



# Remote HRI: a Methodology for Maintaining COVID-19 Physical Distancing and Human Interaction Requirements in HRI Studies

Curtis L. Gittens<sup>1</sup>

Accepted: 16 June 2021

© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2021

## Abstract

Observing how humans and robots interact is an integral part of understanding how they can effectively coexist. This ability to undertake these observations was taken for granted before the COVID-19 pandemic restricted the possibilities of performing HRI study-based interactions. We explore the problem of how HRI research can occur in a setting where physical separation is the most reliable way of preventing disease transmission. We present the results of an exploratory experiment that suggests Remote-HRI (R-HRI) studies may be a viable alternative to traditional face-to-face HRI studies. An R-HRI study minimizes or eliminates in-person interaction between the experimenter and the participant and implements a new protocol for interacting with the robot to minimize physical contact. Our results showed that participants interacting with the robot remotely experienced a higher cognitive workload, which may be due to minor cultural and technical factors. Importantly, however, we also found that whether participants interacted with the robot in-person (but socially distanced) or remotely over a network, their experience, perception of, and attitude towards the robot were unaffected.

**Keywords** Remote-HRI · Remote interaction · Social robots · HRI · COVID-19 pandemic

## 1 Introduction

COVID-19 has disrupted the world since it was declared a pandemic by the World Health Organization (WHO) in March 2020, with governments enforcing strict physical lockdowns in their countries to limit the spread of the SARS-CoV-2 virus. These disruptions have inflicted economic, social, educational, and psychological costs across all strata of society. The global labor market lost 8.8% of its working hours in 2020 – approximately 225 million full-time jobs (International Labour Organization, 2021). Women, and more so minority women, have seen issues with pay inequity, reproductive healthcare and poverty exacerbated by the pandemic (United Nations, 2020). Children have lost on average 66% of the academic year due to COVID-19 school lockdowns (UNESCO, 2021), which negatively affect their health, well-being and their education.

Vaccinations are being distributed and administered globally and some countries are seeing their case numbers fall.

Unfortunately, a majority of countries are still experiencing significant community transmission, while new SARS-CoV-2 variants are being detected and monitored (Cabecinhas et al., 2021). These new variants are making it increasingly difficult to determine how long it will be before the world returns to normal because early evidence indicates that these variants reduce the efficacy of some of the existing vaccines and therapeutics (Wang et al., 2021).

COVID-19 has significant negative outcomes in the elderly, especially those with comorbidities (Killerby et al., 2020; Price-Haywood et al., 2020; Stokes et al., 2020). Unfortunately, as a simple Google Scholar search can show, most HRI research is based on these populations. Recognizing this problem, Feil-Seifer et al. (2020) propose nine research questions that must be addressed in order for human-centered research like HRI to proceed in the face of the pandemic.

One area they identify is the need to establish methodologies that would allow for realistic and ethical HRI research during the pandemic (Feil-Seifer et al., 2020). This is important since most of the work on assistive and companion robotics is conducted with the elderly and communities with existing health problems (Robinson et al., 2013; Martinez-Martin & del Pobil, 2018; Oh & Oh, 2019; Al-Tae et al., 2016; Blanson Henkemans et al., 2017; Schneider &

---

✉ Curtis L. Gittens  
curtis.gittens@cavehill.uwi.edu

<sup>1</sup> The University of the West Indies Cave Hill Campus, P.O. Box 64, Bridgetown, Barbados

Kummert, 2018; Barata, 2019), and it is these populations that are most at risk as the virus spreads (CDC, 2020).

Presently, studies where different participants need to touch the same robot is a challenge because common surfaces must be routinely sanitized, and constantly disinfecting a robot might potentially damage interfaces like its screen and sensors. HRI studies that involve at-risk populations where participants may suffer from any combination of mental, emotional, or physical impairment or disease, entering their living areas is still very limited or restricted - even where residents may be fully vaccinated. To address these pandemic-related problems, this paper proposes a methodology called Remote Human-Robot Interaction (R-HRI) where the participant is either: (a) physically with the robot but separated by no more than one or two meters or (b) the participant is in a physically different location than the robot. The public health measures as stipulated in (World Health Organization and others, 2020) for controlling viral transfer are met under this proposal. This includes reduction in contact with common surfaces, which in this case means not touching the robot, as well as reduction in congregating people from different households in a common indoor space.

## 2 Background

We will examine the two main methods that form the foundation for the R-HRI study: usability test and remote usability testing. These methods have been used effectively to assess user experience with software systems, including robots with remote usability being just as effective as in-person usability testing.

### 2.1 Usability Testing

The term *usability testing*, which we differentiate here by referring to it as *conventional usability testing*, is frequently broadly applied to techniques that test products, devices, websites and processes as in (Kim et al., 2005; Fang & Holsapple, 2011; Dörndorfer & Seel, 2020; Lee et al., 2020). This, however, is usually an incorrect application of the term. Usability testing, in the conventional sense, means gathering the target user-community of the program under test into a lab where a moderator watches them as they perform tasks to expose usability flaws (Rubin & Chisnell, 2008). Processes like system evaluations by experts, or system walkthroughs do not require representative users, and therefore do not qualify as usability testing.

Conventional usability testing uses classical experimental methods ranging from large formal experiments with complex test designs, to informal qualitative studies with one participant. The purpose and resources across these methodologies differ. Typically, participants, following conventional

usability testing protocol, are encouraged to speak aloud as they work. This helps the moderator to gain insight into each participant's thought processes as well as recognize any possible misunderstandings as they interact with the system (Rubin & Chisnell, 2008).

Although this approach has several benefits, one of which is high-quality quantitative data, it also has significant disadvantages. High costs, difficulty designing environments that match end users' environments, and, most significantly, finding a representative group of participants that can be in the lab are all impediments to a rigorous usability test. The fact that software is sold and marketed globally aggravates the last problem because a diverse pool of participants from around the world is a critical requirement - regardless of where the program is sold (Hartson et al., 1996).

### 2.2 The Promise of Remote Usability Testing

The rapid uptick in the accessibility and quality of software designed to promote knowledge sharing and community collaboration lead to Remote Usability Testing (RUT) becoming a viable alternative to conventional usability testing. Computer-based video, online collaborative tools and network technology matured enough to provide the technological foundation for remote testing facilities (Hammontree et al., 1994; Thompson et al., 2004). Instead of in-lab usability testing, interactive prototypes were tested remotely using traditional "thinking aloud" strategies with a combination of online logging, shared windowing tools and recording the verbal responses of the participant via the telephone. All modern video conferencing and online collaborative systems can be used to perform RUT. They easily provide screen sharing and other collaborative features out-of-the-box and facilitate the space and/or time separation requirement between evaluator and participant as defined in (Castillo et al., 1998).

The space/time separation requirement raises a number of interesting remote configuration possibilities that were evaluated in (Andreasen et al., 2007; Bruun et al., 2009). These configurations are referred to as *moderated/unmoderated* or *synchronous/asynchronous moderation*. The terms used to describe these configurations are interchangeable because they indicate whether the user or the evaluator is ultimately reporting on the usability issues. *Moderated* and *synchronous moderation* configurations use real-time evaluators, while *unmoderated* and *asynchronous moderation* configurations either use an evaluator at some point after the session or have the user act as the evaluator themselves. Asynchronous moderation under RUT has seen mixed results (Vasalou et al., 2004; Andreasen et al., 2007; Bruun et al., 2009; Hertzum et al., 2015). Consequently, synchronous methods for RUT are commonly employed because they provide the same level of data quality as conventional usability testing.

The lessons learnt from RUT can be used as a guide for performing remote human-robot interaction experiments especially since it has been shown that RUT is just as effective as conventional in-lab testing. More recently, it has been demonstrated that the physical lab environment can be a simulated environment and usability tests in these conditions trigger no major workload discrepancies when compared to a typical lab or even a remote testing approach (Chalil Madathil & Greenstein, 2017). Given this work, we suggest that Remote Usability Testing (RUT) can address the issue of ethical and realistic HRI studies that has been raised by Feil-Seifer et al. (2020).

### 2.3 Adjusting to a Pandemic - Using Remote Human-Robot Interaction

In our proposed R-HRI methodology that is based on RUT, the moderator acts on behalf of the participant. In conventional and remote usability test scenarios, the moderator observes the user and encourages the user to express his or her thoughts. In the R-HRI methodology however, the moderator is more like that of a proxy for the participant, performing tactile or other physical interactions for the user that would ordinarily violate COVID-19 transmission prevention measures. Additionally, the R-HRI moderator makes no attempt to encourage the participant to vocalize any of his or her thoughts.

To determine the feasibility of our proposal, we created an experiment to evaluate the effect the R-HRI methodology would have on participants' attitudes, cognitive workload, and interaction experience. Specifically, we concentrated on three questions:

1. Does a person involved in an R-HRI experiment express the same attitudes about robots whether they interact with the robot in person or remotely?
2. Will the participant's cognitive workload be the same whether the interaction is in person or remotely?
3. Does the participant have the same user experience independent of an in-person or remote interaction with the robot? Otherwise, is the nature of the R-HRI study so different that it is impossible to use?

