de Araujo et al., Gerontechnology (2024) 23:1
https://doi.org/10.4017/gt.2024.23.1.1027.11

Social robot-based depression screening in older adults: A pilot study

Bruno Sanchez de Araujo MSc[a], Marcelo Fantinato PhD[a,*], Meire Cachioni PhD[a], Mônica Sanches Yassuda PhD[a], Ruth Caldeira de Melo PhD[a], Sarajane Marques Peres PhD[a], Patrick C. K. Hung PhD[b]

*[a]School of Arts, Sciences and Humanities, University of São Paulo, São Paulo, Brazil; [b]Faculty of Business and Information Technology, Ontario Tech University, Oshawa, Canada; *Corresponding author: m.fantinato@usp.br*

Abstract

Background: Depression in older adults is a prevalent issue that can lead to severe consequences including a decline in overall health and even suicide. Early detection and management of depression are crucial for preventing such outcomes. The integration of technology solutions in healthcare represents a promising ap-proach to support prevention, diagnosis, and continuous monitoring of patients.
Research aim: This pilot study aims to evaluate the feasibility of depression screening in older adults through interactions facilitated by social robots, focusing on individuals without severe cognitive impair-ments.
Methods: The study involved five older adults with a minimum score of 24 on the Montreal Cognitive As-sessment (MoCA), ensuring no significant cognitive impairment. The Geriatric Depression Scale (GDS-15) was used as the screening tool. Participants interacted with a social robot and a healthcare professional in alternating sequences for the administration of the GDS-15. Additional assessments using the Positive and Negative Affect Schedule (PANAS) and the Godspeed questionnaire series were conducted to evaluate emo-tional responses and perceptions towards the social robot. Notably, MoCA, PANAS, and Godspeed were not administered by the social robot.
Results: Preliminary data showed that all participants fell within the same depression range when screened by both the social robot and the healthcare professional. The results indicated no adverse effects on partici-pants' emotional states post-interaction with the social robot, as evidenced by PANAS scores. The Godspeed questionnaire revealed that participants generally had a positive perception of the social robot.
Conclusions: The findings suggest that social robots can effectively perform depression screening in older adults without severe cognitive impairments. Their use matches the assessment outcomes of healthcare professionals and does not negatively impact emotional states, indicating their potential as a feasible and positively perceived tool for early depression diagnosis and continuous monitoring.

Keywords: social robots, older adults, elderly, depression, screening.

Introduction

Worldwide, around 5.7% of adults aged 60 years or older have received a diagnosis of depression, amounting to a total of 57.7 million individuals, as per a 2019 report from the Global Health Data Exchange (IHME, 2019). The rise in the percentage of older adults relative to the overall population necessitates a robust response from public health systems. Nevertheless, the debate centered on policies aimed at fostering healthy aging has not been effective, with the extended lifespan not consistently translating into improved health conditions (WHO, 2015).

Depression is a prevalent concern among the older adult population worldwide, and unfortunately, it often goes undiagnosed and untreated (Birrer and Vemuri, 2004; Minallah et al., 2019). This lack of recognition and care can significantly worsen the condition (American Psychiatric Association, 1994; Cuijpers et al., 2004; Wells et al., 1992). Furthermore, severe depression in older adults can elevate the risk of suicide (Minallah et al., 2019), with rates twice as high as in the general population (Alexopoulos et al., 2001). These statistics underscore the importance of finding more effective methods to screen depression in older adults at the earliest stages possible.

Recent advancements in gerontechnology, as outlined by Colnar et al. (2020), emphasize technology's potential to aid various facets of older adults' lives. They emphasize the crucial role of technology, including smart homes, wireless sensor networks, and data analytics, in promoting independence and mitigating health-related risks among older adults. Abdi et al. (2018) spotlight socially assistive robots as a prominent technology in older adult care, spanning areas such as affective therapy, cognitive training, social facilitation, companionship, and physiological assistance. These robots offer a promising avenue for enhancing the overall well-being and quality of life for older adults.

Estimations suggest that the deployment of social robots has the potential to decrease labor costs by up to 65% while simultaneously improving the quality of healthcare services (Eggleston and Lee, 2020). Moreover, social robots can function as valuable assistants, alleviating caregivers' burden level and empowering them to concen-trate on higher-value tasks beyond the capabilities of robots, such as the development of a personalized care plan (Ting et al., 2021; Asprino et al., 2019; Fiorini et al., 2019). These robots are currently undergoing testing in diverse healthcare settings (Lewis, 2014). Nevertheless, they face constraints in terms of their interactive capabilities, which pose a computational challenge in expanding their functionality to address specific issues more effec-tively (Pu et al., 2019).

In developing economies such as Brazil, there is a notable shortage of studies focused on addressing depression among older adults using social robots (de Araujo et al., 2022). Specifically, in the context of Brazil, there is an evident absence of studies with social robots tailored to the Brazilian Portuguese language, as highlighted in the available literature (de Araujo et al., 2022). Additionally, an important consideration in these economies is constrained access to costly resources like physical, social robots, which might promote the adoption of virtual robot-based solutions, such as avatars.

While there is a recognized gap in research on social robots for depression screening in older adults, ongoing studies have explored their application in assessing other health aspects such as cognitive and physical abilities, and even symptoms of Covid-19. These studies have involved implementing various health scales and tests on social robots, ranging from custom-designed tests to the use and adaptation of established scales (Rossi et al., 2018; García-Olaya et al., 2019; Kobayashi et al., 2019; Naranjo-Saucedo et al., 2019; Varrasi et al., 2019; Cor-mons et al., 2020; Chen et al., 2021; Do et al., 2021; Mucchiani et al., 2021; Ting et al., 2021; Chang et al., 2022). Among these, only García-Olaya et al. (2019), Naranjo-Saucedo et al. (2019), and Ting et al. (2021) pre-cisely implemented the intended scale—the Barthel scale—which assesses functional abilities in older adults. Moreover, the few studies that compared results between robot-facilitated and human-facilitated methods did not employ the same scale on the social robot as that used by the human agent, or the implementation on the social robot was merely based on the human application but was not identical. Notably, only Naranjo-Saucedo et al. (2019) administered the same Barthel scale through both the social robot and the human agent. However, the authors concluded that the methods—robot and human—could not be considered equivalent, a determination made through qualitative analysis. Conversely, Kobayashi et al. (2019) found both methods to be equivalent through quantitative analysis, despite the version implemented on the robot differing from that used by the human agent. In the context of emotional assessments, research is even more limited. Only one study has focused specifically on evaluating anxiety, stress, and depression using the DASS-21 scale. However, it is important to note that this study did not include older adults in its participant group (Nandanwar and Dutt, 2023).

