandemic, we were only able to recruit three participants for this session. Group C consisted of three older adults, aged approximately 60 to 95 years, living in an assisted living community, thus providing the viewpoint of persons in a different care setting than in session 1. For active and largely autonomous users such as those of group C, our system may prove advantageous, as its goal is to promote health literacy and support their healthy aging. As opposed to group A, group C showed fewer and less pronounced health impairments. In addition, group C already had extensive experience with the social robot Pepper, as one is permanently deployed in their assisted living community in the framework of a research project. We recruited group C by asking for volunteers in this project who were interested in testing assistive technologies in the context of health monitoring.

In the study scenario, a participant was led into a room, where the testing took place and the research team was present. A member of the research team demonstrated the use of each of the connected health devices and prompted the participant to try it themselves. Upon testing the smartwatch, the thermometer, and the blood pressure monitor, the health assessment with the robot Pepper was conducted. The interaction with the social robot was audio-recorded for later comparison with the system logs. We deliberately refrained from recording videos of the interactions, to avoid possible bias. With the goal of assessing menu navigation and usability of the mobile app, the participants were given the task to find their previously recorded body temperature and blood pressure measurements in the app. The study scenario remained the same for all three sessions. The study setup, including all tested devices, is shown in Figure 5.



**Figure 5.** The study setup is shown. (a) All tested system components, including the connected health devices, the robot, and the app, are shown. (b) The interaction with the robot is demonstrated by one of the members of the research team in the room session 1 took place in.

### 2.5. UX Questionnaires

At the end of each session, the participants completed a questionnaire, rating their experience with the system. As the questionnaire for session 1, we chose the validated UEQ+ modular kit of UX items. This kit includes a wide range of scales, targeting various characteristics of UX. Each scale can be used and evaluated individually and consists of four to six items, each stating terms of opposite meaning, e.g., “useless/useful”. Participants rate their own experience on a 7-point Likert scale, which represents the difference between the terms [29]. A final item prompts participants to rate the relevance of the scale characteristics. For our questionnaire, we chose the scale “perspicuity”, targeting the participants’ overall



impression of handling and usage of the system. With this scale, the characteristics rated by participants are “understandable”, “easy to learn”, “easy” and “clear”. This scale was queried once for each individual component of our system - the health assessment, the use of the connected health devices, and the retrieval of health data in the mobile app.

In an additional item, we queried previous experience with the robot Pepper. Furthermore, we added questions on demographics and previous occupational orientation, to determine whether those factors had an effect on the UX results, and provided a text field for additional comments. For use in sessions 2 and 3, we expanded the questionnaires. We included the previously used UEQ+ scale for an evaluation of each of the system components, along with the system usability scale (SUS) [30] questionnaire for an evaluation of the entire system. The SUS consists of a total of 10 statements on different aspects of usability, such as “I think that I would like to use the system frequently” or “I found the system unnecessarily complex”. Each statement is to be rated on a 5-point Likert scale. The output of SUS is an easy-to-interpret score from 0 to 100 [30].

Furthermore, we added an item to determine to what extent the text display on Pepper’s tablet was used and perceived as helpful. Additionally, we queried the same additional items as in the first questionnaire. Due to various health-related impairments, all participants in sessions 1 and 3 received support from a member of the research team while completing the questionnaires. The help was provided by a member of the research team and consisted of reading the questions out loud, explaining answer options, and partially ticking the checkboxes on the survey form. The caregiver in session 1 and the participants in session 2 did not receive support.

## 2.6. Data Analysis

The data analysis for both the UEQ+ and SUS questionnaires was performed using available templates. For the UEQ+, this includes the calculation of the KPI of each individual component as well as the overall system KPI. The KPI for the UEQ+ are calculated similarly to the KPI for the UEQ questionnaire, from which it is derived. This calculation is described in [45]. The individual  $KPI_j$  for each participant  $j$  can be calculated as follows:

$$KPI_j = \left( \sum_{i=1}^3 M_{i,j} * I_{rel,i,j} \right) - 4. \quad (1)$$

Here,  $i$  corresponds to the UEQ+ parameter of the specific component, namely the robot-assisted assessment, the health devices and the app.  $M$  is the mean of each UX scale per participant, i.e., the perspicuity scale in this study, while  $I_{rel,i,j}$  refers to the relative importance of the specific characteristic of this scale. This number is then subtracted by 4 to normalize the 7-point Likert scale around zero, resulting in values between  $-3$  and  $3$ . The overall system KPI can then be calculated as the mean of all individual participant KPI using the following equation, with  $n$  being the number of all participants:

$$KPI = \frac{1}{n} \sum_{j=1}^n KPI_j. \quad (2)$$

For the scoring of the SUS, we use the following equation, as described in [30], where  $X$  represents the sum of the points for all odd-numbered questions subtracted by 5, while  $Y$  is calculated by subtracting the sum of the points for all even-numbered questions from 25:

$$SUS\ Score = (X + Y) * 2.5. \quad (3)$$

In addition to the UEQ+ and SUS, we evaluated the success and completion rates of the health assessments. We define the success rate as the rate at which both Pepper’s speech recognition and the Rasa dialogue manager correctly identify the user’s intent. The completion rate is the rate at which the entire health assessment, including all questions, is completed. To determine both the success and completion rates, we compared the

Rasa conversation logs with the transcripts we created from the audio recordings of the robot interactions.

### 3. Results

The results of the UX and usability tests of our health monitoring system are described in the following. An overview of the overall questionnaire results is shown in Table 1. For each of the system components, the perspicuity was rated using the UEQ+ questionnaire, for sessions 2 and 3 the SUS questionnaire was additionally used for evaluations. We calculated the means of the UEQ+ ratings and their corresponding rated importances, along with the component and overall system KPI, as well as the overall SUS scores.