Answers to these questions were obtained through a pilot study, which we undertook as a precursor to a larger study. We chose this approach two reasons. First, we did not know what impact the underpinning technologies like the network, device cameras, speakers, etc. would have on the R-HRI experiment. Although our study did not seek to quantify and formalize the technology design, its technological configuration and potential effects were informally observed to evaluate whether an "as-is" configuration was sufficient. We provide more information on this in the Discussion section of the paper. Secondly, we needed to determine whether R-HRI was

irrefutably a negative experience that should be considered as an option. Given these two issues, we concluded that it would be unwise to conduct a full study without first establishing the merit of the idea.

### 3 Related Work

Huber and Weiss (2017) used the term Remote-HRI to explain how participants in their study used a remote-controlled, off-the-shelf robotic arm. The term was used to describe how participants manipulated the robot with a remote control, and contrasted it with the term Physical-HRI, which referred to the user manually manipulating the robot. Our definition focuses on the nature of the physical interaction the user has with the robot during the experiment – specifically the nature of the separation imposed by the experiment – rather than how the robot is controlled robot. This makes the Huber and Weiss definition significantly different from the one given in the work discussed here.

The majority of remote robot interaction research focuses on either how to communicate with a robot in a remote environment (Mackey et al., 2020; Sierra et al., 2019) or teleoperation – how to control the robot from a remote location or by using augmented reality (AR) or other simulated environments (Yanco et al., 2005; Qian et al., 2013; Nagy et al., 2015; Chiou et al., 2016; Chivarov et al., 2019; Xue et al., 2020). Some studies have looked at how to improve HRI outcomes by using the common ground principle to improve communication between humans and robots. Stubbs et al. (2008) investigates using a robot proxy – a device that simulates the robot's responses – to enhance interaction outcomes by creating common ground between the person and the robot. Although HRI was included in the work, the focus of the research was on the efficacy of the human-robot grounding process. Consequently, investigating factors like physical separation, behaviors, and experience is not done, with other work on robot proxies and common grounding following a similar path (Burke & Murphy, 2007; Prasov, 2012).

Honig and Oron-Gilad (2020) recently published work that used online video surveys to determine the consistency of hand movements used to interact with robots. The research was motivated by some of the same factors that drive RUT, namely the expense and effort involved in performing in-person studies, particularly when the robot has not been completely tested for human interaction. The researchers did not use a moderator or use real-time interaction in the study, but they did have participants watch videos and react to them asynchronously through video. The findings reveal that the method of interaction has no impact on seven of the eight movements tested. This suggests that their method could be a viable alternative to in-person testing, especially during the

prototyping stage. Their conclusion is in line with the research discussed in this paper. There is also research under the Remote-HRI category where the interaction is observed by a third party and data is gathered asynchronously (Zhao & McEwen, 2021). This is a variation on what we are proposing in this work. However, this paper is not evaluating the efficacy of the method, they are simply using the remote-asynchronous HRI approach because it is a longitudinal study involving children in their homes.

Human-robot proxemics study the physical and psychological distancing between humans and robots (Mumm & Mutlu, 2011; Walters et al., 2009). The aim of this research is to find out how humans respond to robot actions when they are within a certain range of the user. It also explores how a robot can influence human behavior, human emotions, and other types of affect. The questions we aim to answer in this paper are not explicitly answered by human-robot proxemics. In proxemics, physical distancing is explored, but this is done with the aim of making the robot model social and cultural norms. In our work, physical distancing is a restriction that must be met for experiments to be carried out. It is not a social modelling effort. As a result, rather than exploring physical distance alone, as we would in a human-robot proxemic study (Mumm & Mutlu, 2011), we constrain the physical distance from the robot and conduct the HRI study under that condition.

Finally, there is a form of remote robot interaction; however, the focus is on human-human remote interaction using a robot as an intermediary (Papadopoulos et al., 2013). The robot is a remotely accessible toy that is accessed by a paired robot. These paired toys are used by participants to facilitate remote human-human interaction. In this case, the focus is not on the effect of accessing the robot remotely, but on how that remote interaction can be used to effect human relationships.

In summary, current research looks at how different types of interactions affect the human-robot relationship and in some cases human-human relationships. However, none of these studies have evaluated how a physically constrained environment affects user behavior, cognitive workload, or user experience. In this paper, we discuss these limitations and their effect on interaction by conducting a pilot study to see whether Remote-HRI can be used to perform HRI studies in a global pandemic with minimal physical interaction.

## 4 The Method

### 4.1 Participants

Ten males between the ages of 22 and 46 ( $\mu=32.3$ ,  $\sigma=7.36$ ), two females ( $\mu=24$ ,  $\sigma=1.414$ ) and one moderator (male, 22) participated in the pilot study. The participants were selected from a computer science research group whose primary

research interests are unrelated to social robots. Four of the twelve participants (33%) had already communicated with a robot other than the one used in this research prior to participating in this pilot study. The small sample is based on the advice in (Macefield, 2009), which asserts that a baseline of 5–10 participants is a reasonable starting point for early usability studies.

### 4.2 The Experiment Conditions

The *mode of robot interaction* was the independent variable. There were two levels of evaluation: in-person interaction with a moderator and remote interaction with a moderator.

The in-person environment was made to meet all public health criteria known to minimize the risk of transmission from person-to-person. Face masks were mandatory, and participants were asked to stand at least two meters away from the robot and moderator. Any interaction with the robot that required touching the robot had to go through the moderator - who was the only person allowed to touch it. Even though airborne transmission of the virus is limited, the in-person experiments were performed in an open outdoor courtyard to minimize the risk of aerosolized particles circulating in the test area (World Health Organization and others, 2020). This set up was also used to perform remote interaction with the robot. A laptop with a built-in webcam was used to run the video conferencing software with its speaker system supplemented with an external Bluetooth speaker. The robot was positioned in front of the laptop so that the remote user could see the robot move about while allowing the robot to hear the user's commands.

We used a nearby research lab as the on-site remote location. Six people can work in the lab at a time, but no more than three people, wearing face masks, were permitted in the lab at the same time to ensure adequate spacing. Participants were given gloves and hand sanitizer if they wanted to use the provided laptop to communicate with the robot remotely.

### 4.3 The Robot

In this analysis, the ASUS Zenbo companion robot was used (see Fig. 1). The robot is a multifunctional intelligent assistant that is sometimes described as “cute” (Byrne, n.d.). It is intended to serve as a storyteller for teachers, a companion robot for the elderly, and a retail service assistant in educational, industry, and health and safety settings (Zenbo | Intelligent Robot, n.d.). Given the robot's intended uses, we felt it was a good candidate for first-time users.

We decided to use some of the Zenbo robot's built-in functions since these were employed by other researchers, which somewhat validated these functions as acceptable usability study inputs (Chien et al., 2019). We also chose these applications instead of custom-made applications to prevent



**Fig. 1** The Zenbo intelligent robot (ASUS)

confounding caused by applications that may not be as rigorously tested and could possibly produce unexpected outcomes. We, however, do not compare our results with the work done by Chien et al. (2019) since they used these functions in a completely different study.

Each participant went through a pre-selected list of nine Zenbo functions for each experiment, see Table 1. We chose this list of functions because they were simple to explain to participants and we expected that the users would be able to issue them to the robot easily.

For the users to fully appreciate what Zenbo could do, we ensured the functions included commands that triggered oral and motion responses from the robot. Voice commands and touch were also used to communicate with Zenbo. Five of the nine commands (1, 2, 3, 4, and 9) were voice commands that elicited an oral response from Zenbo. Three voice commands (6, 7, 8) caused Zenbo to respond with an oral and movement response. One command (5) involved physical contact with

the robot that would make Zenbo move. Except for Command 5, the participants communicated with Zenbo directly. However, to execute Command 5, participants had to ask the moderator to perform the operation instead of touching the robot themselves.

#### 4.4 Devices Used for Remote Access

Participants had three options for accessing the robot remotely: (i) use the laptop provided; (ii) use their personal laptop or (iii) use their smartphone. The laptop provided to participants ran the Chrome web browser and the remote desktop application AnyDesk®. We used the Jitsi (‘Jitsi.Org’, 2021) video conferencing website to allow participants to communicate with the Zenbo companion robot, see it move around and to communicate with the moderator. The AnyDesk application was used to give the participants a clear picture of the robot’s face and its expressions because these were hard to discern through the video conferencing software. AnyDesk was installed on Zenbo to make this type of connection possible. Participants who chose to use their own laptop had to install AnyDesk and then use the application to request a connection to the robot. The moderator ensured this was done at the start of the session. The final device option was for participants to use their smartphones. With this option participants had to install the Jitsi application to get the best quality connection. Additionally, using the smartphone prevented participants from using the AnyDesk application since it was not possible to use both the Jitsi and AnyDesk applications simultaneously. Using this arrangement, participants saw the robot’s expressions only through the Jitsi video conferencing application.

**Table 1** List of Zenbo commands numbered 1–9

Command	Function	Expected reaction
1) “Hey, Zenbo”	Prepare to receive a command	Blue “ears” appear and waits for a command
2) What can you do?	Opens list of functions	Display functions
3) “What date is it tomorrow” OR “What date is it tomorrow on the lunar calendar”	Date report	Tomorrow’s date in the calendar / lunar calendar
4) “What is the weather in Barbados today?” OR “What is the weather today?”	Weather report	Reports the weather
5) Stroke its head (ask moderator)	Basic interaction	Shows a shy expression
6) “Follow me”	Following	Follows the user
7) “Tell me a story”	Entertainment	Tells the user a story
8) “I want to take a picture” / “I want to take a selfie”	Photo	Takes a picture of the user
9) “I want to listen to music”	Entertainment	Plays music



## 4.5 The Survey Instruments

For this study we selected five survey instruments that could help us determine whether: (a) participants' mode of interaction with the robot (in-person/remotely) affected their attitudes towards the robot; (b) the effort expended in interacting with the robot depended on interaction modality and (c) the quality of the participants' experience with the robot was affected by the interaction mode.