This pilot study primarily aimed to verify the feasibility of measuring depression levels in older adults using a social robot. This feasibility verification involved comparing the results obtained from administering the same depression screening tool by both a social robot and a healthcare professional five to ten days apart. The Geriat-ric Depression Scale (GDS-15) was employed as the depression screening tool. To ensure that participants would understand the questions from the instruments used, our study was restricted to older adults without severe cognitive impairments, as assessed by the Montreal Cognitive Assessment (MoCA) (Nasred dine et al., 2005). We also investigated whether there were noticeable changes in the emotional states of older adults when subjected to each of the two depression screening administration methods, measured by the Positive and Negative Affect Schedule (PANAS) (Watson et al., 1988; Carvalho et al., 2013). Finally, this pilot study sought to gauge the perceptions of older adults regarding a social robot by the Godspeed questionnaire series (Bartneck et al., 2009). MoCA, PANAS, and Godspeed were not administered by the social robot. GDS-15 was the only task administered by both the social robot and the healthcare professional, whereas MoCA, PANAS, and Godspeed were consistently conducted exclusively by human agents. Given the exploratory nature of this research, a small sample size was appropriate to gather preliminary data and identify potential trends. This approach is typical for pilot studies aimed at inform-ing the design of larger, more comprehensive research projects (Thabane et al., 2010; Pearson et al., 2020).

The remainder of this paper is organized as follows. Section 2 presents the methods and instruments. Section 3 shows the results obtained from this pilot study. Section 4 presents a preliminary analysis of these results. Finally, Section 5 concludes this paper with future directions based on the results of this pilot study.

Methods

This section outlines the methods and instruments employed in conducting the pilot study.

Robot

The social robot selected for this project is the Robios robot, developed by Human Robotics . Robios is an autonomous robot produced in Brazil, offered in two versions: (i) a physical embodiment, standing at a height of approximately 1.22 meters, with the capability to move and detect objects; and (ii) an avatar designed for virtual interaction, mirroring the head design of the physical model, and functioning on a tablet screen. Both versions are visible in *Figure 1*. For this pilot study, we used the avatar version, to maintain a simpler environmental setup for the study.

Robios has autonomous and independent interaction capabilities, enabling communication in Brazilian Portuguese. Additionally, it can autonomously detect individuals and movements, as well as initiate interactions proactively. Equipped with voice recognition capabilities and an interactive screen, it empowers users to input data directly into the robot. Moreover, Robios is proficient in object detection, room mapping, and navigation. In the forthcoming phases following this pilot study, we plan to introduce an embodied version to further explore and har-ness body movement, in addition to the capabilities related to object detection, room mapping, and navigation.

Figure 1. Embodied and avatar versions of the Robios robot.

Geriatric depression scale

The Geriatric Depression Scale (GDS) (Yesavage, 1988), created in 1982, provides a straightforward strategy to screen depression symptoms through a yes/no questionnaire designed specifically for older adults. The GDS con-sists of 30 questions, with each response potentially contributing one point to the total score depending on the provided response (yes/no) and the specific question. For some questions, the answer "yes" is scored as a depression symptom, while for others, the answer "no" is scored as a depression symptom. After completing the questionnaire, the total score derived from older adults' responses is categorized based on depression symptoms into three ranges representing unlikely depression, possible depression, and likely present depression (Yesavage, 1988). This can serve as a valuable diagnostic support tool for expedited depression screening.

In 1986, a shorter version of the GDS with 15 questions was proposed (GDS-15) (Sheikh and Yesavage, 1986). Its purpose was to reduce the screening time for the patient by focusing on the 15 questions that best capture depression symptoms, as indicated by the original 30-question validation study. The 15 questions comprising the GDS-15 are listed in *Table 1*. In the table, "Scoring criteria" refers to the response option (i.e., yes or no) that indicates the presence of a respective depression symptom and hence, if chosen, results in the accumulation of one point in the total score. The score ranges and levels of the GDS-15 are presented in *Table 2*. The GDS-15 was selected for this study to be implemented in the social robot.

The use of the GDS and GDS-15 does not supplant a clinical assessment performed by psychiatrists or psychologists. However, these tools have emerged as a commonly employed screening method for identifying potential cases that should be assessed by a specialist (Montorio and Izal, 1996; Stiles and McGarrahan, 1998). The GDS and GDS-15 are currently in the public domain and have been translated into various languages, including Brazilian Portuguese (Almeida and Almeida, 1999), the language selected for this study.

Participants

Individuals aged 60 or older were eligible to participate in this pilot study. Gender was not a

Table 1. GDS-15 with the response that scores for the tool (Sheikh and Yesavage, 1986) ('X' refers to the answer indicating a depression symptom)

ID	Question	Scoring criteria	
		Yes	No
1	Are you basically satisfied with your life?		X
2	Have you dropped many of your activities and interests?	X	
3	Do you feel that your life is empty?	X	
4	Do you often get bored?	X	
5	Are you in good spirits most of the time?		X
6	Are you afraid that something bad is going to happen to you?	X	
7	Do you feel happy most of the time?		X
8	Do you often feel helpless?	X	
9	Do you prefer to stay at home, rather than going out and doing new things?	X	
10	Do you feel you have more problems with memory than most?	X	
11	Do you think it is wonderful to be alive now?		X
12	Do you feel pretty worthless the way you are now?	X	
13	Do you feel full of energy?		X
14	Do you feel that your situation is hopeless?	X	
15	Do you think that most people are better off than you are?	X	

factor in the selec-tion process. The older adults invited to the study should ideally have a prior diagnosis of at least mild depression to avoid floor effects on the scale and produce variability in the responses obtained. This was ascertained by the question, "Has any doctor told you in the last six months that you may have depression?"

Individuals with cognitive impairment or dementia who might fail to understand the protocol's questions were not included in the study, as this could lead to inaccurate data capture due to memory, language, or time and space orientation difficulties. The Montreal Cognitive Assessment (MoCA) (Nasreddine et al., 2005) was employed by the healthcare professional to assess cognitive impairment. Individuals with a MoCA score of 23 points or lower (out of a maximum 30-point score) were excluded from the study (Carson et al., 2018).

