

# **Equivalence of chatbot and paper-and-pencil versions of the De Jong Gierveld Loneliness Scale**

## **Abstract**

Technological progress provides health professionals with an excellent opportunity to take advantage of these developments and contribute to the development of efficient ways of diagnosing, monitoring, treating and assisting users. The purpose of this work is to present the results of a study conducted to examine the quantitative equivalence of paper-and-pencil and a voice-based conversational assistant, popularly known as a “chatbot”, as means to administer tests. One hundred and eight undergraduate university students completed both versions of the De Jong Gierveld Loneliness Scale. The interval between the first and second administration was set at four days. Validity, internal structure, internal consistency and equivalence of chatbot administration mode were assessed. A confirmatory factor analysis was used to verify the factor structure and provided a two-factor structure. Validity and internal consistency are adequate. These results support the feasibility of using chatbots for loneliness assessment in a sample of undergraduate university students and other populations in future.

**Keywords:** Chabot, loneliness scale, quantitative equivalence, dimensionality, undergraduate university students.

## **Introduction**

The technological progress occurring nowadays provides health professionals with an excellent opportunity to take advantage of these developments and use them to cover certain needs they have with regard to the services they offer (Rabbitt et al. 2015). Technology can contribute to the development of efficient ways of diagnosing, monitoring, treating and assisting users. In the case of conversational agents, multiple

studies have reported the scope of use in health care (Abd-Alrazaq et al. 2020; de Cock et al. 2020).

Traditionally, the evaluation of behavioural and psychological traits, such as loneliness, personality or cognitive functions, is conducted using conventional techniques such as paper-and-pencil, phone and email surveys (Miller 2012). Several studies have explored the equivalence, viability and interchangeability between paper-and-pencil and electronic versions of Health Surveys (White et al. 2018).

Evaluating fluctuating and subjective constructs by using a paper-and-pencil survey has several limitations for data gathering. First of all, it reduces answers to an exact period of time, making it difficult to see evolution over time. Even when scales are administered several times to conduct a follow-up study, time in between these two measurements is indeterminate. Second, it is difficult and expensive to reach some populations, such as people living in isolation or in geographically remote places. Third, it is time consuming to enter all responses in the computer analysis system before processing them. Lastly, paper-and-pencil questionnaires are commonly answered with the help of data collectors or other professionals, which can affect individual responses (Alberdi et al. 2018).

Several researchers contemplate the possibility of smart mobile devices replacing paper-and-pencil, telephone and email surveys, laboratory studies and field studies (Harari et al. 2017; Kim et al. 2019; Okeke et al. 2018). The digitization of scales for measuring psychological constructs would bring several advantages for health professionals and researchers (Salas et al 2018). Some of the limitations mentioned above can be overcome by using and developing technological innovations. Psychological research studies have provided valid reliable data when obtained by using the Internet (Hewson 2014). Ecological Momentary Assessment (EMA) have been used as a method

of data collection and it typically uses prompts administered through a personal electronic device, such as tablet or a smartphone (McDevitt-Murphy et al. 2018).

The digitization of scales such as the DJGLS would encourage their use, and consequently also improve response-rates, as well as contributing to a more automated and efficient collecting method (time, resources, etc.). These advantages are noticed when the evaluation instruments are specifically digitized using mobile applications (Salas et al. 2018) or other electronic devices (White et al. 2018). Several meta-analyses have showed equivalence between mean scores for self-report survey responses gathered using paper-and-pencil and computer data collection methodologies (Weigold et al. 2018).

One way of digitizing surveys, collecting health data and giving health support to patients is by using conversational agents (Kim et al. 2019; Laranjo et al. 2018). “There can be two such types of chatbots, a text-based chatbot is the one that interacts and communicates through text or messaging. Voice-activated chatbots are the one that interacts and communicates through voice. They accept the command in an oral or written form and reply through voice”. These are artificial intelligence (AI) software applications capable of maintaining a conversation with humans based on Natural Language Processing and Artificial Intelligence, Machine Learning and Deep Learning algorithms. This enables them to transcribe verbal language to the written form and to further identify terms and language structure (Hinton et al. 2012). Chatbot’s voice interaction emulates the most natural way humans interact with each other: verbal language communication. This intuitive way of communicating simplifies robot-human interaction. It can turn completing a survey into a social interaction, therefore encouraging the user’s commitment and leading to high-quality data (Kim et al. 2019).

The evolution of technology towards the creation of chatbots or smart devices is a good chance to improve data collection for health and research purposes (Bellegente et al.