**Table 1.** The overall results of the UEQ+ and SUS questionnaires are displayed.

<b>Session 1—Group A in residential care, N=8 + 1, 5 + 1 with previous Pepper experience</b>					
Perspicuity	Mean (SD)	Median	Importance (SD)	KPI (SD)	Range
Pepper	2.88 (0.48)	3	2.00 (1.41)	2.88 (0.33)	2–3
Sensors	2.25 (1.64)	3	2.00 (1.41)	2.25 (1.54)	1.75–3
App	2.71 (1.13)	3	2.00 (1.77)	2.71 (0.60)	0–3
KPI Overall (SD)				2.51 (0.94)	0.13–3
<b>Session 2—Group B in lab, N = 5, 4 with previous Pepper experience</b>					
Perspicuity	Mean (SD)	Median	Importance (SD)	KPI (SD)	Range
Pepper	2.45 (0.59)	2.5	2 (0.63)	2.45 (0.48)	−1.75–3
Sensors	1.55 (1.28)	1.5	2.67 (0.49)	1.55 (0.66)	0.75–2.75
App	2.35 (1.35)	3	2.2 (0.75)	2.35 (1.3)	0–2.5
KPI Overall (SD)				2.1 (0.64)	0.83–2.54
SUS Overall (SD)				85 (4.18)	80–90
<b>Session 3—Group C in assisted living, N = 3, 3 with previous Pepper experience</b>					
Perspicuity	Mean (SD)	Median	Importance (SD)	KPI (SD)	Range
Pepper	1.33 (1.97)	2	1.33 (1.25)	1.33 (1.25)	1.25–3
Sensors	1.42 (1.61)	2	0.75 (1.41)	1.42 (1.05)	−0.25–3
App	1.92 (0.86)	2	0.33 (2.05)	1.92 (0.51)	1.25–2.5
KPI Overall (SD)				1.45 (0.94)	0.36–2.67
SUS Overall (SD)				67.5 (16.20)	45–82.5

#### 3.1. Session 1: User Group A in Residential Care

In session 1, we tested our system with user group A, which consisted of eight residents of a nursing home. Additionally, one of their caregivers also participated in the study in this session. The results in Table 1 include only the eight older participants, as they belong to the target group of older adults. Overall, the user experience of the health monitoring system was rated highly in session 1. The KPIs, with a possible value between −3 and 3, are all above 2, with the interaction with the robot Pepper being rated the highest (2.88) and the connected health devices rated the lowest (2.25), and the app in between at 2.71. This results in an overall system KPI of 2.51. In addition, the caregiver also gave high ratings, with an overall system KPI of 2.50, indicating high UX from the care perspective, also.

Regarding the use of health devices for measurements, we observed that some of the participants struggled to put on the wearable and the blood pressure monitor by themselves. All participants were able to correctly carry out the body temperature measurement, while some had problems with sending the measured value to the server. This was performed using a button on the thermometer. After another explanation and/or demonstration, most of the participants managed to both measure and send the measured value to the server correctly.

When using the data visualization app, some participants tapped the buttons in such a manner, that the app did not respond, e.g., due to tapping the button for too long. They, therefore, had to repeatedly tap the button for it to lead to its corresponding action or page. Overall, most of the participants were able to find the values we asked them to look for, i.e., their body temperature and blood pressure. One participant stated that they would prefer to use the app on a computer. This is also possible with this application, as it was developed using a multi-platform framework. Additionally, almost all participants also stated that they had a positive impression of both the devices and the app.

In the context of a health assessment, which is intended to relieve care professionals of documentation tasks, both data accuracy and data completeness are crucial. We evaluated data accuracy by calculating the success rate, which we define as the number of error-free interaction turns in relation to the total number of turns in a session. In this evaluation, we considered the actual health assessment, only, and excluded the introduction. We defined a turn as the complete sequence of a robot question, the user response, and the subsequent robot feedback.

In session 1, over the course of all assessments of the eight older users, we recorded a total of 148 turns, of which 114 were error-free. This results in an overall success rate of 77.03%. The number of total turns includes 24 items that were queried more than once in the same assessment due to faulty system routing. In three cases, an assessment had to be aborted and restarted. In each of these cases, the participants had directed questions at the research team. The system logs show that these off-topic remarks were picked up by the system, interpreted wrongfully, and caused faulty routing. When the evaluation is cleansed of these aborted assessments, a success rate of 80.01% is calculated from a total of 136 turns and 109 error-free turns. Furthermore, the logs show that participants using non-valid answer options is the most common source of error in session 1. This occurred a total of 15 times and is thus the cause for approximately 44% of the off-script events. We categorized the detected off-script events, and present them along with their absolute frequency, in Table 2.

**Table 2.** The error sources of the robot-assisted health assessments are shown.

Error Categories	Session 1 w/ Aborted Tries	Session 1 w/out Aborted Tries	Session 2	Session 3
No. of questions repeated (for whatever reason)	24	22	0	0
Overall off-script-events	34	27	23	19
Not a valid answer option	15	11	7	3
Off-topic remark/remark directed at researcher	8	6	0	2
System error (correct input matched wrong)	4	4	1	0
User uttering spoken too softly/no speech recognized	3	3	7	10
STT error (system understood correct input wrong)	3	2	2	4
Training data error	1	1	6	0

In terms of the completion rate, an assessment is considered complete, if the required data on all items were collected. Based on the system logs, we found that four of the eight valid assessments in session 1 were completed, resulting in a completion rate of 50.0%. The success and completion rates for all sessions are listed in Table 3. As five of the eight older participants in session 1 had previous experience with the robot Pepper, we also calculated the success and completion rates separately for the experienced (group A1) and the inexperienced (group A2) users, as listed in Table 4. Group A1 achieves 79 successful turns, in relation to a total of 95 turns, resulting in a success rate of 83.16%. Out of the five interactions conducted, three were completed in full, thus bringing the completion rate to 60.00%. Group A2, the users without any previous experience with the robot score a lower success rate of 66.04% from a total of 53 turns with 35 successful ones. One out of these three interactions was completed, resulting in a completion rate of 33.33%. Additionally, the caregiver—not included in this calculation as they are outside of the user

group of older adults—achieved a success rate of 93.75% and were able to complete the health assessment.

**Table 3.** Success and completion rates of all sessions are shown.

	Session 1 w/ Aborted Attempts	Session 1 w/out Aborted Attempts	Session 2	Session 3
Total turns taken (incl. aborted attempts)	148	136	167	103
Successful turns (incl. aborted attempts)	114	109	144	84
<b>Success rate</b>	<b>77.03%</b>	<b>80.01%</b>	<b>86.23%</b>	<b>81.55%</b>
No. of aborted queries	3	-	0	0
Total no. of queries (w/out aborted attempts)		8	5	3
Total no. of completed queries		4	5	3
<b>Completion rate</b>		<b>50%</b>	<b>100%</b>	<b>100%</b>

**Table 4.** The comparison of the success and completion rates of the session 1 robot interactions of the participants with and without prior experience with Pepper are shown.

	Session 1 Participants with Experience (N = 5)	Session 1 Participants without Experience (N = 3)
Total turns (w/aborted attempts)	95	53
Successful turns	79	35
Success rate	83.16%	66.04%
Completion rate	60.00%	33.33%

### 3.1.1. Evaluation 1: Conclusions

Overall in session 1, user experience was rated highly. Some participants struggled with the use of the devices and the handling of the tablet when using the app. Additionally, while the success rate of the interactions with the social robot was 77.03%, the completion rate of the health assessment was only 50.00%. We, therefore, introduced the following optimizations into the system, hoping to increase the completion rate in the subsequent test sessions.