Participants were recruited from a university program for older adults at a Brazilian university. This study involved five older adults, which is typical for pilot studies focused on feasibility. The small sample size was chosen due to the exploratory nature of the research, resource constraints, and the need to ensure participant safety and comfort (Thabane et al, 2010; Pearson et al., 2020). They were presented with an Informed Consent Form, which outlined the minimal risks inherent to the experiment , such as potential fatigue and stress due to the nature of the questions included in the GDS-15 and the MoCA, in addition to other questionnaires presented in the next section. To minimize risks of fatigue and stress, participants were informed that they could interrupt their participation at any time if they wished. Furthermore, given the personal nature of the questions posed during the GDS-15 screening, participants were assured the privacy and impersonal handling of their information.

Procedure and design

Each participant engaged in two sessions, whose steps are described below. The sessions and respective steps are described in more detail as follows.

Health professional-facilitated GDS-15 screening:
1. Administration of the PANAS questionnaire.
2. Administration of a sociodemographic questionnaire and a questionnaire about the presence of common diseases in older adults.
3. Administration of the MoCA test.
4. Implementation of the GDS-15 screening (human-facilitated).
Administration of the PANAS questionnaire.

Social robot-facilitated GDS-15 screening:
1. Administration of the PANAS questionnaire.
2. Familiarization of the participants with the robot through a gaming app.
3. Implementation of the GDS-15 screening (robot-facilitated).
4. Administration of the Godspeed questionnaire series.
5. Administration of the PANAS questionnaire.

All participants were required to engage in two assessment sessions. One session involved the application of the GDS-15 by the social robot, while a healthcare professional facilitated the other session, both using the GDS-15 in Portuguese. The participants were divided into two groups based on convenience sampling. Three participants engaged with the social robot in the first session and the healthcare professional in the second session, while the remaining two participants followed the reverse sequence, engaging with the healthcare professional in the first session and with the social robot in the second session.

Table 2. GDS-15 score ranges and levels (Sheikh and Yesavage, 1986).

Score range	Level (or grade)	Interpretation
0 to 5	Unlikely depression	Not suggestive of depression; a follow-up comprehensive assessment is not required.
6 to 10	Possible depression	Suggestive of depression; a follow-up comprehensive assessment is warranted.
11 to 15	Likely present depression	Almost always indicative of depression; a follow-up comprehensive assessment is warranted.

The time interval between the two sessions ranged from five to ten days. This range was adopted to strike a balance between an interval that is neither too short (to prevent participants from recalling the questions and responses from the first session) nor too long (to prevent participants' depression levels from changing due to factors external to the study). Each session lasted approximately 30 minutes.

An app was implemented for the administration of the GDS-15 through the social robot. This app aims to make the robot simulate the behavior of a healthcare professional as closely as possible while administering the same questionnaire. *Figure 2* illustrates the flow of actions followed by the robot during the interaction with the participant. It allows the participant to stop asking questions and quit the experiment at any time.

A game-style application (app) implemented on the social robot was used to allow familiarization of the participants with the robot before the screening itself. In this game, a segment of a classic popular song's music video was first displayed on the robot's screen, followed by the robot asking some questions about the song's lyrics. This app was used to help participants understand how the interaction with the robot takes place and to reduce any resistance that participants might have due to potential fear of the robot. We made a concerted effort to select songs that were as neutral or cheerful as possible, understanding that these are well-known, older songs. To accommodate individual preferences and avoid any negative associations, we offered participants a choice of four different songs: "Garota de Ipanema" (by Tom Jobim, 1962), "Trem das Onze" (by Adoniran Barbosa, 1964), "O Calhambeque" (by Roberto Carlos, 1966), and "A Banda" (by Chico Buarque, 1966). Consequently, we did not anticipate that this familiarization activity could potentially affect their depression scores.

The Positive and Negative Affect Schedule (PANAS) (Watson et al., 1988; Carvalho et al., 2013) questionnaire was administered by the human agents both as the first step and as the last step of each session, whether the GDS-15 facilitated by the social robot or the healthcare professional (four applications). The PANAS was used to evaluate potential changes in participants' emotional states over time, considering before and after interaction with either the social robot or the healthcare professional. Twenty descriptors are considered for analysis—ten for positive affects (interested, excited, strong, enthusiastic, proud, alert, inspired, determined, attentive, and active) and ten for negative affects (distressed, upset, guilty, scared, hostile, irritable, ashamed, nervous, jittery, and afraid).

The Godspeed questionnaire series (Bartneck et al., 2009) was administered by the human agent after the social robot-facilitated GDS-15 screening. The Godspeed was used to evaluate participants' perceptions of the interaction with the robotic agent. Twenty-three features are considered for analysis, categorized into five domains: anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety.

A questionnaire for collecting sociodemographic data was administered before the health professional-facilitated GDS-15 screening. This questionnaire collected data about age, gender, educational background, marital status, who the person lives with, and the number of medications taken in the last three months. A questionnaire on pre-existing diagnoses of common diseases in older adults (such as hypertension, stroke, diabetes, cancer, arthritis, bronchitis, depression, and osteoporosis) was also administered before the health professional-facilitated GDS-15 screening. It aimed to identify, among possible diseases listed through self-reporting, a previous diagnosis of depression.

As introduced in the previous section, the Montreal Cognitive Assessment (MoCA) was employed to assess cognitive impairment as one of the steps before the health professional-facilitated GDS-15 screening. A threshold of at least 24 points (out of a maximum of 30 points) was used to not exclude the participants (Carson et al., 2018). All questionnaires were administered in Portuguese, including versions validated for this language.

All sessions occurred in the same setting, specifically in the social robotics laboratory. During each session, the older adult was accompanied by either the healthcare professional (for the healthcare professional-facilitated GDS-15 screening) or a researcher specialized in information systems (IS) (for the social robot-facilitated GDS-15 screening). The healthcare professional managed all questionnaires administered during their sessions, which included the GDS-15. In contrast, for the social robot-facilitated GDS-15