2018). Several researchers have observed that when using a conversational agent to talk about health related issues and to complete anonymous surveys people tend to be more truthful and open when sharing private information (Lucas et al. 2017; Pickard et al. 2016). When compared to traditional face-to-face assessment, social robots result in more objective and replicable assessment (Desideri et al. 2019). These bots can serve as tools for health professionals by complementing their work in clinical assistance and having an impact on mental health outcomes (Rabbitt et al. 2015). In a recent review about conversational agents in healthcare, Laranjo et al. (2018) concluded that the use of conversational agents in healthcare is an emerging field of research that may have the potential to benefit health across a broad range of application domains.

Loneliness is a subjective and complex construct. There is no agreed definition, though authors repeat and agree on some common key concepts. It is a multidimensional construct that appears as a result of the negative auto-evaluation of the actual number and quality of social relationships (Fokkema et al. 2012). Loneliness appears when there is a discrepancy between the number and quality of relationships we have and the ones we want (Perlman and Peplau 1982). It is an individual unpleasant feeling that emerges from a subjective judgement of the degree of satisfaction with the actual social relationships. Individual and societal factors influence this process of self-evaluation, which is why there are cultural differences in loneliness (De Jong Gierveld and Tesch-Romer 2012). Everyone, from time to time, feels lonely, and this is part of being human (Cacioppo et al. 2014), but when people feel lonely most or all of the time, it can become a public health problem.

There is growing evidence that unwanted loneliness and social isolation is associated with both the physical and the psychological health of elderly people, related to aspects such as cardiovascular disease, obesity, reduced physical activity and functional capacity,

stress, depression, anxiety, sleep disturbances, cognitive functioning, mortality, mild cognitive decline, coronary heart disease and stroke, risk of various dementias and even Alzheimer's disease (Courtin and Knapp 2017; Khosravi et al. 2016; Lara et al. 2019). Frequently feeling lonely is linked to being readmitted to a hospital or having a longer stay (Valtorta et al. 2018). Due to its magnitude, it is necessary to evaluate loneliness and identify those at risk of suffering from it. An effective evaluation of loneliness would also reduce health care costs and medical services overload (Cacioppo and Cacioppo 2018; Ercole and Parr 2020).

Several studies show that loneliness is highest in adolescence and in those aged 80 or older (Victor and Yang 2012). It is particularly intense in adolescence because this is a period of life when peer relations are very important. Furthermore, in this stage in life loneliness is experienced more strongly during its earlier phases than later on (Ercole and Parr 2020).

Four main scales are used around the world for loneliness evaluation: University of California Los Angeles Scale (UCLA), De Jong Gierveld Loneliness Scale (DJGLS), Social Emotional Loneliness Scale for Adults (SELSA) and Emotional and Social Loneliness Inventory (ESLI). The DJGLS, which is mainly used in Europe is a valid and reliable scale tested in several cultures (De Jong Gierveld and Van Tilburg 2010). It was elaborated in accordance with the distinction between the two subtypes of social and emotional loneliness (Weiss 1973). However, due to its psychometric properties it can also be used as a unidimensional measure for loneliness. Initially it had 11 items, but with the aim of adapting it for use in larger surveys the same authors constructed a short six-item version (De Jong Gierveld and Van Tilburg 2006).

Our objective was to examine the quantitative equivalence of the use of paper-and-pencil and chatbot versions in the administration of the DJGLS in a sample of

undergraduate university students. With this purpose we analysed the psychometric properties of this scale using a chatbot version and a paper-and-pencil version. Our first hypothesis was that the DJGLS in chatbot version is a valid reliable method for evaluating loneliness. Our second hypothesis was that both versions of the scale, paper-and-pencil and chatbot, will be equivalents.

## **Method**

### **Participants and procedures**

We recruited 154 undergraduate university students at the Universitat Jaume I (Spain). All of them had to complete the De Jong Gierveld Loneliness Scale (DJGLS) and the UCLA Loneliness Scale in both paper-and-pencil and chatbot versions. Only 108 of them managed to complete both parts. Therefore, the sample was composed of these 108 participants with ages ranging from 17 to 54 years ( $M = 19.50$ ,  $SD = 4.19$ ) and there were more women ( $n = 73\%$ ) than men ( $n = 24\%$ ), and 3% refused to answer. Most of the students were living with someone ( $n = 94.4\%$ ) and only a few were living alone ( $n = 5.6\%$ ).