We used the Negative Attitude towards Robots Scale (NARS) (Nomura et al., 2006) to determine the level of negative perceptions held by the participants towards robots. NARS is a 14-item scale that has three subscales. The first, six-item subscale, measures negative attitudes towards human-robot social interactions (HRSI). The second, five-item subscale, measures the negative attitudes towards the social influence of robots (SIR) and the third, three-item subscale, measures negative attitudes towards emotional interactions with robots (EIR). Measuring negative perceptions was important since there are very few companion, assistive or service robots available in our region. This meant that any impression participants may have had about robots would likely not have been from prior exposure, but from second-hand sources like popular media. The higher the NARS score, the more negative the participant's attitude towards robots. This means that by using this scale we could determine whether participants' attitudes change after each interaction with the robot (Pretest:  $\alpha = .76$ ; Posttest I:  $\alpha = .73$ ; Posttest II:  $\alpha = .68$ ).

The Robotic Social Attributes Scale (RoSAS) (Carpinella et al., 2017) was used to get a better sense of the perceptions held by participants, especially if NARS showed that participants did not have any negative attitudes towards robots. RoSAS is an 18-item scale with three subscales: *Warmth*, *Competence* and *Discomfort*, each consisting of six items. The higher the score in each subscale, the greater the perception by the user of the robot possessing the characteristics described by the items of the subscale. For example, a robot that is perceived as warm, implies that the participant scored it highly in the six items: happy, feeling, social, organic, compassionate, and emotional that comprises the warmth subscale. We used this scale to determine if any of these perceptions changed based on the interaction modality used with the robot (Pretest:  $\alpha = .89$ ; Posttest I:  $\alpha = .78$ ; Posttest II:  $\alpha = .75$ ).

In addition to human perception, we also assessed use of the robot to capture whether the interaction affected levels of technology adoption. We used the Extended Technology Acceptance Model (TAM2) (Venkatesh & Davis, 2000) scale to determine participants' attitudes towards accepting a companion robot for use if one were available. TAM2 measures technology acceptance, and we regarded it as a viable indicator of the quality of the interaction with the robot. We assessed the factors: Intention to Use (IU); Perceived Usefulness (PU) and Perceived Ease of Use (PEU). The higher the TAM2 score, the more likely the

participant is to accept and/or adopt the technology under assessment (In-Person:  $\alpha = .82$ ; Remotely:  $\alpha = .82$ ).

The NASA Task Load Index (TLX) consists of six subscales: Mental Demand (MD), Physical Demand (PD), Temporal Demand (TD), Performance (P), Effort (E) and Frustration (F) (Hart, 2006). Its total score measures the workload associated with a specific task. In the between subject study, we wanted to determine if there was any effect on all aspects of the user's workload, so we evaluated all subscales. For the within subject study, the key measures we wanted to capture was the level of work the participants' experience with the Zenbo robot within the constraints of the experiment environment. However, since we examined all aspects of the scale in the between-subjects study, we decided to do the same for the within subject study when examining the two modes of interaction (In-Person:  $\alpha = .81$ ; Remotely:  $\alpha = .74$ ).

The short version of the User Experience Questionnaire (Laugwitz et al., 2008) is a seven-point, eight-item inventory that allows subjects to provide a full assessment of their experience using a technology. It has two subscales: *Pragmatic Quality*, which measures how efficiently you can perform the task using the product; and *Hedonic Quality*, which measures how interesting and stimulating it is to use the product to perform the task. Values greater than 0.8 represent a *positive* evaluation, values less than  $-0.8$  represent a *negative* evaluation. Values between  $-0.8$  and  $0.8$  represent a neutral evaluation. Given this, we can determine the quality of the participant's experience after each interaction mode (In-Person Pragmatic:  $\alpha = .95$ ; Remotely Pragmatic:  $\alpha = .87$ ); (In-Person Hedonic:  $\alpha = .97$ ; Remotely Hedonic:  $\alpha = .85$ ).

## 4.6 The Experiment Design

We performed the study with two independent groups. Group 1 consisted of five participants, two females three males, and the experiment investigated the independent variable, interaction mode, at one level – remote interaction. Group 2 consisted of seven males and the experiment employed a counterbalanced measures design with the independent variable (interaction mode) at two levels: in-person and remote. The seven participants were randomly assigned to either Group 2A (which first interacted with the robot in-person and then remotely) or Group 2B (which interacted with the robot remotely first, and then, in-person). Data for the twelve participants was collected over a period of nine hours.

Given the small sample size of the pilot study, we used the Friedman's non-parametric test on the NARS and RoSAS data for within subject analysis of the effects of the interaction modalities against attitudes towards robots. To analyze the other scales that measured technology adoption (TAM2), workload effort (NASA TLX), and user experience (UEQ), we tested for normality using a visual test (Q-Q Plot) verified by a Shapiro-Wilk test, which has more power than the

Kolmogorov-Smirnov normality test (Ghasemi & Zahediasl, 2012). If the data met the normality requirement, we investigated the presence of effects using a repeated measures  $t$ -test, since it functions with small sample sizes (De Winter, 2013). Data samples that failed the normality tests were analyzed using the Wilcoxon Signed-Rank test.

For the between subject study of Group 1 (participants that had one interaction with the robot) and Group 2 (participants that had two interactions with the robot) our analysis had to also account for the small sample size. We tested for normality using the Q-Q plot test and verified normality with the Shapiro-Wilk even though the  $R^2$  values for the trendlines in the Q-Q plots were greater than 0.85, indicating a high correlation with the normal distribution. Once the data passed the normality test, we inspected the data for the presence of effects between the two samples using Welch's  $t$ -test for independent samples.

## 4.7 Procedure

Before starting the experiment, participants were given an online copy of the informed consent form, which they had to read and then click to accept. After giving consent, basic demographic information such as age and whether that had ever interacted with a robot was collected. Participants then filled out the NARS and RoSAS inventories and instructed on how to issue the nine commands, shown previously in Table 1, to interact with the robot. This was the pretest phase of the experiment.

Group 1 participants issued the commands to the robot remotely using the environment described in Experiment Conditions, after which they were instructed to fill out the five inventories. Upon completion of the surveys, they were thanked for their participation and their session ended. For Group 2 participants, after their first interaction, they were also instructed to fill out the five inventories. When they completed these surveys, they used the same commands to interact with the robot a second time but using a different interaction modality. To simplify, a Group 2 participant, who for example may have been assigned to Group 2A, would complete the NARS and RoSAS inventories during the pretest phase, interact with the Zenbo robot *in-person*, and then complete five inventories in the first posttest phase. After their second interaction, which would happen remotely, they would complete the five inventories in the second posttest phase. The session ended after they completed the surveys.

## 5 Results

### 5.1 The between Subject Study - One Interaction vs Two Interactions

In the between subject study, we evaluated the data collected from the NASA TLX and the UEQ inventories. The data

collected from the surveys completed by the Group 1 participants after their only interaction with the robot (posttest phase 1) is evaluated against the survey data from Group 2 participants after their second interaction with the robot (posttest phase 2).

#### 5.1.1 NASA Task Load Index

For the NASA TLX scale, we used Q-Q plots to perform normality visual tests on the survey that indicated the data for all subscales were normally distributed for both groups, Fig. 2. We ran Shapiro-Wilk as a verification test for normality, Table 2, which allowed us to use the Welch's  $t$ -test to analyze the overall workload and all subscales except for Effort and Frustration. Analyzing the data with the Welch's  $t$  and Wilcoxon Signed Rank tests found that the number of interactions had no significant effect for the Overall workload and the subscales, Table 3.

#### 5.1.2 The User Experience Questionnaire

We tested the Pragmatic Quality and Hedonic Quality subscale data for normality using the Shapiro-Wilk test. Both data sets met the normality criteria for both the one interaction and two interactions data set: Pragmatic<sub>1</sub> ( $W = 0.885, p = .33$ ), Pragmatic<sub>2</sub> ( $W = 0.925, p = .51$ ), Hedonic<sub>1</sub> ( $W = 0.974, p = .89$ ) and Hedonic<sub>2</sub> ( $W = 0.916, p = .43$ ). Based on these results, we used the Welch's  $t$ -test for independent samples and found that whether the user had one or two interactions with the robot, it had no effect on either the Pragmatic Quality ( $t(10) = -0.978, p = .35$ ) or the Hedonic Quality ( $t(10) = -0.100, p = .92$ ) of the experience.

## 5.2 The Within Subject Study

In this study we analyzed the data collected from all five surveys taken by the seven participants who interacted with the robot in-person and remotely.