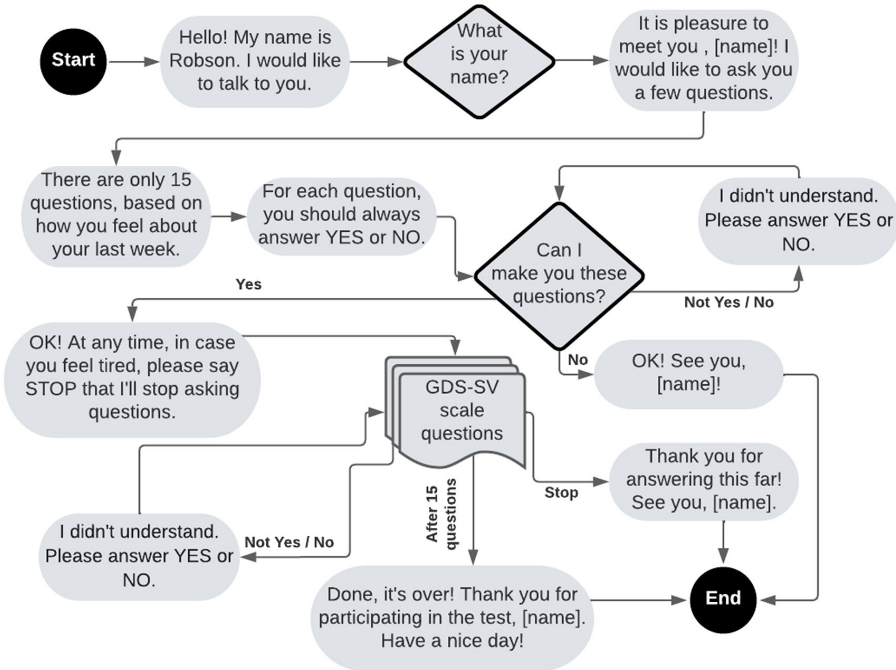

Figure 2. Robot action flow (the sentences for the dialogue were translated from the original Portuguese)

screening, the IS researcher was respon-sible for conducting the PANAS and Godspeed questionnaires and introducing the robot to the participants through the gaming app. Concerning the administration of the GDS-15, the IS researcher prepared the robot for its use and then left the room, allowing the older adult to complete the questionnaire exclusively in the robot's pres-ence. The IS researcher closely monitored the GDS-15 administration from a separate room, using a system log to promptly address any technical issues or aid if required. The sessions took place between May 9th and May 30th, 2023. The following section provides a detailed presentation of the collected data, as well as an analysis and discussion of these data.

It is important to note that this pilot study did not include a longitudinal follow-up due to time and resource con-straints. The primary aim was to gather preliminary data on the feasibility of using social robots for depression screening in older adults. Future research will address this limitation by incorporating long-term follow-up to evaluate the sustained effectiveness of robotic screening.

Results

Table 3 summarizes the most relevant data obtained in this study. First, the table presents the results from the application of the GDS-15, with the social robot and the healthcare professional. The asterisk (*) denotes the initial GDS-15 screening type used for each participant. Then, the key sociodemographic data are listed. Finally, health-related data are presented, including the MoCA score and the prior diagnosis of depression. Two participants had identical GDS-15 scores when applied by the social robot or the health professional. For three participants, scores were numerically higher for the application with the health professional. Yet, even for these three participants, the score ranges did not differ.

Table 4 summarizes the results obtained from the administration of the GDS-15 for each of the five participants for each GDS-15 question. The symbols 'x' and '+' mark the GDS-15 questions whose responses contributed 1.0 point to the total score of each participant. The last row shows the cumulative scores for each participant, con-sidering both the GDS-15 administration facilitated by a social robot and that facilitated by a healthcare profes-sional. The 'x' symbol is consistently displayed in pairs, indicating a consensus in the results obtained through the social robot and the healthcare professional. In contrast, the '+' symbol denotes a discrepancy between the two methods, reflecting a result obtained exclusively by either the social robot or the healthcare professional.

Table 5 summarizes the results of the PANAS questionnaire administered before and after both types of GDS-15 screening. The results are sum-

Table 3. Overall results from the application of the GDS-15, including the GDS-15 results for both social robot and healthcare professional-facilitated methods, as well as demographic and medical data

ID	GDS-15 (score/level)		Demographic data				Medical data	
	Social robot	Healthcare professional	Age	Gender	Lives alone	Education	MoCA	Depression diagnosis
1	0 (unlikely)	2 (unlikely)*	61	Female	No	College	30	No
2	2 (unlikely)	3 (unlikely)*	66	Female	No	Postgraduate	29	No
3	6 (possible)	8 (possible)*	64	Female	No	Postgraduate	30	Yes
4	6 (possible)*	6 (possible)	75	Female	Yes	College	29	No
5	0 (unlikely)*	0 (unlikely)	62	Female	Yes	College	28	No

marized in *Table 5*, considering both the mean and standard deviation ($\mu \pm \sigma$) as well as the median and interquartile range ($\tilde{X}$, Q3 − Q1). The PANAS questionnaire consists of ten positive affect descriptors, each rated on a scale from 1 to 5. Consequently, participants' positive affect scores can range from 10 to 50, with higher scores indicating greater positive affect. Likewise, there are ten negative affect de-scriptors, also rated from 1 to 5, yielding a potential score range of 10 to 50 for negative affect. Lower scores in this context reflect reduced levels of negative affect. While the data is ordinal, the original article's authors (Watson et al., 1988) summarize it using both mean and standard deviation. They additionally aggregate the scores of the ten positive and ten negative descriptors. Summing scores is indeed a common method to simplify and aggregate ordinal data, particularly when aiming to create a composite score representing an underlying con-struct, thereby facilitating interpretation. This approach is predominantly employed in cases like this one, where the data approximates a symmetric and normal distribution without the presence of outliers. Moreover, the differences between scale categories are relatively stable or proportional, revealing a clear order, and the distances between them are reasonably consistent. Nevertheless, the median and interquartile range are also presented for more ro-bust statistical analyses, though they are based on the sums of the ten items (positive or negative) for each case.

Table 6 lists the results of the Godspeed questionnaire series administered immediately after the social robot-facilitated GDS-15 screening. Twenty-three items were employed in the evaluation, categorized into five do-mains: anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety. For each item, participants were required to select a rating score in the range from 1 to 5, where 1 corresponds to the feature listed in the first column, and 5 corresponds to the feature listed in the last column. The obtained values are pre-sented as percentages in the table. For instance, for the first item (Fake vs. Natural), 40% of the participants responded 1 (i.e., consider the robot totally fake), 20% responded 2 (i.e., somewhat fake), 40% responded 3 (i.e., neither fake nor natural), and no participant considered the robot somewhat or totally natural. One of the participants declined to respond to the items in the perceived intelligence domain as they believed the interaction with the robot was too limited to evaluate that aspect.