### **Measures**

**The De Jong Gierveld Loneliness Scale.** Loneliness was assessed with De Jong Gierveld Loneliness Scale (DJGLS). It encompasses three negatively formulated items (“I miss having people around me”, “I experience a general sense of emptiness”, and “I often feel rejected”) and three positively formulated items (“There are many people I can trust completely”, “There are plenty of people I can rely on when I have problems” and “There are enough people I feel close to”). The items had three response categories: (1 = no), (2 = more or less) and (3 = yes). It is a reliable valid instrument to assess overall loneliness in adults of all ages (De Jong Gierveld and Van Tilburg 2010; Hajek and König 2017). The scale can be used either as a unidimensional measure of loneliness or as two

dimensions, emotional and social loneliness (De Jong Gierveld and Van Tilburg 2010). In the present sample, Cronbach alphas for the subscale scores were 0.50 and 0.63, respectively.

**UCLA Loneliness Scale.** Loneliness was also measured with a three-item version of the UCLA Loneliness Scale (Hughes et al. 2004). It encompasses (“How often do you feel that you lack companionship”, “How often do you feel left out” and “How often do you feel isolated from others”). Responses were measured on a 3-point scale: (1 = hardly ever), (2 = some of the time), and (3 = often). Scores on the individual items were added up to produce the scale. Cronbach’s alpha in the present sample was 0.80.

**Chatbot.** The Information and Communication Technologies solution to collect the data from the questionnaires consists of two parts: i) a client application programmed on chatbot technology, which serves as a voice-based user interface; and ii) a server application which controls the users’ input and collects the data they provide.

The data recollection platform consists of two parts: i) the client part, which runs on a smart mobile phone, and ii) the server side, which in turn consists of two main components: a) the Natural Language Processing engine based on Artificial Intelligence algorithms, which runs on the cloud, and b) the validation and storage engine, which runs on UJI premises. The client part runs on top of Google Assistant and was developed using Dialog Flow, which makes it possible to create the structure of a dialogue between a human being and a machine, commonly known as a chatbot; we called our chatbot “Serena”. In our case, the defined dialogue has no bifurcation depending on the user’s answer to a question, the dialogue between the human being and the chatbot is linear, that is to say, the user is always asked the questions in the same time sequence. After invoking the “Serena” chatbot by saying “Talk to my attendant Serena”, the set of questions a user is asked are: i) Sex/Gender, ii) Age, iii) Do you live alone? iv) Questions belonging to the

de Jong questionnaire, and v) Questions belonging to the UCLA questionnaire. The locution given by the user in spoken language is converted to text using powerful AI speech-recognition algorithms. The audio containing the user's speech is sent to Google premises using the Secure Socket Layer (SSL) protocol on top of the HTTP protocol. The recognized text is presented to the users in the Google Assistant user interface. Afterwards, Natural Language Processing (NLP) algorithms are applied to the text to find out data which represent the information of interest, called entities, i.e. in the text "I am sixty-six years old" the entity is the user's age "sixty-six", but the user might have said "I'm sixty-six", in both cases the entity of interest is the same, but the text in which the entity is embedded could be very different. AI algorithms were trained with an exhaustive number of cases in order to improve the accuracy of recognizing entities in spoken language. Once the user's locution has been converted to text and the entities have been recognized, all this information is sent to a server side application running on UJI premises using the SSL protocol. This server side application is in charge of, first, checking whether the answer given by the user is one of the valid answers to a question, this application was developed as a RESTful application using the Java programming language and Spring Boot technologies. Second, it is also responsible for storing all anonymous information in a secure private database. Elasticsearch was chosen as a No-SQL database due to its high searching performance if the data stored on it is text. The RESTful application and the No-SQL database run as Docker containers, which allows the performance of the server side application to be easily scaled up or down according to the number of users' connections, and to be fault tolerant by replicating the server application in various containers.

## **Procedure**



Participants were informed about the objectives of the study and confidentiality of information. Afterward, those who consented were given the study measures to complete. We provided a code to include it in the paper and chatbot questionnaires to analyse participant data. First, participants completed the paper-and-pencil version of DJGLS and UCLA in the university classroom. Three members of the research team were involved in the process. Second, they were given instructions on how to access the DJGLS and UCLA in chatbot version to complete it in their own home during the next four days and using their mobile device. The research was carried out in accordance with the Declaration of Helsinki ethical principles and approved by the University's Ethics Committee.