### 3.1.2. Optimizations

Following an iterative development model in the UX design process, we implemented the optimizations before expert users were confronted with the system in session 2. The main factor influencing the completion rate of the health assessment was the number of overall off-script-events at 34, which led to wrong turns taken in the conversation with the robot, and then subsequent abortions to the interaction. To improve this, we implemented the health assessment questionnaire using the Rasa feature forms, as well as custom actions, for test sessions 2 and 3. If during an interaction with the robot Pepper, the user's verbal input can not be recognized or matched with an intent, a question mark is displayed on the robot's tablet, to prompt the user to repeat and/or rephrase their input. As part of the optimization, we added this information to the introduction of the health assessment. Furthermore, we observed that participants frequently gave premature voice commands, while the robot was transitioning from the introduction to the assessment. As a result, only a part of the input was captured by the robot, leading to false routing between items. To prevent this and improve the success rate of the interactions, we implemented a robot prompt that asked the user to wait two seconds while the robot switches to the health assessment. In session 1 we included only 14 questions plus one conditional question of the health assessment questions in order to first test a simplified version of the assessment. After session 1, we implemented the full 31 questions for the subsequent tests, which are listed in Appendix A.

We also made several design adjustments to the mobile app. These include an introduction with explanations of how to use the app and visual support to use the app

in both portrait and landscape mode. Finally, we implemented stronger haptic feedback when tapping buttons to increase the awareness of users if they had actually tapped the button correctly.

### 3.2. Session 2: User Group B in Lab Tests

In session 2, we conducted lab tests with five expert users who have previous experience not only with using but also developing chatbots or robots. The user experience was also rated highly for all system components by this group. The KPI for the health assessment (2.45) and the app (2.35) were rated higher than for the sensors (1.55), resulting in an overall system KPI of 2.1. The system usability, evaluated with the SUS, was rated highly at 85, indicating very good usability at a maximum possible value of 100. Additionally, based on our observations the participants had only very minor difficulties using the health devices, and all of them were able to find their measured body temperature and blood pressure in the app. Some of the participants also stated that the app had a clear layout, and was intuitive to use.

In session 2, we recorded a total of 167 turns over the course of five assessments. Based on logs and transcripts, 144 turns can be considered error-free, resulting in a success rate of 86.2%. The most common reasons for errors were (1) participants deviating from the predefined answer options and (2) the system registering no speech input, each occurring seven times over the course of this session. There was no need to abort any assessments, on the contrary, each of the five assessments conducted was completed in full, so the completion rate was 100%. A number of participants stated that the health assessment was too long and sometimes lacked variety in terms of the robot's replies. We also observed that some of the participants tried to test the robot, by giving more complex answers than the predefined answers. When this approach was unsuccessful, they went back to using short answers. Additionally, in session 2, we identified several errors in the training data, which we subsequently resolved before session 3.

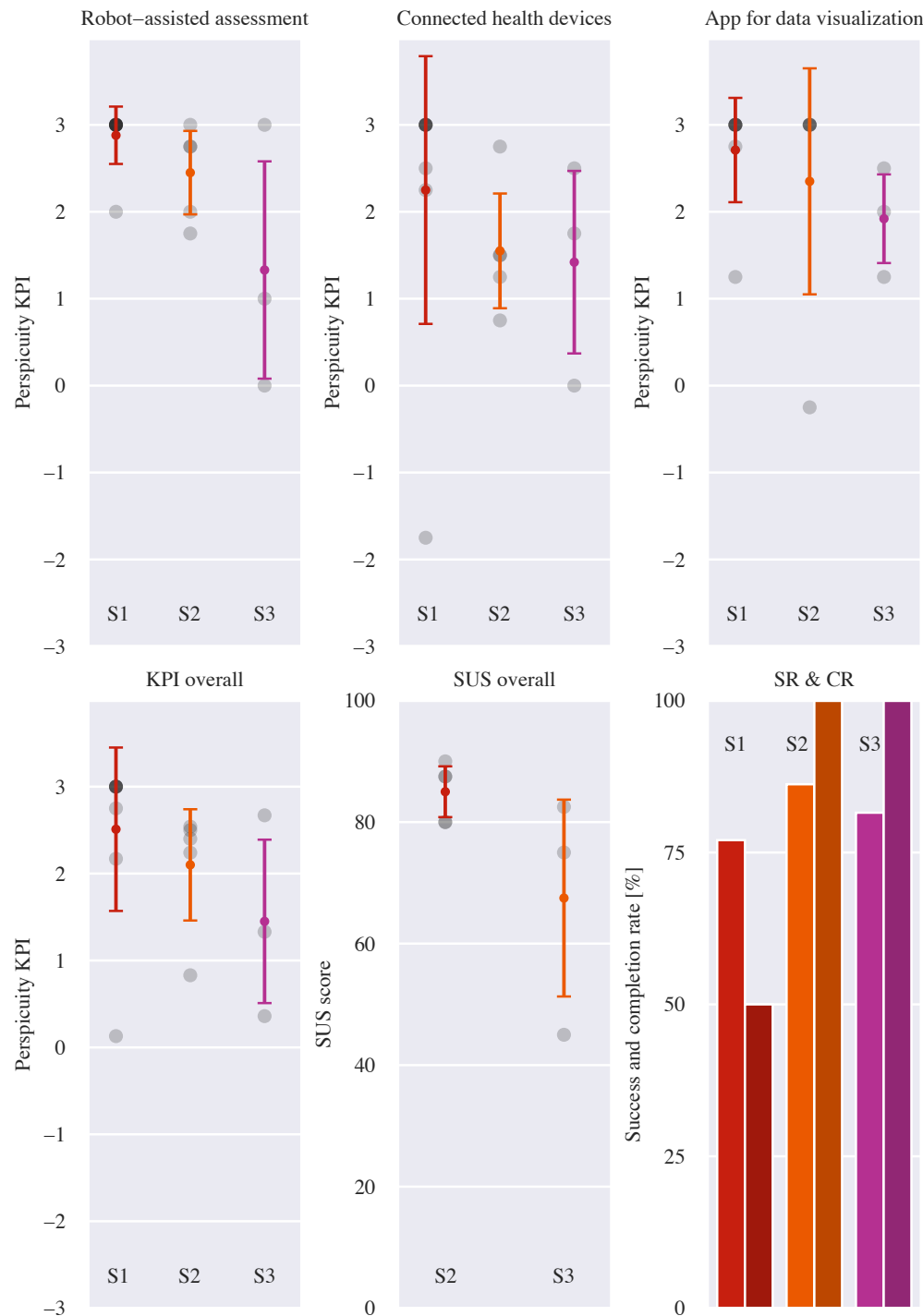
### 3.3. Session 3: User Group C in Assisted Living

In session 3, we tested our system with user group C, who are residents of assisted living communities and have extensive previous experience with the social robot Pepper, as one lives in their community. Again, the KPI of the robot (1.33), the devices (1.42), and the app (1.92) were rated highly, albeit lower than in sessions 1 and 2. The app was rated best, with the devices rated only slightly higher than the robot. The overall system KPI is rated well at 1.45, while the overall system usability is rated at 67.5 of 100, indicating average usability. Regarding the health devices, the participants had difficulties on the first try of the measurements. However, when attempting a second measurement, all participants were able to conduct the measurements by themselves. Some of the participants also mentioned their interest in using the devices and liked their modern design. When using the mobile app, we observed that the participants were unsure which category to find the measured values in. However, after finding the first value we asked them to find (body temperature), they were quick to also find the other values.