Discussion

Initial findings and insights

As shown in *Table 3*, the depression levels according to the GDS-15 score ranges for each participant remain con-sistent across the GDS-15 screenings facilitated by both the social robot and the healthcare professional. While this study involved a small sample size of five participants, this is typical for pilot studies. The primary aim was to gather preliminary data and assess the feasibility of using social robots for depression screening. Despite the small sample, the consistency of the GDS-15 scores between the robot and healthcare professional screenings suggests promising initial evidence of the method's effectiveness. All participants fell within the same score range for both methods: participants 1, 2, and 5 were classified as unlikely depression in both GDS-15 screenings, while participants 3 and 4 were classified as possible depression in both cases as well. This consistent pattern suggests that the administration of the GDS-15 by the social robot yielded similar results to those obtained by the healthcare professional. Although the overall consistency is apparent, there is some variability in the GDS-15 scores among participants. Participants 1, 2, and 3 scored up to 2 points higher when assisted by the healthcare professional. It is noteworthy that the GDS-15 screening with the health professional was the second one for these three cases. This slight variability might be attributed to individual differences in the perception of the interaction with the social robot or the healthcare professional. The only previous diagnosis of depression, referring to Partici-pant 3, was captured by both GDS-15 screening methods. As the five participants scored very high on the MoCA test, cognitive impairment would not be a potential issue for interacting with technology.

The data for each question, as detailed in *Table 4*, shows variations in participant responses be-

Table 4. Breakdown of GDS-15 data organized both by participant and by question

ID	Participant 1		Participant 2		Participant 3		Participant 4		Participant 5	
	Social robot	Hlth. Profes.	Social robot	Hlth. Profes.	Social robot	Hlth. Profes.	Social robot	Hlth. Profes.	Social robot	Hlth. Profes.
1				+	x	x	x	x		
2							+			
3					+		x	x		
4		+	x	x		+				
5					x	x		+		
6		+	x	x	x	x	+			
7					x	x	+			
8					x	x		+		
9						+				
10								+		
11										
12							x	x		
13										
14						+				
15										
∑	0	2	2	3	6	8	6	6	0	0

tween the social robot and healthcare professional GDS-15 screenings. Some questions, like questions 4 and 6, show discrepancies in the responses of at least two participants between the two GDS-15 screenings. This suggests that certain questions might be perceived differently when administered by a social robot compared to a human healthcare profes-sional. There is also participant-specific variability in how their responses align between the two GDS-15 screen-ing methods. For example, participants 2 and 5 had consistent responses for all or most questions, while participants 3 and 4 showed more variations, mainly in questions related to mood and emotional well-being. This variabil-ity highlights the importance of understanding the role of human-robot interaction in depression screening. While designed to be empathetic and supportive, social robots may not capture all nuances of human interaction, poten-tially affecting the way participants express their emotions and feelings. In addition, certain questions in the GDS-15 may be particularly sensitive to variations in screening. For example, questions relat-ed to social engage-ment or mood changes might elicit different responses depending on whether the interaction is with a social robot or a human.

Regarding the PANAS results (*Table 5*), the baseline scores indicate that participants already had relatively good emotional well-being before the GDS-15 screenings, as evidenced by their elevated positive affect scores and low negative affect scores, regardless of the GDS-15 screening method. Both GDS-15 screening methods resulted in an improvement in participants' emotional well-being, which is noteworthy, considering that GDS-15 was origi-nally designed to screen depression levels rather than to influence emotional states. Furthermore, it is interesting to highlight that the improvements in participants' emotional well-being were slightly more pronounced following interactions facilitated by the social robot compared to those led by the healthcare professional. This could be attributed to participants potentially feeling more at ease and less judged during their interactions with the social robot. The novelty of engaging with a robot might have also contributed to these observed differences. These find-ings raise intriguing questions about the potential advantages of human-robot interactions within healthcare settings. However, it is essential to acknowledge that this study consisted of single-session GDS-15 screenings for each method, and further research is needed to explore the long-term effects of such interactions.

The results of the Godspeed questionnaire series (*Table 6*) indicated that participants had an overall positive perception of the social robot used in the study. The detailed data indicate that participants tended to perceive the robot as more machine-like, artificial, and fake rather than human-like, lifelike, and natural. This suggests that the robot's design and behavior may not have successfully created a strong anthropomorphic impression, which could be due to limitations in its appearance or interactions, although it is not necessarily an option to want the robot to get closer to a human. In terms of animacy, participants balanced between characteristics such as apathetic and responsive, as well as inert and interactive, showing room to improve their inter-

Table 5. PANAS means and standard deviations plus medians and interquartile ranges for each test administration moment

Types of affect	Administration moment	n	Healthcare professional		Social robot	
			μ±σ	$\tilde{X}$, Q3-Q1	μ±σ	$\tilde{X}$, Q3-Q1
Positive affects	Before the GDS-15 screening	5	38.4±4.1	39, 3	37.0±2.5	38, 5
	After the GDS-15 screening	5	39.6±2.0	41, 2	40.0±1.5	41, 1
Negative affects	Before the GDS-15 screening	5	18.8±7.3	16, 4	17.0±4.3	17, 6
	After the GDS-15 screening	5	17.8±6.3	16, 10	14.2±3.5	13, 2

action capacity. Participants had very positive perceptions of the robot's likeability, including friendliness, kindness, pleasantness, and niceness. This indicates that the robot was generally well-received and elicited positive emotions. Participants had good positive perceptions of the robot's intelligence, mainly in terms of competence and intelligence itself. This suggests that the robot was seen as intelligent and capable of fulfilling its role effectively. Finally, participants generally felt emotionally stable (relaxed and calm) towards the robot, indicating they did not see it as a threat.

Ethical and emotional considerations

The integration of social robots into healthcare, particularly for sensitive tasks such as depression screening, ne-cessitates a careful examination of ethical and emotional considerations. Privacy and informed consent are para-mount when using robots for depression screening. Participants must be fully aware of how their data will be used, stored, and protected. Ensuring transparency about the robot's capabilities and limitations is crucial to maintain-ing trust (Vitale et al., 2018). The emotional effects of interacting with social robots can be both positive and negative. On the one hand, robots can provide a non-judgmental and supportive presence, potentially making participants feel more comfortable and less stigmatized when discussing sensitive topics. On the other hand, there is a risk that robots might not fully capture the nuances of human emotions, which could affect the accuracy of the screening and the emotional experience of the participants (García et al., 2020). It is important to consider the design and functionality of the robot to maximize its empathetic capabilities while being mindful of these limitations.