### **Statistical Analysis**

After the data collection phase was finished, all data gathered by the system was processed to compare the results provided using chatbot technology with the results provided by the paper-and-pencil procedure. Data were analysed with the Statistical Package for the Social Sciences (SPSS 25) (IBM SPSS® Statistics) and EQS 6.3 (Bentler 2006). The paper-and-pencil and the chatbot versions of the DJGLS were compared with Wilcoxon's signed-rank test. A confidence interval procedure for assessing mean equivalence was calculated, as it has been recommended over other equivalence testing procedures (Weigold et al., 2016; 2018). The existence of two dimensions was examined by means of confirmatory factor analysis. Because the scores on the items were dichotomous, tetrachoric correlations were computed and arbitrary generalized least squares (AGLS) estimation was applied. The main advantage of the AGLS estimator is that it does not require multivariate normality (Browne 1984). Three models were tested: (1) a single-factor model; (2) two-factor uncorrelated model; and (3) two-factor correlated model. The model fit was evaluated by considering the Chi-square significance value –

values greater than .01 indicate a good fit; Comparative Fit Index (CFI); Non Normed Fit Index (NNFI) – values equal to or greater than .95 indicate a good fit; Standardized Root Mean Square Residual (SRMR) and Root Mean Square Error of Approximation (RMSEA) – values below .08 indicate a good fit (Hu and Bentler 1999). Reliability was calculated using internal consistency with Cronbach's alpha coefficient. To estimate test-retest reliability between two versions the Pearson coefficient was used. As a second measure of test-retest reliability we used the coefficient of Weighted Kappa, using Landis and Koch's (1977) standards for its interpretation. Convergent validity was calculated with the correlation with the three-item UCLA in the paper-and-pencil and chatbot versions.

## **Results**

Mean and standard deviation was ( $M = 2.97$ ,  $SD = 1.49$ ) in paper-and-pencil condition and ( $M = 3.05$ ,  $SD = 1.64$ ) in chatbot condition. Equivalence interval at  $\pm 20\%$  was  $\pm 0.12$  and the lower and upper CI as (- 0.305, 0.143). The results of the present study supported quantitative equivalence between chatbot and paper-and-pencil conditions. This implies that self-report survey-based measures can generally be administered through the chatbot with good (i.e., equivalent to paper-and-pencil) results.

## **Reliability**

Means, standard deviations, skewness, kurtosis, corrected item-total correlation and Cronbach's alpha coefficient for the emotional and social dimensions are shown in Table 1. Correlation between the paper-and-pencil and chatbot versions of the DJGLS was positive, large and statistically significant ( $r = 0.76$ ,  $p < .01$ ). The strength of agreement between the 6 items in the chatbot and the paper-and-pencil versions of the DJGLS is at least moderate. For item 3 and item 1 the strength of agreement is substantial (Table 2).

**Table 1**

Means (M), standard deviations (SD), skewness, kurtosis, corrected item-total correlation and alpha in DJGLS using a chatbot

Item	Mean	SD	Skewness	Kurtosis	Corrected item-total correlation	Alpha
G1	0.38	0.48	0.50	-1.78	0.40	
G2	0.64	0.36	-1.90	1.67	0.22	0.50
G3	0.36	0.48	0.58	-1.68	0.30	
G4	0.56	0.49	-0.22	-1.98	0.45	
G5	0.63	0.48	-0.54	-1.73	0.39	0.63
G6	0.26	0.44	1.11	-0.77	0.38	

**Table 2**

Agreement between items of DJGLS using paper-and-pencil and chatbot

Item	Weighted Kappa	<i>p</i>
G1	0.72 <sup>b</sup>	0.00
G2	0.45 <sup>a</sup>	0.00
G3	0.64 <sup>b</sup>	0.00
G4	0.53 <sup>a</sup>	0.00
G5	0.57 <sup>a</sup>	0.00
G6	0.43 <sup>a</sup>	0.00

<sup>a</sup> Moderate strength of agreement; <sup>b</sup> Substantial strength of agreement

### Convergent validity

As an external criterion of convergent validity a different well-established measure of loneliness was used, namely, the three-item UCLA loneliness scale in the paper-and-pencil version and the chatbot version (Table 3). The correlation between the DJGLS chatbot version and the UCLA paper-and-pencil version was positive, large and statistically significant ( $r = 0.69, p < .01$ ). The correlation between the DJGLS chatbot version and the UCLA chatbot version was, as expected, also positive, large and statistically significant ( $r = 0.72, p < .01$ ).