#### 5.2.1 NARS

We performed the Friedman's Test on the three subscales of the NARS separately. The three treatments were the pre-interaction versus in-person interaction versus remote interaction against the independent variable *interaction modality* at two levels: in-person and remote. For the HRSI subscale there was no significant effect of the interaction modality on user attitudes,  $X^2(2, N = 7) = 3.43, p = 0.18$ . For the SIR subscale, there was no significant effect of the interaction modality on social interaction attitudes,  $X^2(2, N = 7) = 0.5, p = 0.78$ . Finally, for the EIR subscale, there was no significant effect of the interaction modality on attitudes towards emotional interactions with robots,  $X^2(2, N = 7) = 0.28, p = 0.87$ .

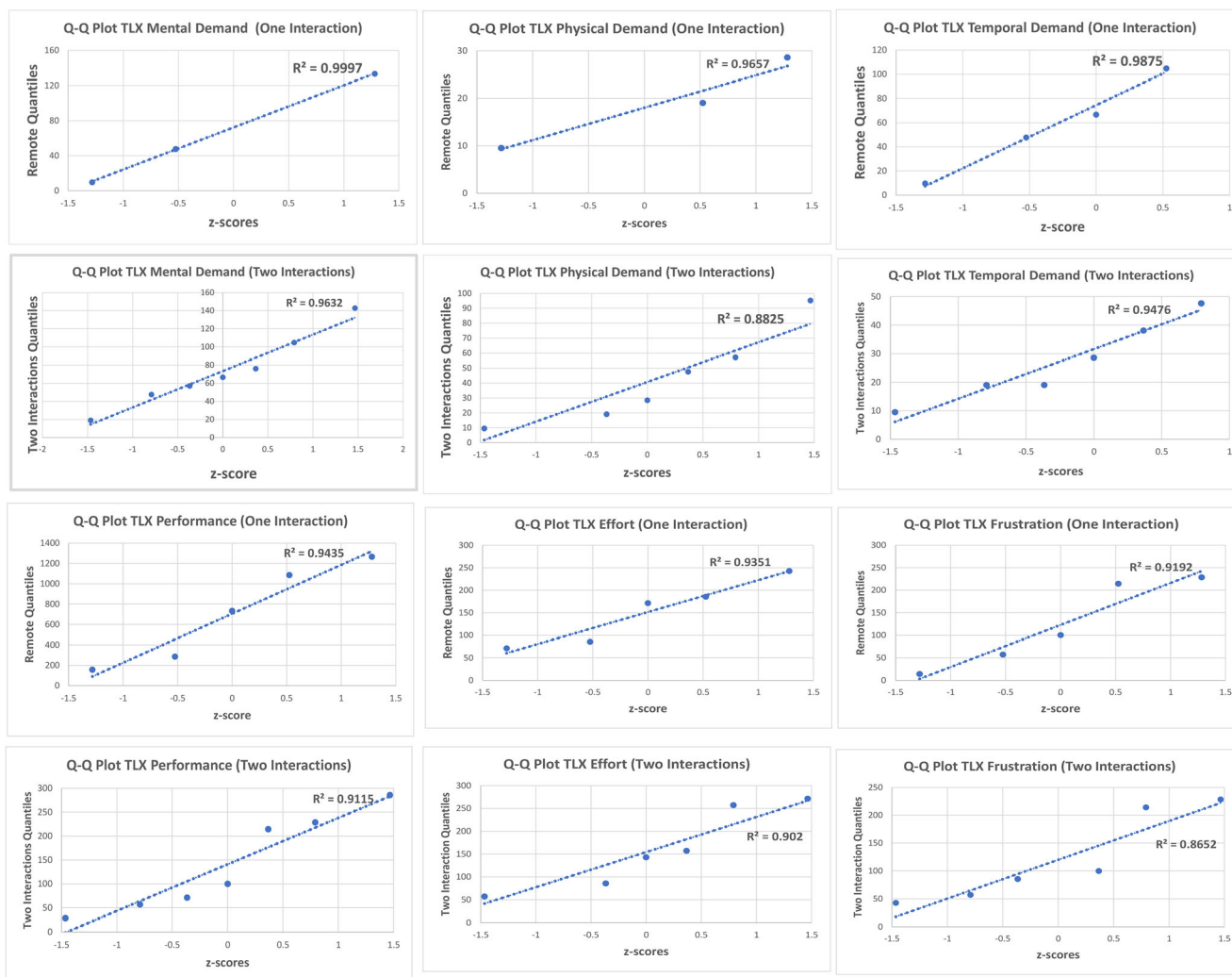


Fig. 2 Q-Q Plot visual test for normality for the NASA TLX inventory

### 5.3 RoSAS

We performed a Friedman’s test on the three subscales of the RoSAS inventory. The three treatments were the pre-interaction versus the in-person interaction versus the remote interaction against the interaction modality at two levels: in-person and remotely. For the *Warmth* subscale there was no significant effect of the interaction modes on this attitude,  $X^2(2, N = 7) = 0.5, p = 0.78$ . For the *Competent* subscale there was no significant effect of the interaction modalities on the attitude,  $X^2(2, N = 7) = 2, p = 0.37$ . For the *Discomfort* subscale, there was no significant effect of the interaction modalities on this attitude,  $X^2(2, N = 7) = 1.36, p = 0.51$ .

### 5.4 TAM2

For the TAM2 scale, we used Q-Q plots to perform normality visual tests. They appeared to indicate that the data for all three

subscales were normally distributed, Fig. 3. We ran the Shapiro-Wilk normality for additional verification, which allowed us to use the *t*-test for all subscale analysis. We performed a repeated measures *t*-tests on the *Intention to Use* (IU) subscale from the TAM2 inventory. The results indicated that the interaction modality had no effect on Intention to Use  $t(6) = .733, p = .49$ .

We performed a repeated measures *t*-test on the *Perceived Usefulness* (PU) subscale and the results indicated that the interaction modality had no effect on Perceived Usefulness  $t(6) = .668, p = .53$ . Finally, we performed a repeated measures *t*-test on the *Perceived Ease of Use* (PEU) subscale and the results indicated that the interaction modality had no effect on Perceived Ease of Use  $t(6) = -1.056, p = .33$ .

### 5.5 NASA Task Load Index

For the NASA TLX scale, we used Q-Q plots to perform normality visual tests that appeared to indicate that all data,



**Table 2** Shapiro-Wilk normality tests for one- and two interaction NASA TLX data

	One interaction (N=5)	Two interactions (N=7)
<b>Overall Workload</b>		
Mean	74.60	30.57
Standard Deviation	40.71	17.83
Shapiro-Wilk Normality Test Results		
W	0.879	0.923
p value	0.376	.492
<b>Mental Demand</b>		
Mean	57.14	58.50
Standard Deviation	45.67	43.44
Shapiro-Wilk Normality Test Results		
W	0.811	0.886
p value	0.128	.257
<b>Physical Demand</b>		
Mean	15.24	24.49
Standard Deviation	8.52	16.36
Shapiro-Wilk Normality Test Results		
W	0.771	.842
p value	0.068	.104
<b>Temporal Demand</b>		
Mean	66.67	28.57
Standard Deviation	40.41	15.55
Shapiro-Wilk Normality Test Results		
W	0.909	0.848
p value	0.586	0.118
<b>Performance</b>		
Mean	705.71	161.22
Standard Deviation	484.03	100.24
Shapiro-Wilk Normality Test Results		
W	0.924	0.843
p value	0.7	0.106
<b>Effort</b>		
Mean	141.43	104.08
Standard Deviation	71.86	78.12
Shapiro-Wilk Normality Test Results		
W	0.923	0.788
p value	0.693	< .05
<b>Frustration</b>		
Mean	122.86	81.63
Standard Deviation	95.08	68.37
Shapiro-Wilk Normality Test Results		
W	0.899	0.718
p value	0.506	< .05

except the Physical Demand In-Person and Effort In-Person subscales, were normally distributed, Fig. 4. We ran the Shapiro-Wilk as a verification test for normality, Table 4, which allowed us to use the *t*-test to analyze the

**Table 3** Results of the Welch’s *t*-tests and Wilcoxon Signed Rank test for the NASA TLX inventory - one interaction and two interaction groups

	Welch’s <i>t</i> -test	
	p value	t Statistic
Overall Workload	.07	2.268
Mental Demand	.96	−0.052
Physical Demand	.23	−1.274
Temporal Demand	.10	2.004
Performance	.06	2.478
	Wilcoxon Signed Rank Test	
	p value	W Statistic
Effort	.19	26
Frustration	.68	20.5