Session 3 was conducted with three participants so that a lower number of 103 turns was recorded. In relation to the number of 84 error-free turns, a success rate of 81.6% results. Based on transcripts and system logs, we identified the most common cause of error to be deviating from the ones of the previous sessions. In session 3, the system not recognizing speech input is the most prevalent issue, occurring 10 times and thus causing over 52% of the errors. The main consequence is that users have to repeat themselves. However, two turns were identified, in which the system did not register any user input in the logs, yet triggered a robot uttering. In session 3, again all assessments were completed in full, resulting in a completion rate of 100%. Additionally, we occasionally observed slow responses in the robot, which can be indicative of a slow internet connection, and also has an impact on how well or fast the robot's speech recognition functions.

### 3.4. Overall Results

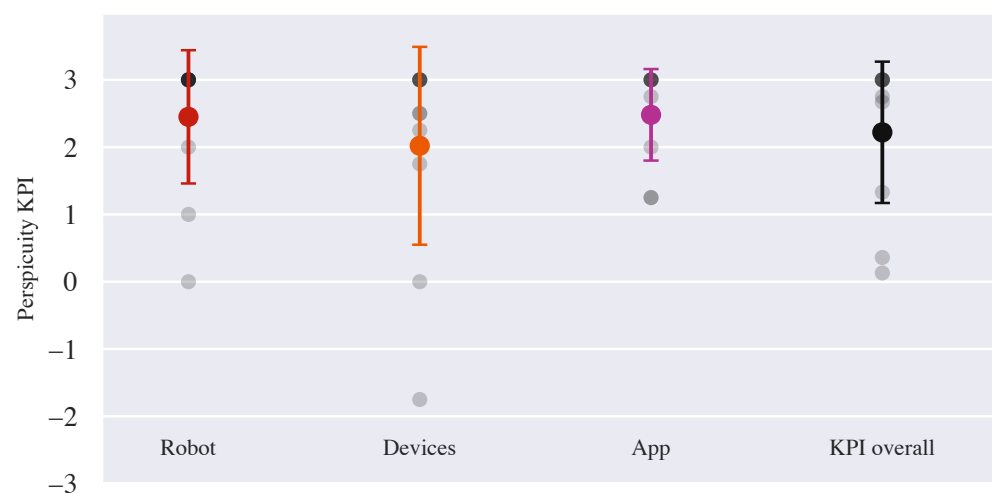
The overall results of the study are visualized in Figure 6. The participants of session 1 rated the user experience of each individual component and the overall system the highest. The UX evaluation of session 2 was lower, with the lowest values for session 3. Especially the rating of the robot-assisted health assessment was noticeably lower in session 3 than in sessions 1 and 2. Additionally, the system usability was rated as high in session 2 and average in session 3.



**Figure 6.** The overall results of the study, including the individual values, means, and standard deviations for the KPIs of the individual components and the whole system, the system usability score, and the success (SR) and completion rates (CR), are visualized.

Regarding the success rates of the conversation turns in the interaction with the social robot for the health assessment, small differences can be seen. The success rates in sessions 1 and 3 are similar at around 80%. The success rate in session 2 with the expert users is slightly higher at 86.23%. Large differences can be seen in the completion rates of the health assessment after the first session. While only five of nine assessments could be completed in session 1, all assessments were completed in sessions 2 and 3. Overall, the results indicate good user experience and usability of our health monitoring system.

Additionally, the combined results of sessions 1 and 3 are shown in Figure 7. Minor changes, mostly of technical nature, were introduced into the system after session 1, as described in Section 2.4. Both sessions 1 and 3 were conducted with the target group of older adults; however, giving good indications of the rated UX of the system for this group. The combined KPIs of all system components, as well as the overall system, are rated very highly in sessions 1 and 3, with all values above 2.



**Figure 7.** The combined KPI results (N = 11) of the sessions with older adults, namely session 1 in residential care (N = 8) and session 3 in assisted living (N=3), are visualized. All combined KPIs are rated highly with values above 2, indicating an excellent UX of our health monitoring system among older adults.

#### 4. Discussion

The overall UX of our health monitoring system was rated highly, both by older adults and younger expert users. However, there were differences in the ratings between the three sessions, which are discussed in the following.

##### 4.1. Influence of Prior Experience with Robots

The rating of the system declined with each of the test sessions. This difference was the most pronounced with the robot-assisted health assessment. A possible explanation for this result is the novelty effect [19], as described in Section 1.1. The participants in session 3 regularly interact with a Pepper robot and are therefore much more experienced with the robot than the participants in the other sessions. The novelty of the robot may have worn off, leading to less excitement in interacting with it, thus also resulting in slightly lower UX evaluations. It needs to be considered, however, that session 3 had only three participants. The results are, therefore, not statistically significant, but rather exploratory in nature. Additionally, the KPI were very high in session 1, where the participants had greater difficulties with the devices and the completion rates of the health assessment were much lower. This also supports the explanation that the novelty effect may have led to more favorable results in sessions 1 and 2.

In session 1, a total of five older users with previous experience with the robot Pepper and three without tested the system. The overall success rate of participants with experience

is at 83.2% considerably higher than the 66.0% achieved by inexperienced users. The difference between these user groups becomes even more apparent when considering the completion rates. Three out of five assessments were successfully completed by experienced users, while the same applies to only one assessment with an inexperienced user. It must be considered, that the users from session 1 have only interacted with the robot on very few occasions, therefore, the experience can not be seen as extensive. With the success rate increasing noticeably with only little prior experience, we conclude that it is easy for older adults to learn how to interact with the robot.