**Table 3**

Correlations between the different scales and different versions of loneliness measurement

	DJGLS Chatbot	DJGLS Paper-and-pencil	UCLA Paper-and-pencil
DJGLS Paper-and-pencil	0.76**		
UCLA Paper-and-pencil	0.69**	0.66**	
UCLA Chatbot	0.72**	0.61**	0.82**

\*\*  $p < .01$

### Confirmatory factorial analysis

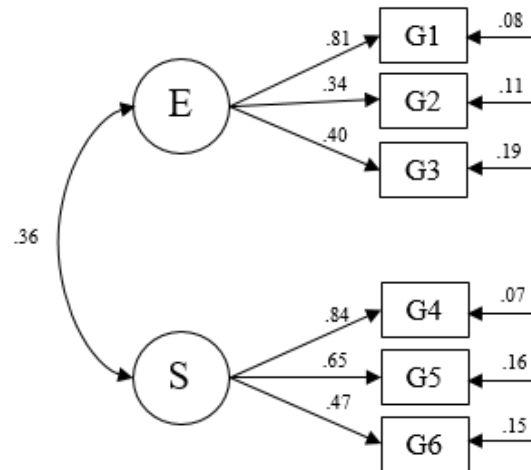
The fit indices for the three models examined are presented in Table 4. As has been mentioned in the data analysis section, a model was considered acceptable if CFI was greater than 0.95, NNFI was greater than 0.95, and SRMR and RMSEA were less than 0.08. We checked the differences between models and found the best fitting model was clearly the two-factor correlated one: an emotional and a social loneliness dimension. Standardized regression weights and error variances can be seen in Figure 1. All the standardized regression weights are significant. All errors were significant, with the exception of the errors in G1 and G4, which suggests that no errors were made in those items.

**Table 4**

Fit indices for the different models

Models	$\chi^2$	<i>df</i>	<i>p</i>	<i>CFI</i>	<i>NNFI</i>	<i>SRMR</i>	<i>RMSEA</i>	<i>RMSEA 90% CI</i>
Single-factor	22.60	9	0.01	0.78	0.59	0.08	0.13	[0.06, 0.19]
Two-factor uncorrelated	15.59	9	0.07	0.87	0.72	0.02	0.10	[0.03, 0.17]
Two-factor correlated	9.83	8	0.28	0.97	0.95	0.05	0.04	[0.00, 0.12]

$\chi^2$  = chi-square goodness-of-fit statistic; *df* = degrees of freedom; *CFI* = comparative fit index; *NNFI* = non normed fit index; *SRMR* = standardized root mean-square residual; *RMSEA* = root mean square error of approximation; *CI* = confidence interval



**Figure 1**

Standardized solution for the two factor-correlated model

## Discussion

Our first hypothesis was that the DJGLS in chatbot version was a valid and reliable method for measuring loneliness and this was confirmed by the internal consistency, test-retest reliability and confirmatory factor analysis. The results showed the existence of an emotional and a social loneliness scale according to the distinction established by Weiss (Weiss 1973), and also in line with results obtained in other studies by De Jong Gierveld and Van Tilburg (2010). Cronbach's alpha coefficient for the emotional dimension was 0.50 and for the social dimension was 0.63. Our second hypothesis was that both versions of the scale (paper-and-pencil and chatbot) were equivalents, and we found that differences in scores in loneliness between the two versions were minor. Furthermore, the analysis agreement between the six items of the two methods was adequate; these results coincide with those observed by Barrigón et al. (2017).

## Conclusions and future directions

Mobiles are affordable and ubiquitous devices. Developing an open access e-health chatbot and making it available for anyone with Internet access is an opportunity to

improve the quality of health care, expand coverage and reduce health costs. It is a possible way to reach sub-populations that are often underserved or even those living in developing countries where sometimes alternative healthcare devices are unattainable. Regarding research practice, using a chatbot to complete scales provides the opportunity to gather data in real-life situations while reducing physical and temporal barriers. This type of technology is an opportunity to reduce research costs and increase the users' participation. The automation of the process contributes to a more efficient way of managing data (collecting, processing, saving or sharing among health professionals). Vaidyam et al. (2019) explored the current evidence for conversational agents in the field of mental health and their role in the screening, diagnosis and treatment of mental illnesses. Early evidence showed that the mental health field could use conversational agents not only in diagnosis but also in psychiatric and psychological treatment.

In an increasingly technological society where many people have mobile devices with Internet access, evaluating loneliness by using personal mobile phones seems an appropriate possibility. Some of the inconveniences of using chatbots are that unintended human-like biased algorithms can lead to prejudiced outcomes (Obermeyer et al. 2019). Machines endowed with artificial intelligence can produce the same biases as those present in the data they are trained with, sometimes as a consequence of stereotypes and inequalities that are widespread in society. Human-like semantic biases, such as cultural or gender stereotypes, may be replicated and appear when applying machine learning to human language data (Caliskan et al. 2017; Lee et al. 2019).