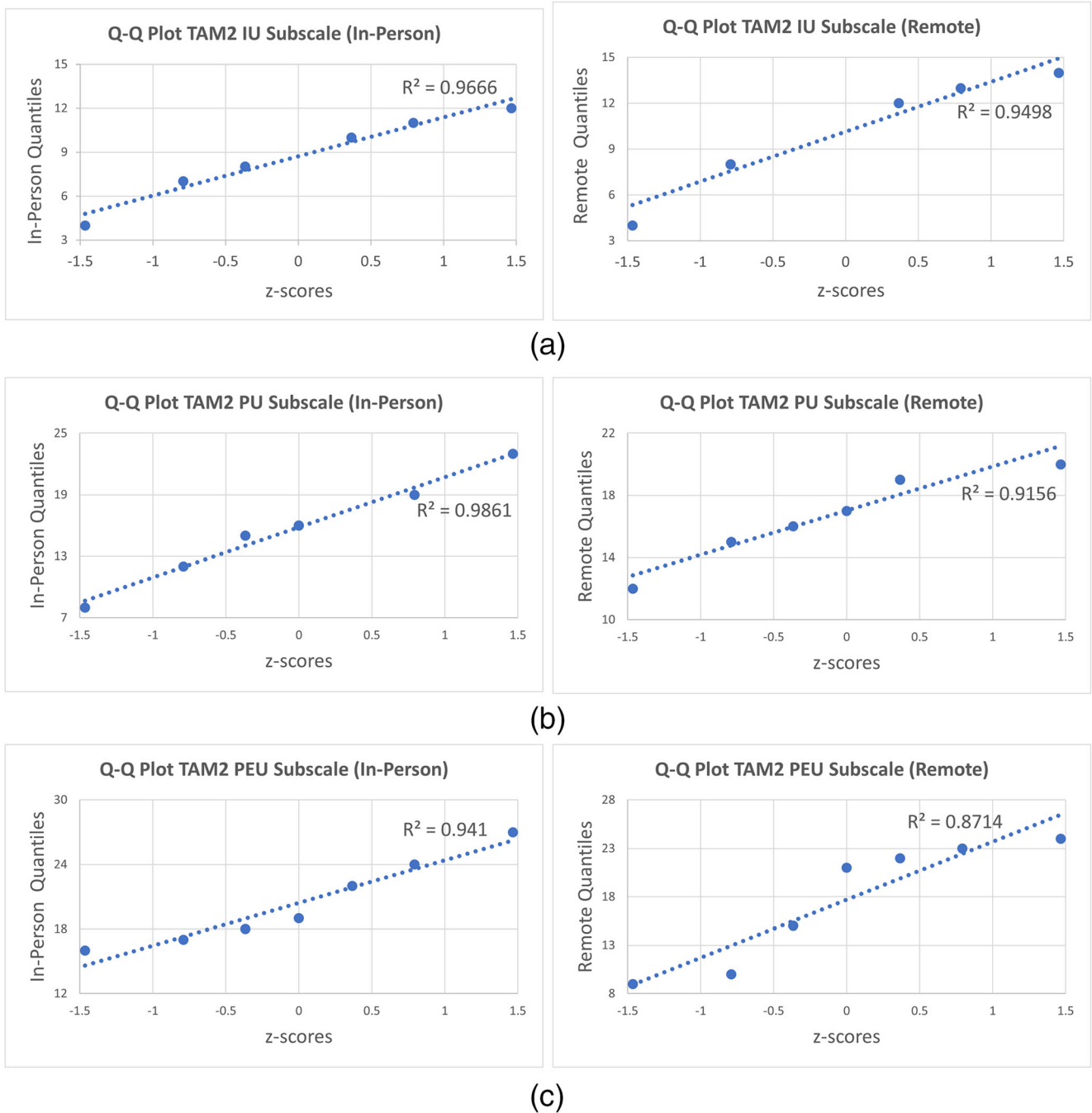
subscales that passed both tests. For the Physical Demand and Effort subscales we used the Wilcoxon Signed-Rank test.

The result of the repeated measures *t*-test for the interaction modality showed that it influenced the Overall workload ( $t(6) = 3.69, p < .05$ ), Mental Demand ( $t(6) = 3.73, p < .05$  and Frustration ( $t(6) = 2.53, p < .05$ ). It also showed that it did not influence Temporal Demand ( $t(6) = 0, p = 1$ ), or Performance ( $t(6) = 1.456, p = .19$ ). These findings imply that interacting with the robot remotely required more effort than interacting with the robot in-person although physically separated from the robot. Temporal demands and performance were the same for both interaction modes, which indicated that the participants never felt hurried and achieved their goals in both interaction modes.

The results of the Wilcoxon Signed-Rank test showed that the interaction modality had no effect on Physical Demand ( $V = 1, p = .20$ ) but affected Effort ( $V = 1, p < .05$ ). These preliminary results reflect the fact that there was no physical interaction with the robot because the commands were verbal, and any physical interaction that required was done by the moderator. Additionally, the effort required for in-person interaction was lower than that required for remote interaction. Caution is noted here however, because the sample was small, and a normal approximation was used to calculate the *p* value.

### 5.6 UEQ

The Q-Q Plots appear to indicate that the data for the UEQ subscales were normally distributed, Fig. 5. Both subscales passed the Shapiro-Wilk normality test, see Table 5, which meant we could use the *t*-test. For the *Pragmatic* subscale, the repeated measures *t*-test ( $t(6) = -0.254, p = .81$ ) showed there was no effect of the interaction modality on the pragmatic quality of the experience. For the *Hedonic* subscale the repeated measure *t*-test ( $t(6) = .950, p = .38$ ), showed that there



**Fig. 3** Q-Q plots of **a** TAM2 IU subscale; **b** TAM2 PU subscale and; **c** TAM2 PEU subscale showing that the subscale data may fit a normal distribution, with minimum  $R^2$  value of 0.871, 2 (c), and maximum  $R^2$  value of 0.9861, 2 (b)

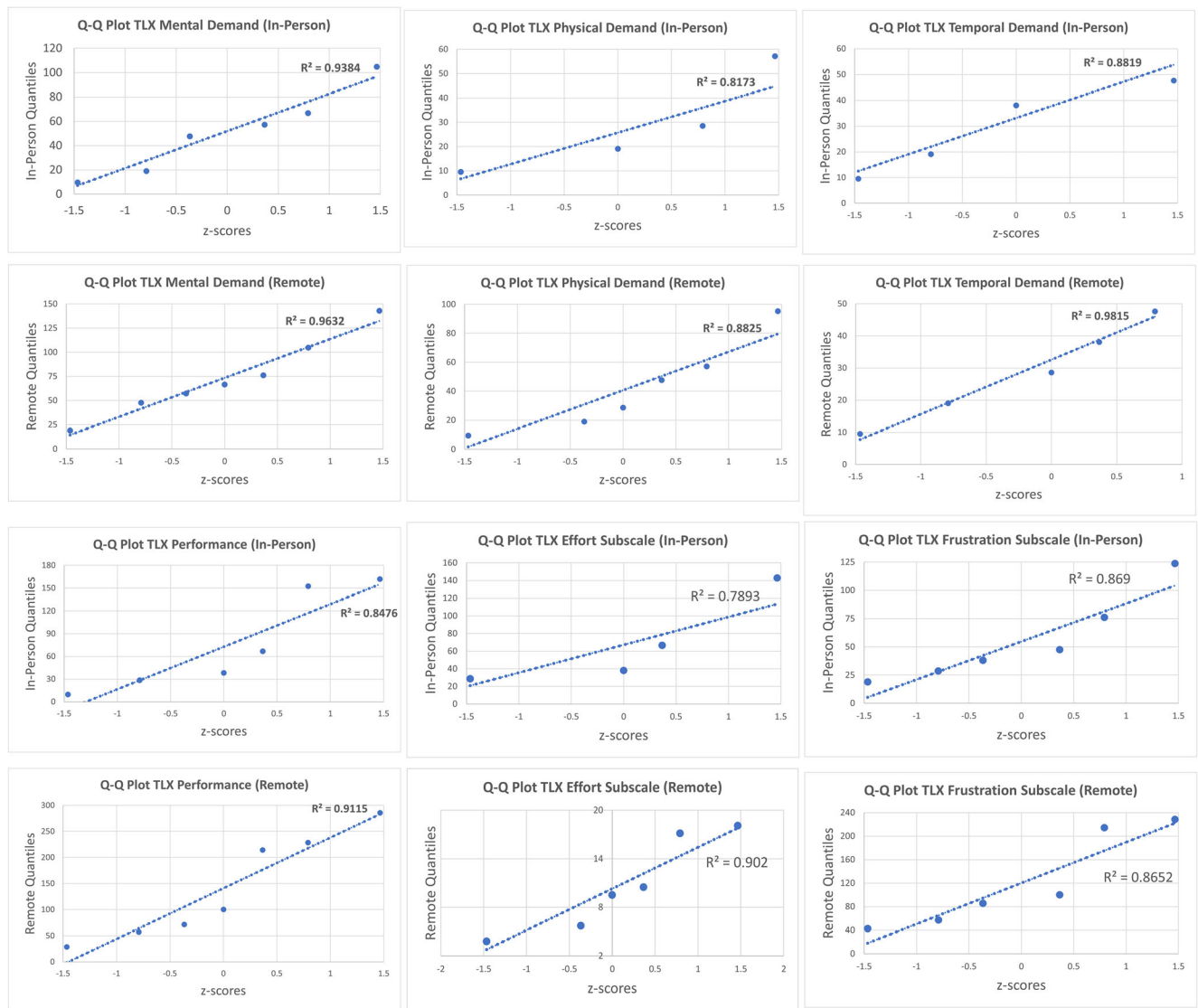
was no effect by the interaction mode on the hedonic quality of the experience.

## 6 Discussion

In this paper we presented a pilot study of a Remote-HRI methodology in which we examined whether:

1. The number of interactions affected the user’s workload.
2. The participants’ mode of interaction (in-person versus remotely) with the robot affected their attitudes towards, and perceptions of, the robot.
3. The quality of the participants’ experience with the robot was affected by the interaction mode.