Out of the participants in this study, group C had by far the most previous experience with the robot Pepper. Nevertheless, the success rate of session 3 differs only slightly from that in session 1—especially when considering the success rate, cleansed from the failed attempts. On the testing day, however, we repeatedly observed difficulties and delays in the robot's speech recognition, possibly caused by a slow internet connection. This is also reflected in the comparison of the system logs with the transcripts. The speech recognition seemed to fail more frequently, resulting in users having to repeat their input, false routing, and a reduced success rate. It is therefore likely that the success rate in session 3 would have been higher with a better connection. Finally, despite the described internet issues, the success and completion rates for session 3 with older adults were only slightly lower than in session 2 with younger expert users. This again indicates that older adults, especially those with prior experience with the robot, perform well when interacting with the robot.

#### *4.2. Influence of Care Setting*

Another factor to be taken into account leading to possible differences between sessions 1 and 3 are the different care settings. Both user groups include older adults. Session 1 was conducted in a residential care facility with users with overall more significant health impairments. Therefore, for the regular checks of the health status of many of these users, the assistance of caregivers is required. Using our health monitoring system may promote the independence of this user group, as they do not have to wait for the caregiver and can work with the system on their own time. This may have an effect on the perceived usefulness and thus the reported UX of this user group.

Session 3 was conducted in an assisted living facility where participants are not in need of full-time care. The participants in this facility also tend to lead more active social lives and are more mobile, which may also influence the overall perception of the need for such a health monitoring system, as opposed to participants in full-time care. This is an important factor that needs to be further investigated in future work.

#### *4.3. Influence of Technological Expertise*

Testing the system with younger users in session 2, the health assessment achieved its highest success rate at 86.23%. The participants all have a technical career orientation and can thus be described as tech-savvy, and four out of five had previous experience with the robot Pepper. Out of twenty-three off-script-events recorded in this session, seven pertained to participants responding with something other than the displayed answer options. It should be noted that some of the participants stated afterwards that they intentionally deviated from the answer options, to test the limits of the dialogue system. Hence, the actual success rate might have been in fact higher than the calculated one. Overall, the high UX and SUS ratings indicate a very good UX and usability for this group of expert users.

The caregiver's results in terms of UX were quite similar to the other participants' results from session 1. However, as a younger and more tech-savvy person, with additional experience with the robot Pepper, they were excluded from our calculations for session 1 and viewed separately, instead. Additionally, they achieved a high success rate of 93.75% in their interaction. Furthermore, the system usability scores were lower in session 3 than in session 2. A factor that may have influenced this is that the participants in session 2 had fewer difficulties using the health devices and the tablet app, as they were a younger user group with professional backgrounds in technology and engineering fields. Overall,



this indicates that prior technological experience, which is often correlated with the age or respective generation of the user, has an effect on the UX of our system.

#### 4.4. Influence of Optimizations

Aside from participants using answer options deviating from the predefined ones, one main cause of error in session 1 was off-topic-remarks. A total of eight times, participants directed questions about the interaction with the robot at the research team. This led us to the task of further optimizing the health assessment after session 1, by means of participatory methods, so it can be used more intuitively by members of the target group. When comparing the completion rates over all sessions, it becomes apparent that incomplete health assessments occurred in session 1—before the optimizations—only. The same is valid for the unwanted repetition of individual items or even complete parts of the assessment. This indicates that optimizing the dialogue system, by using Rasa forms with custom actions, was a successful approach to avoiding false routing between items as well as to ensuring data completion.

After the introduction of the optimizations between sessions 1 and 2, the UX rating was lower in sessions 2 and 3. We assume, however, that this is related to the influence of the prior experience with the robot, the care setting, and the technological expertise, rather than the optimizations themselves, as they had the intended effect of higher success and completion rates for the robot interactions.

#### 4.5. Limitations

While we assume that our study provides promising insights about the UX of our system, it does have several limitations, which are discussed in the following. The younger participants of session 2 completed the questionnaires mostly by themselves. Only one question was raised about the interpretation of an SUS item; a research team member answered it neutrally. However, all participants of sessions 1 and 3 received help with the questionnaires. Care was taken to use neutral language when explaining questions and answer options, so as not to create a bias among the participants. Nevertheless, this approach may have had an impact on the results—for example, as a result of sympathy or antipathy towards the researchers. In addition, social desirability bias might have played a role. Bogner and Landrock describe in [46] that participants often show a tendency towards better ratings, as they assume these are desired by the interviewer. Additionally, in session 3, we noted that completing the SUS items of our questionnaire seemed challenging to the participants. For several items, the participants seemed unsure how to answer and required an explanation. While completing the SUS items, one participant even commented on their tendency to select the neutral option.

After session 1, we optimized the system and implemented various changes, which were only tested with experienced chatbot and/or robot users in sessions 2 and 3. In the future, we will also test the optimized version with users who have none or only a little experience with robots. Additionally, the internet connection in session 3 may have had an impact on the reported UX of the participants, as it led to difficulties and delays in the robot's speech recognition. In the future, we will conduct this test again with a user group with extensive Pepper experience and an optimally working internet connection.

In this study, we conducted single tests, so each participant used the system only once. Such a limited test duration does not allow any conclusions to be drawn about the long-term use of the system. Furthermore, sessions 1 and 2 were conducted with a number of participants generally deemed as adequate for testing UX [44]. However, even when taking into account the reduced population the samples were drawn from, the sample sizes are too small for valid statistical analysis. Nevertheless, we chose to use validated UX questionnaires, as we felt that the questionnaires explore our target figures, i.e., UX and usability, in full. While we are aware that our results are not statistically significant, they provide a promising indication of our system's UX. We also included only one caregiver in session 1, and the user group in session 2 consisted of members of our lab as a convenience

sample to confirm that our system works correctly and to evaluate the usability of our system among expert users. Additionally, due to the direct effects of the current pandemic situation, we were able to include three participants in session 3 only. While this number of test users and the inclusion of a convenience sample cannot provide significant and transferable results, the feedback from all user groups was nevertheless very valuable as a first indication and resulting implications for future work.

## 5. Conclusions and Future Work

In this work, we evaluated the UX and usability of our health monitoring system and achieved high results in both. The ratings were the highest in the users with the least robotic experience, while they also had the most difficulties in using the system, indicating that the novelty effect may have positively influenced the UX results. The ratings were also high in the experienced and tech-savvy group of younger expert users, while the older adults with the most experience with Pepper specifically were the most critical in their responses. Based on these results, we draw the conclusion that previous robotic experience and technological expertise play an important role in the evaluation of the UX of robot-assisted health assessments. The care setting of the older adults may also play a role in this evaluation, though these needs to be further investigated. All of these factors provide important indications as to what needs to be considered when developing, but also evaluating, health monitoring systems that provide a positive UX, specifically for older adults.