Social robots can reduce the burden on healthcare professionals, provide consistent and standardized screening, and offer companionship that can alleviate feelings of loneliness and isolation (Sharkey and Sharkey, 2012). These benefits highlight the potential for robots to enhance healthcare delivery, particularly in settings where human resources are limited. However, the deployment of social robots must be approached with caution. Poten-tial challenges include ensuring the emotional safety of participants, addressing any feelings of discomfort or unease, and managing the ethical implications of robotic interactions (Felzmann, 2020). Ongoing evaluation and adaptation of the technology are necessary to mitigate these challenges and improve the effectiveness of social robots in healthcare. By addressing these ethical and emotional considerations, we aim to provide a more com-prehensive understanding of the potential challenges and benefits associated with using social robots for depression screening in older adults. Future research should continue to explore these aspects to ensure the responsible and effective use of robotic technologies in healthcare.

While the benefits of human touch in healthcare are well-documented, providing comfort, reducing anxiety, and fostering emotional well-being through the release of oxytocin, it is not without its complexities. Studies, such as those by Field (2010) and Linden (2015), have highlighted the positive impacts of human touch. However, the perception of touch can vary significantly among individuals, particularly among those with past traumas or dis-comfort with physical contact, potentially exacerbating stress and depressive symptoms. Conversely, social robots offer a non-invasive alternative for emotional support and interaction, which can be particularly beneficial for individuals uncomfortable with human touch. Research by Pu et al. (2019) and Chen et al. (2021) indicates that social robots can improve emotional well-being and reduce feelings of loneliness. Pu et al. (2019) provide a systematic review and meta-analysis, presenting a broad overview of the effectiveness of social robots in enhancing emotional well-being among older adults, while Chen et al. (2021) offer insights into the practical applications of social robots in healthcare settings. Nonetheless, robots may lack the nuanced empathy inherent in human interactions, which can limit their effectiveness in fully replicating the benefits of human touch. Moreover, it is im-portant to acknowledge the possibility that interactions with robots, if perceived as inadequate or impersonal, could potentially increase feelings of depression in some older adults. This presents a complex issue that requires careful consideration and analysis. While it is not the primary focus of this study, a brief exploration of these factors underscores the need for further research to balance

Table 6. Godspeed questionnaire series results

Rating score	1	2	3	4	5	Rating score
Anthropomorphism						
Fake	40%	20%	40%	–	–	Natural
Machine-like	40%	40%	20%	–	–	Human-like
Unconscious	–	40%	20%	40%	–	Conscious
Artificial	20%	40%	40%	–	–	Lifelike
Animacy						
Dead	–	20%	40%	40%	–	Alive
Stagnant	–	40%	40%	20%	–	Lively
Mechanical	–	80%	–	20%	–	Organic
Artificial	20%	40%	20%	20%	–	Lifelike
Inert	–	20%	60%	20%	–	Interactive
Apathetic	–	20%	60%	20%	–	Responsive
Likeability						
Dislike	–	–	20%	20%	60%	Like
Unfriendly	–	–	–	80%	20%	Friendly
Unkind	–	–	20%	40%	40%	Kind
Unpleasant	–	–	20%	20%	60%	Pleasant
Awful	–	–	20%	20%	60%	Nice
Perceived intelligence						
Incompetent	–	–	25%	50%	25%	Competent
Ignorant	–	25%	50%	–	25%	Knowledgeable
Irresponsible	–	–	75%	–	25%	Responsible
Unintelligent	–	–	75%	25%	–	Intelligent
Foolish	–	25%	50%	–	25%	Sensible
Perceived safety						
Anxious	–	–	20%	20%	60%	Relaxed
Agitated	–	20%	–	20%	60%	Calm
Surprised	20%	–	20%	20%	40%	Quiescent

the emotional needs and comfort levels of patients in depression screening and management.

Conclusion

This pilot study provides preliminary evidence that social robots may be viable for administering the GDS-15 for depression screening in older adults. The alignment of scores between screenings conducted by the social robot and healthcare professionals suggests that social robots have the potential to automate the process of depression screening. However, variations in GDS-15 scores among participants suggest that responses to robot-facilitated screenings may differ from those conducted by human professionals. Further research with larger samples is nec-essary to understand the factors influencing these variations, which could optimize the deployment of social robots in depression screening.

These results underscore the utility of social robots as tools for depression screening in older adults, demonstrating that they can potentially enhance the accessibility and efficiency of healthcare services. Additionally, variations in participant responses at the question level between the two screening methods suggest that certain aspects of the GDS-15 might need refinement or further validation for robot-facilitated administration. The observed discrepancies highlight the influence of human-robot interaction on the perception of depression symptoms, suggesting a need for further exploration into the psychological factors driving these differences.

Despite the promising findings, it 'is important to acknowledge that these were derived from a small sample size of only five participants. A larger study is planned to validate these results comprehensively before progressing to broader research applications. Moreover, the Positive and Negative Affect Schedule (PANAS) data suggests that screenings facilitated by social robots had a more pronounced positive impact on the emotional well-being of participants compared to those conducted by healthcare professionals. Although the GDS-15 is primarily designed for depression screening, the observed changes in affect scores point to an improvement in emotional well-being, particularly with the aid of social robots.

The Godspeed questionnaire series offered insights into the participants' perceptions of the social robot. Although only the avatar version was used, participants typically perceived the robot as machine-like, artificial, and fake, yet friendly, kind, pleasant, intelligent, responsible, and non-threatening. This dual perception indicates that while the robot lacks pronounced human-like qualities—which may not be entirely desirable—it was nonetheless well-received and seen as capable and emotionally supportive. Improvements in anthropomorphism and a reduction in machine-like qualities could further enhance the robot's effectiveness in future healthcare applications.

Ultimately, these preliminary findings are insightful for understanding how social robots could be further integrat-ed into mental health assessments, requiring additional research to establish the consistency of these results across multiple sessions and to explore the long-term effects of human-robot interactions on the emotional well-being of older adults.

Future research will focus on expanding the sample size to include a more diverse population of older adults, incorporating longitudinal follow-up to assess the long-term effectiveness of social robots in depression screening, and conducting comparative studies with traditional screening methods. These steps will build on the initial feasibility demonstrated in this pilot study and contribute to a better understanding of the role of social robots in supporting the mental health and well-being of older adults. Moreover, while this pilot study focused on quantitative assessments, future research will incorporate qualitative methods, such as interviews and focus groups, to provide a more comprehensive understanding of participants' emotional responses and experiences. Combining qualitative and quantitative data will enhance the depth of our findings and offer richer insights into the use of social robots for depression screening in older adults.