This study is part of a larger project which aims to evaluate loneliness in the elderly population. The ageing of the world's population and the increasing risk of loneliness over the age of 65 are two key factors for directing this chatbot to them in future phases of the project. In the development of the next phase our aim is to research the equivalence

of paper-and-pencil and chatbot administered scales in samples of older adults. Co-design will be used, which implies that older people and caregivers will be involved in improving and implementing the chatbot design through discussion groups (Bazzano et al. 2017). This perspective is in line with that set out by the World Health Organization (WHO, 2016), as it will empower people to play an active role in the development of the device and to participate in the improvement of their own health. We aim to build co-creation groups formed by people older than 55 years and to encourage their active participation in this citizen science project. This will contribute to the creation of an age-friendly chatbot with high ecological validity. By adapting this tool to the elder and future elder generations we seek to reduce the technology gap and introduce an innovative and intuitive tool for loneliness evaluation at all ages. In the next phase of the project, i.e. the training of the chatbot with natural language as input data, we will pay special attention to this in order to prevent unintended human-like biases. One way in which we will do this is by making the chatbot known among older people with different socio-demographic aspects like age, gender, cultural and acquisitive level, religion, etc. so that the data feeding our chatbot is diverse and representative of the whole population.

This study has two important limitations. The first one is that the research was conducted among undergraduate university students. The second one is that participants always completed the paper-and-pencil format first, so this is a possible confound in the results of the research. However, the objective was not to generalize the results to a university population, but to investigate the functioning of the chatbot. The results of this research will allow us to adapt the chatbot to the elderly population, through their collaboration.

Most of the literature regarding health-related chatbots comes from outside the health field (e.g. engineering and information systems). There is a need to create synergies

between technology and health professionals in order to merge their knowledge and apply it to improved technologies (Vaidyam et al. 2019). For example, psychologists can contribute with their technical knowledge to accomplish more valid and reliable technologies (Chamorro-Premuzic et al. 2016). We believe these synergies are essential in strengthening digital health research. Due to the small amount of literature examining paper-and-pencil and chatbot equivalence, this is an important area for more specific future research on equivalence testing methodology.

### **Conflict of interest**

On behalf of all authors, the corresponding author states that there is no conflict of interest.

### **Ethical responsibilities**

This project was approved by the Ethical Commission at Universitat Jaume I with reference number: CD/58/2019.

### **Funding**

This research was supported by the Ministry of Science and Innovation and the Spanish Foundation for Science and Technology (FECYT) with reference number: FCT-18-13677.

### **Acknowledgements**

We are grateful to the undergraduate university students from Universitat Jaume I who voluntarily participated in the study.

### **References**

Abd-Alrazaq, A. A., Rababeh, A., Alajlani, M., Bewick, B. M., & Househ, M. (2020).  
Effectiveness and Safety of Using Chatbots to Improve Mental Health:



Systematic Review and Meta-Analysis. *Journal of Medical Internet Research*, 22(7), e16021. <https://doi.org/10.2196/16021>.

Alberdi, A., Weakley, A., Aztiria, A., Schmitter-Edgecombe, M., & Cook, D. J. (2018). Automatic assessment of functional health decline in older adults based on smart home data. *Journal of Biomedical Informatics*, 81, 119-130. <https://doi.org/10.1016/j.jbi.2018.03.009>.

Bazzano, A. N., Martin, J., Hicks, E., Faughnan, M., & Murphy, L. (2017). Human-centred design in global health: A scoping review of applications and contexts. *PloS one*, 12(11), e0186744. <https://doi.org/10.1371/journal.pone.0186744>.

Barrigón, M. L., Rico-Romano, A. M., Ruiz-Gómez, M., Delgado-Gómez, D., Barahona, I., Aroca, F., & Baca-García, E. (2017). Comparative study of pencil-and-paper and electronic formats of GHQ-12, WHO-5 and PHQ-9 questionnaires. *Revista de Psiquiatría y Salud Mental*, 10, 160-167. <https://doi.org/10.1016/j.rpsm.2016.12.002>.

Bellagente, P., Crema, C., Depari, A., Ferrari, P., Flammini, A., Lanfranchi, G., Lenzi, G., Maddiona, M., Rinaldi, S., Sisinni, E., & Ziliani, G. (2018, March). Remote and non-invasive monitoring of elderly in a smart city context. In 2018 IEEE Sensors Applications Symposium (SAS) (pp. 1-6). IEEE. <https://doi.org/10.1109/SAS.2018.8336732>.