To address these questions, we first examined data taken from the two groups as independent samples. The



**Fig. 4** Q-Q Plot visual normality test for all NASA TLX subscales

difference in the overall workload experienced by the two groups was statistically insignificant. This is an interesting result because, as we will discuss below, participants reported lower overall effort for in-person interaction versus remoter interaction. This may indicate that remotely interacting with requires significantly larger effort that is not mitigated by in-person interaction. This requires further study since four of the seven participants remotely interacted with the robot for the second interaction and with the small sample size effects would be unclear. However, the quality of experience was similar across the number of interactions, which is encouraging because, even though more effort is reported for remote interactions, it does not have a negative effect on the overall experience.

In the within subject study, we examined participants' attitudes towards robots using the NARS and RoSAS inventories and found that the mode of interaction had no effect. We then

used the TAM2 inventory to examine the acceptance of social robot technology and the intention to adopt such technology. We also found that the interaction mode had no effect on these attitudes.

In this study we also examined the interaction using the NASA TLX scale and we found that the interaction mode affected all aspects of the workload except Physical and Temporal demands as well as Effort as mentioned above. This is an important finding, and we believe that this merits further investigation on two fronts: (i) the voice recognition capabilities of the robot and (ii) the technological infrastructure of the experiment environment. Regarding the first issue, the Zenbo robot showed difficulty in recognizing the wake-up command "Hey Zenbo" in the in-person interaction mode with participants sometimes shouting or changing their pronunciation to get the robot to respond. This effect was noticed for most of the other nine commands. We suspect that this

**Table 4** NASA TLX Statistics and normality tests for in-person and remote interaction

	In-Person ( <i>N</i> =7)	Remotely ( <i>N</i> =7)
<b>Overall Workload</b>		
<i>Mean</i>	18.78	36.37
<i>Standard Deviation</i>	10.72	16.40
<b>Shapiro-Wilk Normality Test Results</b>		
<i>W</i>	0.824	0.979
<i>p value</i>	.081	.971
<b>Mental Demand</b>		
<i>Mean</i>	50.34	73.47
<i>Standard Deviation</i>	47.62	40.30
<b>Shapiro-Wilk Normality Test Results</b>		
<i>W</i>	0.951	0.969
<i>p value</i>	.844	.968
<b>Physical Demand</b>		
<i>Mean</i>	21.77	38.09
<i>Standard Deviation</i>	17.14	31.10
<b>Shapiro-Wilk Normality Test Results</b>		
<i>W</i>	0.768	8.889
<i>p value</i>	< .05	0.302
<b>Temporal Demand</b>		
<i>Mean</i>	29.93	29.93
<i>Standard Deviation</i>	13.94	14.99
<b>Shapiro-Wilk Normality Test Results</b>		
<i>W</i>	0.888	0.913
<i>p value</i>	0.295	0.471
<b>Performance</b>		
<i>Mean</i>	69.38	140.82
<i>Standard Deviation</i>	62.38	100.14
<b>Shapiro-Wilk Normality Test Results</b>		
<i>W</i>	0.814	0.896
<i>p value</i>	0.07	0.345
<b>Effort</b>		
<i>Mean</i>	57.14	146.94
<i>Standard Deviation</i>	41.51	89.10
<b>Shapiro-Wilk Normality Test Results</b>		
<i>W</i>	0.751	0.875
<i>p value</i>	< .05	0.227
<b>Frustration</b>		
<i>Mean</i>	53.06	116.33
<i>Standard Deviation</i>	35.99	74.43
<b>Shapiro-Wilk Normality Test Results</b>		
<i>W</i>	0.845	0.823
<i>p value</i>	.125	0.080

may be a result of the dataset used to train the robot's voice recognition system combined with the fact that the experiment was performed at a Caribbean university where the

participants spoke with accents from three different Caribbean countries. The second issue is simpler to rectify, and that is to ensure a reliable wireless or wired network connection, and a speaker that produces a clear sound so that the robot can hear and process the remote voice commands. We observed that when there was slight distortion in the speaker because a participant was speaking too loudly into the remote device, the Zenbo robot could not process the command. As we discussed earlier, since no physical interaction was permitted with the robot, we expected that the Physical Demand subscale would show no effect. The Effort subscale reflected the findings that overall, remote interactions exacted a higher workload on participants than in-person interactions.

We investigated the quality of the participants' experience using the Short UEQ inventory and found that the experiences were neutral and that there was no significant difference in the quality of experience based on the interaction mode.

## 6.1 Limitations

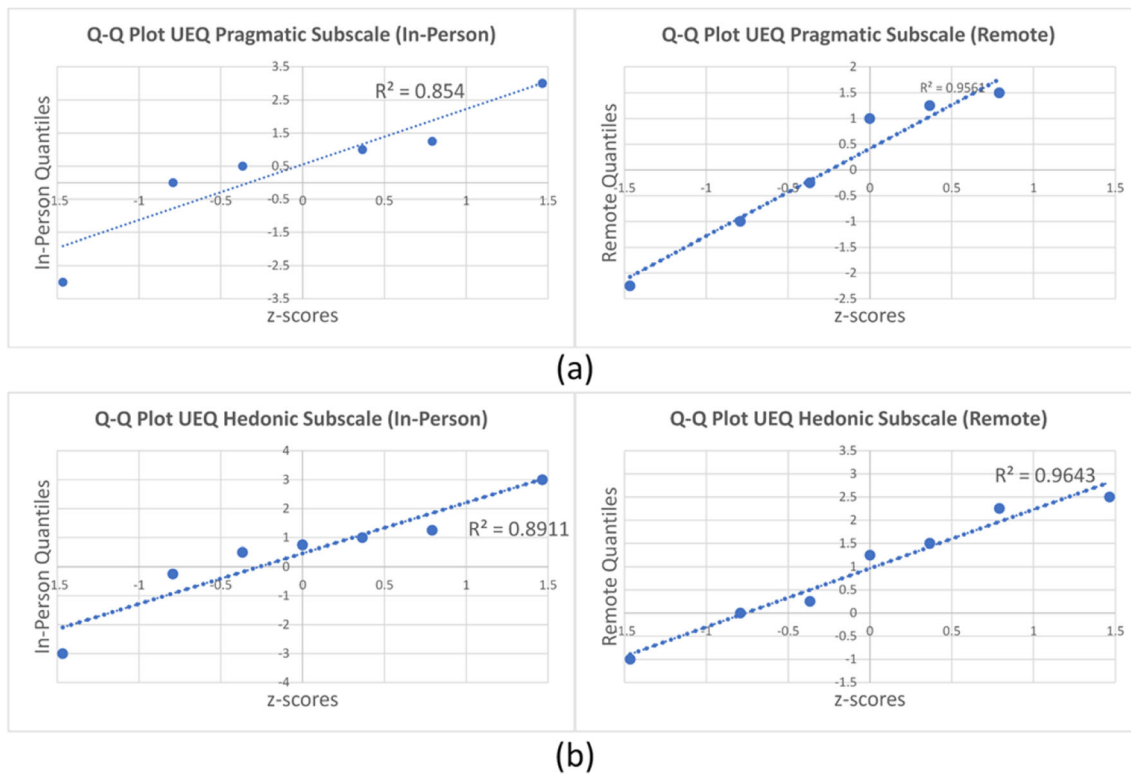
These results have limitations beyond the small sample size and the inconsistent response of the robot due to its voice recognition system. First, we intend to undertake a larger study to obtain clarity on the results we have obtained here. Additionally, the extent to which command repetition affected effort and frustration cannot be quantified since we did not gather that data and only observed this phenomenon as part of the experiment. Second, the effect of the supporting technology was not examined since we focused on other aspects of the Remote-HRI methodology. We also did not conduct a clear comparison between a full interactive high-touch experiment and a completely remote experiment. This prevented us from making a clear comparison between these methodologies.

The quality of the network connection may have influenced aspects of the interaction since there were some points when performance degraded and affected sound and/or video quality. These factors need to be captured so that the remote environment can be adequately designed, and its effects mitigated. Lastly, we did not capture the effect of the smartphone remote device versus the laptop remote device and whether that affected any of our measures. We also must examine this in our full study.

## 7 Conclusions and Future Work

The ability to conduct HRI studies has been constrained by the current COVID-19 pandemic. Finding a way to conduct these studies in an ethical and safe way is important, especially since most HRI research focuses on using social robots to improve quality of life for two of the most at-risk populations – the elderly and those with underlying health conditions.





**Fig. 5** Q-Q Plot of UEQ **a** Pragmatic subscale and **b** Hedonic subscale showing that the data seems to follow a normal distribution

In this paper we presented the results of a pilot study designed to assess the possibility of a Remote-HRI methodology. The results indicated that for our first and third questions—the participants’ perceptions of the robot and their user experiences—were unaffected by interaction modes. This may imply that experiments focused on evaluating user attitudes towards robots can be performed remotely once voice commands are used with limited physical contact. The results for our second question require further analysis due to two other factors that were not examined as part of this study: the supporting technology and the robot’s voice recognition system. We do not believe that these two factors would prevent adoption of Remote-HRI, indeed, we are confident that upon fuller study, we will be able to provide mitigation strategies to reduce the effects of these factors.