In the future, the system should be able to respond more individually to the user's condition. Here, the rather generic assessment we tested in this study will be used to determine current health issues and warning signs. As users are intended to undergo the assessment on a daily basis, we will implement different versions of our health questionnaire to improve the UX. If a condition can be determined based on the user's responses to our questionnaire, e.g., signs of hypertension, the robot will then conduct a condition-specific, medically validated PROM assessment.

Subsequently, we will test the system in a long-term study. Thus, the influence of the novelty effect, if any, will become visible, while the continuous use of the system will contribute to the broadening of the health database available to us. While conducting this long-term study we will also continuously optimize the system based on user feedback, to guarantee a positive UX and thus promote the long-term motivation of older adults to use this system, which can then contribute to their autonomy and independence by increasing their health literacy while supporting them in managing their own health. In the short term, the system should then be able to classify a user's health status based on a combination of the subjective and objective data collected. Finally, a long-term goal of the project is the use of machine learning to predict a user's health status and provide adequate advice and recommendations to the users.

**Author Contributions:** Conceptualization, C.N.; methodology, C.N. and K.L.; software, C.N. and K.L.; validation, C.N. and K.L.; formal analysis, C.N. and K.L.; investigation, C.N. and K.L.; resources, A.R.; data curation, C.N. and K.L.; writing—original draft preparation, C.N. and K.L.; writing—review and editing, C.N., K.L. and A.R.; visualization, C.N.; supervision, A.R.; project administration, C.N. and A.R. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was partially supported by the German Federal Ministry of Education and Research (BMBF) in the project GeneRobot (project number 16GDI102A).

**Institutional Review Board Statement:** The study was reviewed and approved by the Ethics Research Committee of the TH Köln—University of Applied Sciences (application no. THK-2022-0004).

**Informed Consent Statement:** Informed consent was obtained from all subjects involved in the study.

**Data Availability Statement:** Not applicable.

**Acknowledgments:** The authors would like to thank Marco Voigt, Katinka Rosenfeld, Leon Munz, Nathalie Weßels, and all participants for their valuable input to this work.

**Conflicts of Interest:** The authors declare no conflict of interest.

### Abbreviations

The following abbreviations are used in this manuscript:

AI	artificial intelligence
API	application programming interface
app	application
EKG	electrocardiogram
KPI	key performance indicator
PROMs	patient-reported outcome measures
UX	user experience
UEQ	User Experience Questionnaire
UEQ+	modular extension of the User Experience Questionnaire
UI	user interface
SAS	Smiley-Analogue-Scale
SUS	System Usability Scale

### Appendix A. Health Assessment Questionnaire

No.	Session	Question	Possible Answer Options
1	S123	On a scale of 0 to 10: How would you rate your health today? (0 = very poor/10 = very healthy)	numbers between 0 and 10
2	S123	Are you experiencing any pain right now?	yes/no
opt.	S23	Where is your pain located at the moment?	everywhere, back, neck, legs, knee, head, ...
opt.	S123	How severe is your pain?	mild/moderate/strong/very strong
3	S123	Do you have or have you had a headache today?	yes/no
4	S123	Are you feeling or have you felt dizzy today?	yes/no
5	S123	How fit are you feeling today?	absolutely fit/rather fit/somewhat tired/very tired
6	S23	Have you noticed any redness of your face today?	yes/sometimes/no
7	S23	Do you have or have you had intermittent swelling of your hands or feet today?	yes/sometimes/no
8	S23	Do you have or have you had heart palpitations today?	yes/no
9	S123	Are you feeling stressed or tense today?	yes, slightly, not at all
10	S123	Are you feeling sad or down today?	yes, slightly, not at all
11	S123	When you think about this past week, how often did you feel sad or down?	always/usually/often/rarely/never
12	S23	Are you feeling nauseous today?	yes/no
13	S23	Are you currently experiencing pressure in your chest?	yes/no
14	S23	Do you feel as if you had a fever today?	yes/no
15	S23	Do you have or have you had the chills today?	yes/no
16	S123	Did you eat with a good appetite today, respectively yesterday?	yes/no
17	S23	Have you already had a serving of fruit or vegetables today?	yes/no
18	S123	Did you already drink some water today?	yes/no
19	S23	Are you currently experiencing a dry mouth or have chapped lips?	yes/no
20	S123	On a scale of 0 to 10: How would you rate your sleep last night? (0 = very bad/10 = very good)	numbers between 0 and 10
21	S123	Did you feel well-rested when you got up this morning?	yes/no
22	S123	Do you have difficulties with your memory?	yes/no
23	S23	How would you rate your mood today?	balanced/somewhat tense/rather irritated/unbalanced
24	S23	How often did you get up last night?	not at all/1–2 times/3–4 times/more often
25	S23	Were you able to go straight to sleep last night?	yes/took me a little while/was still awake for a long time
26	S123	Have you already taken a walk today, or do you still plan on going outside?	yes/maybe/no
27	S23	Have you done any physical activity already today, or do you still plan to do so?	yes/maybe/no
28	S23	Can you go up and down the stairs fully independently?	yes/no
29	S23	Do you regularly take the stairs instead of the elevator?	yes/no
30	S23	Have you already been in contact with your family or friends today? Even if only by phone?	yes/I still plan on doing so/I might do it (later)/no
31	S23	Do you have any plans on meeting your family or friends today?	yes/not yet/no