Bentler, P. M. (2006). *EQS 6 Structural Equations Program Manual* (Vol. 6). Encino, CA: Multivariate software.

Bonsaksen, T., Opseth, T. M., Misund, A. R., Geirdal, A. Ø., Fekete, O. R., & Nordli, H. (2019). The de Jong Gierveld Loneliness Scale used with Norwegian clubhouse

- members: Psychometric properties and associated factors. *International Journal of Psychosocial Rehabilitation*, 22(2), 88-100.
- Browne, M. W. (1984). Asymptotically distribution-free methods for the analysis of covariance structures. *British Journal of Mathematical and Statistical Psychology*, 37(1), 62-83. <https://doi.org/10.1111/j.2044-8317.1984.tb00789.x>.
- Cacioppo, J. T., & Cacioppo, S. (2018). The growing problem of loneliness. *The Lancet*, 391, 426. [https://doi.org/10.1016/S0140-6736\(18\)30142-9](https://doi.org/10.1016/S0140-6736(18)30142-9).
- Cacioppo, J. T., Cacioppo, S., & Boomsma, D. I. (2014). Evolutionary mechanisms for loneliness. *Cognition & Emotion*, 28(1), 3-21. <https://doi.org/10.1080/02699931.2013.837379>.
- Caliskan, A., Bryson, J. J., & Narayanan, A. (2017). Semantics derived automatically from language corpora contain human-like biases. *Science*, 356(6334), 183-186. <https://doi.org/10.1126/science.aal4230>.
- Chamorro-Premuzic, T., Winsborough, D., Sherman, R. A., & Hogan, R. (2016). New talent signals: Shiny new objects or a brave new world? *Industrial and Organizational Psychology*, 9(3), 621-640. <https://doi.org/10.1017/iop.2016.6>.
- Courtin, E., & Knapp, M. (2017). Social Isolation, loneliness and health in old age: a scoping review. *Health and social Care in the Community*, 25(3), 799-812. <https://doi.org/10.1111/hsc.12311>.
- de Cock C., Milne-Ives M., van Velthoven, M.H., Alturkistani, A., Lam, C. & Meinert, E. (2020). Effectiveness of Conversational Agents (Virtual Assistants) in Health Care: Protocol for a Systematic Review. *JMIR Research Protocols*, 9(3): e16934. <https://doi.org/10.2196/16934>

- De Jong Gierveld, J., & Tesch-Romer, C. (2012). Loneliness in old age in Eastern and Western European societies: theoretical perspectives. *European Journal of Aging*, 9(4), 285-295. <https://doi.org/10.1007/s10433-012-0248-2>.
- De Jong Gierveld, J., & Van Tilburg, T. (2006). A 6-item scale for overall, emotional, and social loneliness: Confirmatory tests on survey data. *Research on Aging*, 28, 582-598. <https://doi.org/10.1177/0164027506289723>.
- De Jong Gierveld, J., & Van Tilburg, T. (2010). The De Jong Gierveld short scales for emotional and social loneliness: tested on data from 7 countries in the UN generations and gender surveys. *European Journal of Ageing*, 7(2), 121–130. <https://doi.org/10.1007/s10433-010-0144-6>.
- De Jong Gierveld, J., Van Tilburg, T., & Dykstra, P. A. (2006). Loneliness and social isolation. In D. Perlman & A. Vangelisti (Eds.), *The Cambridge Handbook of Personal Relationships* (pp. 485-500). Cambridge, UK: Cambridge University Press.
- Desideri, L., Ottaviani, C., Malavasi, M., di Marzio, R., & Bonifacci, P. (2019). Emotional processes in human-robot interaction during brief cognitive testing. *Computers in Human Behavior*, 90, 331–342. <https://doi.org/10.1016/j.chb.2018.08.013>.
- Ercole, V., & Parr, J. (2020). Problems of Loneliness and Its Impact on Health and Well-Being. In: W. Leal, T. Wall, A. Azul, L. Brandli & P. Özuyar (Eds.), *Good Health and Well-Being. Encyclopedia of the UN Sustainable Development Goals*. Basel, Switzerland: Springer Nature Switzerland AG. [http://doi.org/10.1007/978-3-319-69627-0\\_127-1](http://doi.org/10.1007/978-3-319-69627-0_127-1).