In summary, these results indicate that performing HRI studies remotely can be a feasible alternative to face-to-face HRI studies – especially for studies that have a large voice-command component. Perception and attitude studies are good candidates for this methodology, and low-touch studies can be performed using the R-HRI in-person option by making use of the moderator since this process does not negatively affect the perception of or attitudes towards robots.

Given the promising initial results, we will undertake a full study to validate these results and explore whether the participants’ age and dialect has some effect on attitudes and

experiences within the interaction modes. We will also evaluate whether different types of robotic applications are more suited for different types of Remote-HRI interaction beyond those that are primarily voice driven, since successfully facilitating physical interactions will have a significant impact on the feasibility of R-HRI studies.

**Table 5** Statistics and Normality Test Results UEQ

	In-Person (N=7)	Remotely (N=7)
Subscale: Pragmatic		
Mean	0.464	0.25
S.D.	1.805	1.458
Shapiro-Wilk Normality Test		
W	0.897	0.857
p value	0.351	0.157
Subscale: Hedonic		
Mean	0.464	0.96
S.D.	1.82	1.27
Shapiro-Wilk Normality Test		
W	0.912	0.954
p value	0.468	0.869

## References

- Al-Taei, M. A., Kapoor, R., Garrett, C., & Choudhary, P. (2016). Acceptability of robot assistant in management of Type 1 diabetes in children. *Diabetes Technology & Therapeutics*, 18(9), 551–554.
- Andreasen, M.S., Nielsen, H.V., Schröder, S.O., & Stage, J. (2007). What Happened to Remote Usability Testing? An Empirical Study of Three Methods. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 1405–14. CHI '07. San Jose, California, USA: Association for Computing Machinery. <https://doi.org/10.1145/1240624.1240838>.
- Barata, A.N. (2019). Social Robots as a Complementary Therapy in Chronic, Progressive Diseases. In *Robotics in Healthcare*, 95–102. Springer.
- Bruun, A., Gull, P., Hofmeister, L., & Stage, J. (2009). Let your users do the testing: A comparison of three remote asynchronous usability testing methods. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 1619–28. CHI '09. New York, NY, USA: Association for Computing Machinery. <https://doi.org/10.1145/1518701.1518948>.
- Burke, J., & Murphy, R. (2007). RSVP: An investigation of remote shared visual presence as common ground for human-robot teams. In *Proceedings of the ACM/IEEE International Conference on Human-Robot Interaction*, 161–68. HRI '07. Arlington, Virginia, USA: Association for Computing Machinery. <https://doi.org/10.1145/1228716.1228738>.
- Byrne, S. (n.d.). Zenbo family robot destroys Computex with cuteness before it even begins. CNET. Accessed 26 July 2020. <https://www.cnet.com/news/zenbo-family-robot-destroys-computex-with-cuteness-before-it-even-begins/>
- Cabecinhas, A.R.G., Roloff, T., Stange, M., Bertelli, C., Huber, M., Ramette, A., Chen, C. et al (2021) SARS-CoV-2 N501Y introductions and transmissions in Switzerland from beginning of October 2020 to February 2021 – Implementation of Swiss-wide diagnostic screening and whole genome sequencing. *MedRxiv*, January, 2021.02.11.21251589. <https://doi.org/10.1101/2021.02.11.21251589>.
- Carpinella, C.M., Wyman, A.B., Perez, M.A., Stroessner, S.J. (2017). The robotic social attributes scale (RoSAS) development and validation. In *Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction*, 254–262.
- Castillo, J.C., Rex Hartson, H., & Hix, D. (1998). Remote usability evaluation: Can users report their own critical incidents? In CHI 98 conference summary on human factors in Computing systems - CHI '98, 253–254. : ACM Press. <https://doi.org/10.1145/286498.286736>.
- CDC. (2020). Coronavirus disease 2019 (COVID-19). Centers for Disease Control and Prevention. 11 February 2020. <https://www.cdc.gov/coronavirus/2019-ncov/need-extra-precautions/people-with-medical-conditions.html>
- Chalil Madathil, K., & Greenstein, J. S. (2017). An investigation of the efficacy of collaborative virtual reality systems for moderated remote usability testing. *Applied Ergonomics*, 65(November), 501–514. <https://doi.org/10.1016/j.apergo.2017.02.011>.
- Chien, S.-E., Chu, L., Lee, H.-H., Yang, C.-C., Lin, F.-H., Yang, P.-L., Wang, T.-M., & Yeh, S.-L. (2019). Age difference in perceived ease of use, curiosity, and implicit negative attitude toward robots. *ACM Transactions on Human-Robot Interaction*, 8(2), 9:1–9:19. <https://doi.org/10.1145/3311788>.
- Chiou, M., Biaksaite, G., Hawes, N., Stolkin, R. (2016). Human-initiative variable autonomy: An experimental analysis of the interactions between a human operator and a remotely operated Mobile robot which also possesses autonomous capabilities. In *AAAI Fall Symposia*.
- Chivarov, N., Chikurtev, D., Chivarov, S., Pleva, M., Ondas, S., Juhar, J., & Yovchev, K. (2019). Case study on human-robot interaction of the remote-controlled service robot for elderly and disabled care. *Computing and Informatics*, 38(5), 1210–1236.
- De Winter, J. C. F. (2013). Using the Student's t-test with extremely small sample sizes. *Practical Assessment, Research, and Evaluation*, 18(1), 10. <https://doi.org/10.7275/e4r6-dj05>.
- Dörndorfer, J., & Seel, C. (2020). Context modeling for the adaption of Mobile business processes – An empirical usability evaluation. *Information Systems Frontiers*. <https://doi.org/10.1007/s10796-020-10073-w>.
- Fang, X., & Holsapple, C. W. (2011). Impacts of navigation structure, task complexity, and users' domain knowledge on web site usability—An empirical study. *Information Systems Frontiers*, 13(4), 453–469. <https://doi.org/10.1007/s10796-010-9227-3>.
- Feil-Seifer, D., Haring, K. S., Rossi, S., Wagner, A. R., & Williams, T. (2020). Where to next? The impact of COVID-19 on human-robot interaction research. *ACM Transactions on Human-Robot Interaction*, 10(1), 1–7. <https://doi.org/10.1145/3405450>.
- Ghasemi, A., & Zahediasl, S. (2012). Normality tests for statistical analysis: A guide for non-statisticians. *International Journal of Endocrinology and Metabolism*, 10(2), 486–489. <https://doi.org/10.5812/ijem.3505>.
- Hammontree, M., Weiler, P., & Nayak, N. (1994). Remote usability testing. *Interactions*, 1(3), 21–25.
- Hart, S.G. (2006). NASA-task load index (NASA-TLX); 20 years later. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 50:904–8. Sage Publications Sage CA.
- Hartson, H. R., Castillo, J.C., Kelso, J., Neale, W.C. (1996). Remote evaluation: The network as an extension of the usability laboratory. In *Proceedings of the SIGCHI conference on human factors in Computing systems common ground - CHI '96*, 228–235. Vancouver, British Columbia, Canada: ACM Press. <https://doi.org/10.1145/238386.238511>.
- Henkemans, B., Olivier, A., Van der Pal, S., Werner, I., Neerinx, M.A., & Looije, R. (2017). Learning with Charlie: A robot buddy for children with diabetes. In *Proceedings of the Companion of the 2017 ACM/IEEE International Conference on Human-Robot Interaction*, 406–406.
- Hertzum, M., Borlund, P., & Kristoffersen, K. B. (2015). What do thinking-aloud participants say? A comparison of moderated and unmoderated usability sessions. *International Journal of Human-Computer Interaction*, 31(9), 557–570. <https://doi.org/10.1080/10447138.2015.1065691>.
- Honig, S., & Oron-Gilad, T. (2020). Comparing laboratory user studies and video-enhanced web surveys for eliciting user gestures in human-robot interactions. In *Companion of the 2020 ACM/IEEE International Conference on Human-Robot Interaction*, 248–50. HRI '20. Cambridge, United Kingdom: Association for Computing Machinery. <https://doi.org/10.1145/3371382.3378325>.
- Huber, A., Weiss, A. (2017). Developing human-robot interaction for an industry 4.0 robot: How industry workers helped to improve remote-HRI to physical-HRI. In *Proceedings of the Companion of the 2017 ACM/IEEE International Conference on Human-Robot Interaction*, 137–38. HRI '17. Vienna, Austria: Association for Computing Machinery. <https://doi.org/10.1145/3029798.3038346>.
- International Labour Organization. (2021). ILO Monitor: COVID-19 and the World of Work. 7th Edition. Briefing note. 25 January 2021. [http://www.ilo.org/global/topics/coronavirus/impacts-and-responses/WCMS\\_767028/lang%2D%2Den/index.htm](http://www.ilo.org/global/topics/coronavirus/impacts-and-responses/WCMS_767028/lang%2D%2Den/index.htm).
- Jitsi.Org (2021). Jitsi. 2021. <https://jitsi.org/>.
- Killerby, M.E., Link-Gelles, R., Haight, S.C., Schrodt, C.A., England, L. (2020). Characteristics Associated with Hospitalization Among Patients with COVID-19 — Metropolitan Atlanta, Georgia, March–April 2020. *MMWR. Morbidity and Mortality Weekly Report* 69 (June). <https://doi.org/10.15585/mmwr.mm6925e1>.