Acknowledgments

The São Paulo Research Foundation (Fapesp), Brazil (2019/16352-0), the Center for Artificial Intelligence (C4AI-USP) with the support of the São Paulo Research Foundation (Fapesp) (2019/07665-4) and the IBM Corporation, and Human Robotics for granting us licenses to test the Robios robot avatar.

References

Abdi, J., Al-Hindawi, A., Ng, T., and Vizcaychipi, M. P. (2018). Scoping review on the use of socially assistive robot technology in elderly care. BMJ Open, 8(2):1–20. https://doi.org/10.1136/bmjopen-2017-018815

Alexopoulos, G. S., Katz, I. R., III, C. F. R., Carpenter, D., and Docherty, J. P. (2001). The expert consensus guide-line series. Pharmacotherapy of depressive disorders in older patients. Postgraduate Medicine, Spec No Phar-macotherapy:1–86.

Almeida, O. P. and Almeida, S. A. (1999). Reliability of the Brazilian version of the geriatric depression scale (GDS) short form. Arquivos de Neuro-psiquiatria, 57(2B):421–426. https://doi.org/10.1590/s0004-282x1999000300013

American Psychiatric Association (1994). Diagnostic and statistical manual of mental disorders (DSM-IV). APA, Washington, D.C., 4th edition.

Asprino, L., Gangemi, A., Nuzzolese, A. G., Presutti, V., Recupero, D. R., and Russo, A. (2019). Ontology-based knowledge management for comprehensive geriatric assessment and reminiscence therapy on social robots. In Consoli, S., Recupero, D. R., and Petkovic´, M., editors, Data Science for Healthcare: Methodologies and Applications, pages 173–193. Springer. https://doi.org/10.1007/978-3-030-05249-2_6

Bartneck, C., Kulic´, D., Croft, E., and Zoghbi, S. (2009). Measurement instruments for the anthropomorphism, ani-macy, likeability, perceived intelligence, and perceived safety of robots. International Journal of Social Ro-botics, 1:71–81. https://doi.org/10.1007/s12369-008-0001-3

Birrer, R. B. and Vemuri, S. P. (2004). Depression in later life: A diagnostic and therapeutic challenge. American Family Physician, 69(10):2375–2382.

Carson, N., Leach, L., and Murphy, K. J. (2018). A re-examination of Montreal cognitive assessment (MoCA) cutoff scores. International Journal of Geriatric Psychiatry, 33(2):379–388. https://doi.org/10.1002/gps.4756

Carvalho, H. W., Andreoli, S. B., Lara, D. R., Patrick, C. J., Quintana, M. I., Bressan, R. A., de Melo, M. F., de J. Mari, J., and Jorge, M. R. (2013). Structural validity and reliability of the positive and negative affect schedule – PANAS: Evidence from a large Brazilian community sample. Revista Brasileira de Psiquiatria, 35(2):169–172. https://doi.org/10.1590/1516-4446-2012-0957

Chang, Y.-L., Luo, D.-H., Huang, T.-R., Goh, J. O., Yeh, S.-L., and Fu, L.-C. (2022). Identifying mild cognitive impairment by using human-robot interactions. Journal of Alzheimer's Disease, 85(3):1129–1142. https://doi.org/10.3233/JAD-215015

Chen, Y.-C., Yeh, S.-L., Huang, T.-R., Chang, Y.-L., Goh, J. O. S., and Fu, L.-C. (2021). Social robots for evaluating attention state in older adults. Sensors, 21(21). https://doi.org/10.3390/s21217142

Colnar, S., Dimovski, V., Grah, B., Rogelj, V., and Bogataj, D. (2020). Gerontechnology: Literature review and research agenda. In 59th Annual Conference of the Society of Instrument and Control Engineers of Japan, pages 391–396, Piscataway. IEEE. https://doi.org/10.23919/SICE48898.2020.9240225

Cormons, L., Poulet, C., Pellier, D., Pesty, S., and Fiorino, H. (2020). Testing social robot acceptance: What if you could be assessed for dementia by a robot? A pilot study. In 6th International Conference on Mechatron-ics and Robotics Engineering, pages 92–98. https://doi.org/10.1109/ICMRE49073.2020.9065021

Cuijpers, P., de Graaf, R., and van Dorsselaer, S. (2004). Minor depression: Risk profiles, functional disability, health care use and risk of developing major depression. Journal of Affective Disorders, 79(1-3):71–79. https://doi.org/10.1016/S0165-0327(02)00348-8

de Araujo, B. S., Fantinato, M., Peres, S. M., de Melo, R. C., Batistoni, S. S. T., Cachioni, M., and Hung, P. C. K. (2022). Effects of social robots on depressive symptoms in older adults: A scoping review. Library Hi Tech, 40(5):1108–1126. https://doi.org/10.1108/LHT-09-2020-0244

Do, H. M., Sheng, W., Harrington, E. E., and Bishop, A. J. (2021). Clinical screening interview using a social robot for geriatric care. IEEE Transactions on Automation Science and Engineering, 18(3):1229–1242. https://doi.org/10.1109/TASE.2020.2999203

Eggleston, K. and Lee, Y. S. (2020). The impact of robots on staffing in nursing homes. https://www-csli.stanford.edu/events/hai-virtual-community-impact-robots-staffing-nursing-homes. HAI Vir-tual Community.

Felzmann, H. (2020). Can a robot provide the answer? Ethical considerations in using social robots for assessment and monitoring in healthcare. In: Charalambous, A. (ed.) Developing and Utilizing Digital Technology in Healthcare for Assessment and Monitoring. Springer, Cham. https://doi.org/10.1007/978-3-030-60697-8_3

Field, T. (2010). Touch for socioemotional and physical well-being: A review. Developmental Review, 30(4):367-383. https://doi.org/10.1016/j.dr.2011.01.001

Fiorini, L., D'onofrio, G., Rovini, E., Sorrentino, A., Coviello, L., Limosani, R., Sancarlo, D., and Cavallo, F. (2019). A robot-mediated assessment of Tinetti balance scale for sarcopenia evaluation in frail elderly. In 28th IEEE International Conference on Robot and Human Interactive Communication. https://doi.org/10.1109/RO-MAN46459.2019.8956439

García-Olaya, A., Fuentetaja, R., García-Polo, J., González, J. C., and Fernández, F. (2019). Challenges on the applica-tion of automated planning for comprehensive geriatric assessment using an autonomous social robot. Advances in Intelligent Systems and Computing, 855:179–194. https://doi.org/10.1007/978-3-319-99885-5_13