## References

1. Smith, R.O.; Scherer, M.J.; Cooper, R.; Bell, D.; Hobbs, D.A.; Pettersson, C.; Seymour, N.; Borg, J.; Johnson, M.J.; Lane, J.P.; et al. Assistive Technology Products: A Position Paper from the First Global Research, Innovation, and Education on Assistive Technology (GREAT) Summit. *Disabil. Rehabil. Assist. Technol.* **2018**, *13*, 473–485. [CrossRef] [PubMed]
2. Maresova, P.; Režný, L.; Bauer, P.; Fadeyi, O.; Eniyewu, O.; Barakovic, S.; Barakovic Husic, J. An Effectiveness and Cost-Estimation Model for Deploying Assistive Technology Solutions in Elderly Care. *Int. J. Healthc. Manag.* **2022**, 1–16. [CrossRef]
3. Dobner, S.; Musterd, S.; Droogleever Fortuijn, J. ‘Ageing in Place’: Experiences of Older Adults in Amsterdam and Portland. *GeoJournal* **2016**, *81*, 197–209. [CrossRef]
4. Ollevier, A.; Aguiar, G.; Palomino, M.; Simpelaere, I.S. How Can Technology Support Ageing in Place in Healthy Older Adults? A Systematic Review. *Public Health Rev.* **2020**, *41*, 26. [CrossRef]
5. Wildenbos, G.A.; Peute, L.; Jaspers, M. Aging Barriers Influencing Mobile Health Usability for Older Adults: A Literature Based Framework (MOLD-US). *Int. J. Med. Inform.* **2018**, *114*, 66–75. [CrossRef]
6. Neef, C.; Schneider, S.; Richert, A. An Architecture for Social Robot-assisted Subjective and Objective Health Monitoring. In Proceedings of the 2022 IEEE International Conference on Advanced Robotics and Its Social Impacts (ARSO), Long Beach, CA, USA, 28–30 May 2022; pp. 1–6. [CrossRef]
7. Durán-Vega, L.A.; Santana-Mancilla, P.C.; Buenrostro-Mariscal, R.; Contreras-Castillo, J.; Anido-Rifón, L.E.; García-Ruiz, M.A.; Montesinos-López, O.A.; Estrada-González, F. An IoT System for Remote Health Monitoring in Elderly Adults through a Wearable Device and Mobile Application. *Geriatrics* **2019**, *4*, 34. [CrossRef]
8. Dratsiou, I.; Varella, A.; Romanopoulou, E.; Villacañas, O.; Cooper, S.; Isaris, P.; Serras, M.; Unzueta, L.; Silva, T.; Zurkühlen, A.; et al. Assistive Technologies for Supporting the Wellbeing of Older Adults. *Technologies* **2022**, *10*, 8. [CrossRef]
9. Wu, M. Wearable Technology Applications in Healthcare: A Literature Review. *Online J. Nurs. Inform.* **2019**, *23*. Available online: <https://www.proquest.com/docview/2621329056/abstract/1AFDA83C2684939PQ/1> (accessed on 27 November 2022).
10. Moore, K.; O’Shea, E.; Kenny, L.; Barton, J.; Tedesco, S.; Sica, M.; Crowe, C.; Alamäki, A.; Condell, J.; Nordström, A.; et al. Older Adults’ Experiences With Using Wearable Devices: Qualitative Systematic Review and Meta-synthesis. *JMIR mHealth uHealth* **2021**, *9*, e23832. [CrossRef]
11. Looije, R.; Neerinx, M.A.; Cnossen, F. Persuasive Robotic Assistant for Health Self-Management of Older Adults: Design and Evaluation of Social Behaviors. *Int. J.-Hum.-Comput. Stud.* **2010**, *68*, 386–397. [CrossRef]
12. Pandey, A.K.; Gelin, R. A Mass-Produced Sociable Humanoid Robot: Pepper: The First Machine of Its Kind. *IEEE Robot. Autom. Mag.* **2018**, *25*, 40–48. [CrossRef]
13. Paletta, L.; Schüssler, S.; Zuschnegg, J.; Steiner, J.; Pansy-Resch, S.; Lammer, L.; Prodromou, D.; Brunsch, S.; Lodron, G.; Fellner, M. AMIGO—A Socially Assistive Robot for Coaching Multimodal Training of Persons with Dementia. In *Social Robots: Technological, Societal and Ethical Aspects of Human-Robot Interaction*; Korn, O., Ed.; Springer International Publishing: Cham, Switzerland, 2019; pp. 265–284. [CrossRef]
14. Fattal, C.; Cossin, I.; Pain, F.; Haize, E.; Marissael, C.; Schmutz, S.; Ocnareescu, I. Perspectives on Usability and Accessibility of an Autonomous Humanoid Robot Living with Elderly People. *Disabil. Rehabil. Assist. Technol.* **2022**, *17*, 418–430. [CrossRef]
15. Robinson, N.L.; Cottier, T.V.; Kavanagh, D.J. Psychosocial Health Interventions by Social Robots: Systematic Review of Randomized Controlled Trials. *J. Med Internet Res.* **2019**, *21*, e13203. [CrossRef]
16. Piasek, J.; Wieczorowska-Tobis, K. Acceptance and Long-Term Use of a Social Robot by Elderly Users in a Domestic Environment. In Proceedings of the 2018 11th International Conference on Human System Interaction (HSI), Gdansk, Poland, 4–6 July 2018; pp. 478–482. [CrossRef]
17. Mucchiani, C.; Sharma, S.; Johnson, M.; Sefcik, J.; Vivio, N.; Huang, J.; Cacchione, P.; Johnson, M.; Rai, R.; Canoso, A.; et al. Evaluating Older Adults’ Interaction with a Mobile Assistive Robot. In Proceedings of the 2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Vancouver, BC, Canada, 24–28 September 2017; pp. 840–847. [CrossRef]
18. Biswas, M.; Romeo, M.; Cangelosi, A.; Jones, R.B. Are Older People Any Different from Younger People in the Way They Want to Interact with Robots? Scenario Based Survey. *J. Multimodal User Interfaces* **2020**, *14*, 61–72. [CrossRef]
19. Ostrowski, A.K.; Breazeal, C.; Park, H.W. Mixed-Method Long-Term Robot Usage: Older Adults’ Lived Experience of Social Robots. In Proceedings of the 2022 17th ACM/IEEE International Conference on Human-Robot Interaction (HRI), Sapporo, Japan, 7–10 March 2022; pp. 33–42. [CrossRef]
20. van der Putte, D.; Boumans, R.; Neerinx, M.; Rikkert, M.O.; de Mul, M. A Social Robot for Autonomous Health Data Acquisition Among Hospitalized Patients: An Exploratory Field Study. In Proceedings of the 2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI), Daegu, Republic of Korea, 11–14 March 2019; pp. 658–659. [CrossRef]
21. Boumans, R.; van Meulen, F.; van Aalst, W.; Albers, J.; Janssen, M.; Peters-Kop, M.; Huisman-de Waal, G.; van de Poll, A.; Hindriks, K.; Neerinx, M.; et al. Quality of Care Perceived by Older Patients and Caregivers in Integrated Care Pathways With Interviewing Assistance From a Social Robot: Noninferiority Randomized Controlled Trial. *J. Med. Internet Res.* **2020**, *22*, e18787. [CrossRef]
22. Mann, J.A.; MacDonald, B.A.; Kuo, I.H.; Li, X.; Broadbent, E. People Respond Better to Robots than Computer Tablets Delivering Healthcare Instructions. *Comput. Hum. Behav.* **2015**, *43*, 112–117. [CrossRef]
23. Manh Do, H.; Sheng, W.; Harrington, E.E.; Bishop, A.J. Clinical Screening Interview Using a Social Robot for Geriatric Care. *IEEE Trans. Autom. Sci. Eng.* **2021**, *18*, 1229–1242. [CrossRef]