- Kim, H., Kim, J., & Lee, Y. (2005). An empirical study of use contexts in the Mobile internet, focusing on the usability of information architecture. *Information Systems Frontiers*, 7(2), 175–186. <https://doi.org/10.1007/s10796-005-1486-z>.
- Laugwitz, B., Held, T., Schrepp, M. (2008). Construction and evaluation of a user experience questionnaire. In *HCI and Usability for Education and Work*, edited by Andreas Holzinger, 63–76. Lecture notes in computer science. Berlin, Heidelberg: Springer. [https://doi.org/10.1007/978-3-540-89350-9\\_6](https://doi.org/10.1007/978-3-540-89350-9_6).
- Lee, K., Lee, K. Y., & Sheehan, L. (2020). Hey Alexa! A magic spell of social glue?: Sharing a smart voice assistant speaker and its impact on users' perception of group harmony. *Information Systems Frontiers*, 22(3), 563–583. <https://doi.org/10.1007/s10796-019-09975-1>.
- Macefield, R. (2009). How to specify the participant group size for usability studies: A practitioner's guide. *Journal of Usability Studies*, 5(1), 34–45.
- Mackey, B.A., Bremner, P.A., Giuliani, M. (2020). The Effect of Virtual Reality Control of a Robotic Surrogate on Presence and Social Presence in Comparison to Telecommunications Software. In *Companion of the 2020 ACM/IEEE International Conference on Human-Robot Interaction*, 349–51. HRI '20. Cambridge, United Kingdom: Association for Computing Machinery. <https://doi.org/10.1145/3371382.3378268>.
- Martinez-Martin, E., & del Pobil, A.P. (2018). Personal robot assistants for elderly care: An overview. In *Personal Assistants: Emerging Computational Technologies*, 77–91. Springer.
- Mumm, J., & Mutlu, B. (2011). Human-robot proxemics: Physical and psychological distancing in human-robot interaction. In *Proceedings of the 6th International Conference on Human-Robot Interaction*, 331–338.
- Nagy, G.M., Young, J.E., Anderson, J.E. (2015). Are tangibles really better? Keyboard and joystick outperform TUIs for remote robotic locomotion control. In *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction Extended Abstracts*, 41–42. HRI'15 extended abstracts. Portland, Oregon, USA: Association for Computing Machinery. <https://doi.org/10.1145/2701973.2701978>.
- Nomura, T., Suzuki, T., Kanda, T., & Kato, K. (2006). Altered attitudes of people toward robots: Investigation through the negative attitudes toward robots scale. In *Proc. AAI-06 Workshop on Human Implications of Human-Robot Interaction*, 2006:29–35.
- Oh, S., & Oh, Y.H. (2019). Understanding the preference of the elderly for companion robot design. In *International Conference on Applied Human Factors and Ergonomics*, 92–103. Springer.
- Papadopoulos, F., Dautenhahn, K., & Ho, W.C. (2013). Behavioral analysis of human-human remote social interaction mediated by an interactive robot in a cooperative game scenario. In *Handbook of Research on Technoself: Identity in a Technological Society*, 637–65. IGI Global.
- Prasov, Z. (2012). Shared gaze in remote spoken HRI during distributed military operation. In *Proceedings of the Seventh Annual ACM/IEEE International Conference on Human-Robot Interaction*, 211–12. HRI '12. Boston, Massachusetts, USA: Association for Computing Machinery. <https://doi.org/10.1145/2157689.2157760>.
- Price-Haywood, E. G., Burton, J., Fort, D., & Seoane, L. (2020). Hospitalization and mortality among black patients and white patients with Covid-19. *New England Journal of Medicine*, 382, 2534–2543. <https://doi.org/10.1056/NEJMsa2011686>.
- Qian, K., Niu, J., & Yang, H. (2013). Developing a gesture based remote human-robot interaction system using Kinect. *International Journal of Smart Home*, 7(4), 203–208.
- Robinson, H., MacDonald, B., Kerse, N., & Broadbent, E. (2013). The psychosocial effects of a companion robot: A randomized controlled trial. *Journal of the American Medical Directors Association*, 14(9), 661–667.
- Rubin, J., & Chisnell, D. (2008). *Handbook of usability testing: How to plan*. John Wiley & Sons.
- Schneider, S., & Kummert, F. (2018). Comparing the effects of social robots and virtual agents on exercising motivation. In *International Conference on Social Robotics*, 451–61. Springer.
- Sierra, S.D., Jiménez, M.F., Múnera, M.C., Frizzera-Neto, A., Cifuentes, C.A. (2019). Remote-operated multimodal Interface for therapists during Walker-assisted gait rehabilitation: A preliminary assessment. In *Proceedings of the 14th ACM/IEEE International Conference on Human-Robot Interaction*, 528–29. HRI '19. Daegu, Republic of Korea: IEEE press.
- Stokes, E.K., Zambrano, L.D., Anderson, K.N., Marder, E.P., Raz, K.M., Felix, S.El B., Tie, Y., & Fullerton, K.E. (2020). Coronavirus disease 2019 case surveillance — United States, January 22–May 30, 2020. *MMWR. Morbidity and Mortality Weekly Report* 69 (June). <https://doi.org/10.15585/mmwr.mm6924e2>.
- Stubbs, K., Wettergreen, D., Nourbakhsh, I. (2008). Using a robot proxy to create common ground in exploration tasks. In *Proceedings of the 3rd ACM/IEEE International Conference on Human Robot Interaction*, 375–82. HRI '08. Amsterdam, the Netherlands: Association for Computing Machinery. <https://doi.org/10.1145/1349822.1349871>.
- Thompson, K.E., Rozanski, E.P., & Haake, A.R. (2004). Here, there, anywhere: Remote usability testing that works. In *Proceedings of the 5th conference on information technology education*, 132–137.
- UNESCO. (2021). UNESCO figures show two thirds of an academic year lost on average worldwide due to Covid-19 school closures. UNESCO. 25 January 2021. <https://en.unesco.org/news/unesco-figures-show-two-thirds-academic-year-lost-average-worldwide-due-covid-19-school>
- United Nations. (2020). UN secretary-General's policy brief: The impact of COVID-19 on women | digital library: Publications. UN Women. 4 September 2020. <https://www.unwomen.org/en/digital-library/publications/2020/04/policy-brief-the-impact-of-covid-19-on-women>.
- Vasalou, A., Ng, B.D., Wiemer-Hastings, P., & Oshlyansky, L. (2004). Human-Moderated Remote User Testing: Protocols and Applications. In *8th ERCIM Workshop, User Interfaces for All, Wien, Austria*. Vol. 19. sn.
- Venkatesh, V., & Davis, F. D. (2000). A theoretical extension of the technology acceptance model: Four longitudinal field studies. *Management Science*, 46(2), 186–204. <https://doi.org/10.1287/mnsc.46.2.186.11926>.
- Walters, Michael L., Kerstin Dautenhahn, René Te Boekhorst, Kheng Lee Koay, Dag Sverre Syrdal, and Chrystopher L. Nehaniv. (2009). An empirical framework for human-robot proxemics. *Procs of New Frontiers in Human-Robot Interaction*.
- Wang, Zijun, Fabian Schmidt, Yiska Weisblum, Frauke Muecksch, Christopher O. Barnes, Shlomo Finkin, Dennis Schaefer-Babajew, et al. 2021. 'MRNA vaccine-elicited antibodies to SARS-CoV-2 and circulating variants'. *BioRxiv*, January, 2021.01.15.426911. <https://doi.org/10.1101/2021.01.15.426911>.
- World Health Organization and others. (2020). *Transmission of SARS-CoV-2: Implications for infection prevention precautions: Scientific brief, 09 July 2020*. World Health Organization.
- Xue, C., Qiao, Y., Murray, N. (2020). Enabling human-robot-interaction for remote robotic operation via augmented reality. In *2020 IEEE 21st International Symposium on 'A World of Wireless, Mobile and Multimedia Networks' (WoWMoM)*, 194–96. <https://doi.org/10.1109/WoWMoM49955.2020.00046>.
- Yanco, H.A., Baker, M., Casey, R., Chanler, A., Desai, M., Hestand, D., Keyes, B., & Thoren, P. (2005). 'Improving human-robot interaction for remote robot operation'. In *AAAI*, 5:1743–1744.
- Zenbo | Intelligent Robot. (n.d.) ASUS Global. <https://www.asus.com/Commercial-Intelligent-Robot/Zenbo/>

Zhao, Z., & McEwen, R. (2021). 'Luka Luka - investigating the interaction of children and their home Reading companion robot: A longitudinal remote study'. In *Companion of the 2021 ACM/IEEE International Conference on Human-Robot Interaction*, 141–43. HRI '21 companion. New York, NY, USA: Association for Computing Machinery. <https://doi.org/10.1145/3434074.3447146>.

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Dr. Curtis L. Gittens** is a Lecturer at the University of the West Indies - Cave Hill Campus and has been in the information technology field for over twenty-five years. He has extensive industrial research and development experience in computer science within several verticals including healthcare, utilities, eLearning and eBusiness. Dr. Gittens is a researcher in human-robot interaction, intelligent virtual agents, psychologically realistic cognitive architectures, artificial societies, and web-based software architectures. His research has led to publications in journals and conference proceedings and presentations at several international conferences. As part of his ongoing research efforts, Dr. Gittens is working on developing psycho-social robotic surrogates