González-González, C. S., Verónica V.-H., and Gil-Iranzo, R. M. (2021). Social robots in hospitals: A systematic review. Applied Sciences, 11(13): 5976. https://doi.org/10.3390/app11135976

Vitale, J., Tonkin, M., Herse, S., et al. (2018). Be more transparent and users will like you: A robot privacy and user experience design experiment. In 2018 ACM/IEEE International Conference on Human-Robot Interaction, pages 379–387. https://doi.org/10.1145/3171221.3171269

Kobayashi, T., Sameshima, N., Imai, T., Arai, K., Watanabe, T., Miyazaki, T., and Tsujino, A. (2019). Dementia screening system based on SNS agency robot. In IEEE International Conference on Consumer Electronics, pages 306–309. https://doi.org/10.1109/ICCE-Berlin47944.2019.8966145

Lewis, L. (2014). Avatars and robots as social companions in healthcare: Requirements, engineering, adoption and ethics. International Journal of Enterprise Information Systems, 10(2):21–39. https://doi.org/10.4018/ijeis.2014040102

Linden, D. J. (2016). Touch: The Science of Hand, Heart, and Mind. Penguin Books.

Minallah, A., Azam, N., and Merani, I. (2019). Frequency of depression with associated risk factors among elder-ly in two tertiary care hospitals in Rawalpindi. Pakistan Armed Forces Medical Journal, 69(2):S317–S321. https://doi.org/10.4103/jfmpc.jfmpc_947_22

Montorio, I. and Izal, M. (1996). The geriatric depression scale: A review of its development and utility. Interna-tional Psychogeriatrics, 8(1):103–112. https://doi.org/10.1017/S1041610296002505

Mucchiani, C., Cacchione, P., Johnson, M., Mead, R., and Yim, M. (2021). Deployment of a socially assistive robot for assessment of Covid-19 symptoms and exposure at an elder care setting. In 30th IEEE International Conference on Robot and Human Interactive Communication, pages 1189–1195. https://doi.org/10.1109/RO-MAN50785.2021.9515551

Nandanwar, A., and Dutt, V. (2023). Assessing stress, anxiety, and depression with social robots via conversation-al AI. 16th International Conference on Pervasive Technologies Related to Assistive Environments, pages 732–738. https://doi.org/10.1145/3594806.3596589

Naranjo-Saucedo, A. B., Duen˜as-Ruiz, A., Suarez-Mejias, C., Iglesias, A., Marfil, R., and Bandera, A. (2019). CLARC: Robotic platform to assist the CGA evaluation in elderly patients. In Habib, M. K., editor, Handbook of Research on Advanced Mechatronic Systems and Intelligent Robotics, pages 19–41. IGI Global. https://doi.org/10.4018/978-1-7998-0137-5.ch002

Nasreddine, Z. S., Phillips, N. A., Bédirian, V., Charbonneau, S., Whitehead, V., Collin, I., Cummings, J. L., and Chertkow, H. (2005). The Montreal cognitive assessment, MoCA: A brief screening tool for mild cognitive impairment. Journal of the American Geriatrics Society, 53(4):695–699. https://doi.org/10.1111/j.1532-5415.2005.53221.x

Pearson, N., Naylor, P. J., Ashe, M.C. et al. (2020) Guidance for conducting feasibility and pilot studies for im-plementation trials. Pilot Feasibility Studies, 6(167). https://doi.org/10.1186/s40814-020-00634-w

Pu, L., Moyle, W., Jones, C., and Todorovic, M. (2019). The effectiveness of social robots for older adults: A systematic review and meta-analysis of randomized controlled studies. Gerontologist, 59(1):E37–E51. https://doi.org/10.1093/geront/gny046

Rossi, S., Santangelo, G., Staffa, M., Varrasi, S., Conti, D., and Nuovo, A. D. (2018). Psychometric evaluation supported by a social robot: Personality factors and technology acceptance. In 27th IEEE International Sym-posium on Robot and Human Interactive Communication, pages 802–807. https://doi.org/10.1109/ROMAN.2018.8525838

Sharkey, A., Sharkey, N. (2012). Granny and the robots: Ethical issues in robot care for the elderly. Ethics and Information Technology, 14:27–40. https://doi.org/10.1007/s10676-010-9234-6

Sheikh, J. I. and Yesavage, J. A. (1986). Geriatric depression scale (GDS): Recent evidence and development of a shorter version. Clinical Gerontologist: The Journal of Aging and Mental Health, 5(1-2):165–173. https://doi.org/10.1300/J018v05n01_09

Stiles, P. G. and McGarrahan, J. F. (1998). The geriatric depression scale: A comprehensive review. Journal of Clinical Geropsychology, 4(2):89–110. https://doi.org/10.1037/t00930-000

Thabane, L., Ma, J., Chu, R. et al. (2010). A tutorial on pilot studies: The what, why and how. BMC Medical Re-search Methodology, 10(1). https://doi.org/10.1186/1471-2288-10-1

The Institute for Health Metrics and Evaluation (IHME) (2019). Global Burden of Disease (GBD). University of Washington, Seattle. https://www.healthdata.org/research-analysis/gbd.

Ting, K. L. H., Voilmy, D., Roll, Q. D., Iglesias, A., and Marfil, R. (2021). Fieldwork and field trials in hospitals: Co-designing a robotic solution to support data collection in geriatric assessment. Applied Sciences, 11(7). https://doi.org/10.3390/app11073046

Varrasi, S., Nuovo, S. D., Conti, D., and Nuovo, A. D. (2019). Social robots as psychometric tools for cognitive assessment: A pilot test. In 10th International Workshop on Human Friendly

Robotics, volume 7, pages 99–112. https://doi.org/10.1007/978-3-319-89327-3_8

Watson, D., Clark, A. L., and Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS scales. Journal of Personality and Social Psychology, 54(6):1063–1070. https://doi.org/10.1037/0022-3514.54.6.1063

Wells, K. B., Burnam, M. A., Rogers, W., Hays, R., and Camp, P. (1992). The course of depression in adult outpatients: Results from the medical outcomes study. Archives of General Psychiatry, 49(10):788–794. https://doi.org/10.1001/archpsyc.1992.01820100032007

World Health Organization (WHO) (2015). World report on ageing and health. World Health Organization, Gene-va. https://apps.who.int/iris/handle/10665/186463.

Yesavage, J. A. (1988). Geriatric depression scale. Psychopharmacology Bulletin, 24(4):709–711.