24. van Greunen, D. User Experience for Social Human-Robot Interactions. In Proceedings of the 2019 Amity International Conference on Artificial Intelligence (AICAI), Dubai, United Arab Emirates, 4–6 February 2019; pp. 32–36. [\[CrossRef\]](#)
25. Phiriyapokanon, T. Is a Big Button Interface enough for Elderly Users? Towards User Interface Guidelines for Elderly Users. Master's Thesis, Mälardalen University, Västerås, Sweden, 2011.
26. Shourmasti, E.S.; Colomo-Palacios, R.; Holone, H.; Demi, S. User Experience in Social Robots. *Sensors* **2021**, *21*, 5052. [\[CrossRef\]](#)
27. Laugwitz, B.; Held, T.; Schrepp, M. Construction and Evaluation of a User Experience Questionnaire. In Proceedings of the HCI and Usability for Education and Work, Graz, Austria, 20–21 November 2008; Holzinger, A., Ed.; Springer: Berlin/Heidelberg, Heidelberg, 2008; pp. 63–76. [\[CrossRef\]](#)
28. Gerłowska, J.; Skrobas, U.; Grabowska-Aleksandrowicz, K.; Korchut, A.; Szklener, S.; Szczeńniak-Stańczyk, D.; Tzovaras, D.; Rejdak, K. Assessment of Perceived Attractiveness, Usability, and Societal Impact of a Multimodal Robotic Assistant for Aging Patients With Memory Impairments. *Front. Neurol.* **2018**, *9*, 392. [\[CrossRef\]](#)
29. Schrepp, M.; Thomaschewski, J. Design and Validation of a Framework for the Creation of User Experience Questionnaires. *Int. J. Interact. Multimed. Artif. Intell.* **2019**, *5*, 88–95. [\[CrossRef\]](#)
30. Brooke, J. SUS: A 'Quick and Dirty' Usability Scale. In *Usability Evaluation In Industry*; CRC Press: Boca Raton, FL, USA, 1996.
31. Grier, R.A.; Bangor, A.; Kortum, P.; Peres, S.C. The System Usability Scale: Beyond Standard Usability Testing. *Proc. Hum. Factors Ergon. Soc. Annu. Meet.* **2013**, *57*, 187–191. [\[CrossRef\]](#)
32. Borsci, S.; Federici, S.; Lauriola, M. On the Dimensionality of the System Usability Scale: A Test of Alternative Measurement Models. *Cogn. Process.* **2009**, *10*, 193–197. [\[CrossRef\]](#) [\[PubMed\]](#)
33. Bevilacqua, R.; Felici, E.; Marcellini, F.; Glende, S.; Klemcke, S.; Conrad, I.; Esposito, R.; Cavallo, F.; Dario, P. Robot-Era Project: Preliminary Results on the System Usability. In Proceedings of the Design, User Experience, and Usability: Interactive Experience Design, Los Angeles, CA, USA, 2–7 August 2015; Marcus, A., Ed.; Springer International Publishing: Cham, Switzerland, 2015; pp. 553–561. [\[CrossRef\]](#)
34. Di Nuovo, A.; Varrasi, S.; Conti, D.; Bamsforth, J.; Lucas, A.; Soranzo, A.; McNamara, J. Usability Evaluation of a Robotic System for Cognitive Testing. In Proceedings of the 2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI), Daegu, Republic of Korea, 11–14 March 2019; pp. 588–589. [\[CrossRef\]](#)
35. Poberznik, A.; Merilampi, S. Older Adults' Experiences with Pepper Humanoid Robot. *Tutkimusfoorumi* **2019**, *13*, 148.
36. Normenausschuss, D. *DIN EN ISO 9241-110; Ergonomics of Human-System-Interaction—Part 110: Interaction Principles*. Beuth Verlag GmbH: Berlin, Germany, 2020.
37. Stolee, P.; Hillier, L.M.; Esbaugh, J.; Bol, N.; McKellar, L.; Gauthier, N. Instruments for the Assessment of Pain in Older Persons with Cognitive Impairment: pain assessment in persons with cognitive impairment. *J. Am. Geriatr. Soc.* **2005**, *53*, 319–326. [\[CrossRef\]](#)
38. Rasa. Introduction to Rasa Open Source & Rasa Pro. 2022. Available online: <https://rasa.com/docs/rasa/> (accessed on 2 December 2022).
39. Paraschiv, E.A.; Petrache, C.M.; Bica, O. On the Continuous Development of IoT in Big Data Era in the Context of Remote Healthcare Monitoring & Artificial Intelligence. In Proceedings of the 2022 14th International Conference on Electronics, Computers and Artificial Intelligence (ECAI), Ploiesti, Romania, 30 June–1 July 2022; pp. 1–6. [\[CrossRef\]](#)
40. Giggins, O.M.; Doyle, J.; Smith, S.; Crabtree, D.R.; Fraser, M. Measurement of Heart Rate Using the Withings ScanWatch Device During Free-living Activities: Validation Study. *JMIR Form. Res.* **2022**, *6*, e34280. [\[CrossRef\]](#)
41. Edouard, P.; Campo, D.; Bartet, P.; Yang, R.Y.; Bruyneel, M.; Roisman, G.; Escourrou, P. Validation of the Withings Sleep Analyzer, an under-the-Mattress Device for the Detection of Moderate-Severe Sleep Apnea Syndrome. *J. Clin. Sleep Med.* **2021**, *17*, 1217–1227. [\[CrossRef\]](#)
42. SA, W.F. Withings-API Developer Documentation. Available online: <https://developer.withings.com/api-reference/> (accessed on 27 November 2022).
43. SA, W.F. Public API Integration Guide. Available online: <https://developer.withings.com/developer-guide/v3/integration-guide/public-health-data-api/public-health-data-api-overview> (accessed on 28 November 2022).
44. Nielsen, J.; Landauer, T.K. A Mathematical Model of the Finding of Usability Problems. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems—CHI '93, Amsterdam, The Netherlands, 24–29 April 1993; ACM Press: Amsterdam, The Netherlands, 1993; pp. 206–213. [\[CrossRef\]](#)
45. Hinderks, A.; Schrepp, M.; Domínguez Mayo, F.J.; Escalona, M.J.; Thomaschewski, J. Developing a UX KPI Based on the User Experience Questionnaire. *Comput. Stand. Interfaces* **2019**, *65*, 38–44. [\[CrossRef\]](#)
46. Bogner, K.; Landrock, U. *Antworttendenzen in Standardisierten Umfragen; SDM Survey Guidelines; GESIS—Leibniz Institut für Sozialwissenschaften*: Mannheim, Germany, 2015; pp. 1–9. [\[CrossRef\]](#)

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.