- Fokkema, T., De Jong Gierveld, J., & Dykstra, P. A. (2012). Cross-national differences in older adult loneliness. *The Journal of Psychology, 146*, 201-228. <https://doi.org/10.1080/00223980.2011.631612>.
- Harari, G. M., Müller, S. R., Aung, M. S., & Rentfrow, P. J. (2017). Smartphone sensing methods for studying behavior in everyday life. *Current Opinion in Behavioral Sciences, 18*, 83-90. <https://doi.org/10.1016/j.cobeha.2017.07.018>.
- Hewson, C. (2014). Conducting research on the internet-a new era. *The Psychologist, 27*(12), 946-950.
- Hinton, G., Deng, L., Yu, D., Dahl, G. E., Mohamed, A. R., Jaitly, N., Senior, A., Vanhoucke, V., Nguyen, P., Sainath, T., & Kingsbury, B. (2012). Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups. *IEEE Signal Processing Magazine, 29*(6), 82-97.
- Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal, 6*(1), 1-55. <https://doi.org/10.1080/10705519909540118>.
- Hughes, M. E., Waite, L. J., Hawkey, L. C., & Cacioppo, J. T. (2004). A short scale for measuring loneliness in large surveys: Results from two population-based studies. *Research on Aging, 26*(6), 655-672. <https://doi.org/10.1177/0164027504268574>.
- Khosravi, P., Rezvani, A., & Wiewiora, A. (2016). The impact of technology on older adults' social isolation. *Computers in Human Behavior, 63*, 594-603. <https://doi.org/10.1016/j.chb.2016.05.092>.

- Kim, S., Lee, J., & Gweon, G. (2019, May). Comparing data from chatbot and web surveys: Effects of platform and conversational style on survey response quality. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (pp. 1-12).
- Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, *33*, 159.
- Lara, E., Martín-María, N., De la Torre-Luque, A., Koyanagi, A., Vancampfort, D., Izquierdo, A., & Miret, M. (2019). Does loneliness contribute to mild cognitive impairment and dementia? A systematic review and meta-analysis of longitudinal studies. *Ageing Research Reviews*, *52*, 7-16. <https://doi.org/10.1016/j.arr.2019.03.002>.
- Laranjo, L., Dunn, A. G., Tong, H. L., Kocaballi, A. B., Chen, J., Bashir, R., Surian, D., Gallego, B., Magrabi, F., Lau, A., & Coiera, E. (2018). Conversational agents in healthcare: A systematic review. *Journal of the American Medical Informatics Association*, *25*(9), 1248–1258. <https://doi.org/10.1093/jamia/ocy072>.
- Lee, N., Madotto, A., & Fung, P. (2019, August). Exploring Social Bias in Chatbots using Stereotype Knowledge. In *Proceedings of the 2019 Workshop on Widening NLP* (pp. 177-180).
- Lucas, G. M., Rizzo, A. S., Gratch, J., Scherer, S., Stratou, G., Boberg, J., & Morency, L. P. (2017). Reporting mental health symptoms: breaking down barriers to care with virtual human interviewers. *Frontiers in Robotics and AI*, *4*, 51. <https://doi.org/10.3389/frobt.2017.00051>.

- McDevitt-Murphy, M. E., Luciano, M. T., & Zakarian, R. J. (2018). Use of ecological momentary assessment and intervention in treatment with adults. *Focus, 16*(4), 370-375. <https://doi.org/10.1176/appi.focus.20180017>.
- Miller, G. (2012). The Smartphone Psychology Manifesto. *Perspectives on Psychology Science, 7*(3), 221-237.
- Obermeyer, Z., Powers, B., Vogeli, C., & Mullainathan, S. (2019). Dissecting racial bias in an algorithm used to manage the health of populations. *Science, 366*(6464), 447-453. <https://doi.org/10.1126/science.aax2342>.
- Okeke, F., Sobolev, M., & Estrin, D. (2018). Towards a Framework for Mobile Behavior Change Research. Proceedings of the Technology, Mind, and Society, Washington, DC, USA.
- Perlman, D., & Peplau, L. (1982). Theoretical approaches to loneliness. In: D. Perlman y L. Peplau (Eds.), *Loneliness: A Sourcebook of Current Theory, Research and Therapy* (pp 123-134). New York: Wiley.
- Pickard, M. D., Roster, C. A., & Chen, Y. (2016). Revealing sensitive information in personal interviews: is self-disclosure easier with humans or avatars and under what conditions? *Computers in Human Behavior, 65*, 23–30. <https://doi.org/10.1016/j.chb.2016.08.004>.
- Rabbitt, S. M., Kazdin, A. E., & Scassellati, B. (2015). Integrating socially assistive robotics into mental healthcare interventions: Applications and recommendations for expanded use. *Clinical Psychology Review, 35*, 35-46. <https://doi.org/10.1016/j.cpr.2014.07.001>.

- Salas, A. J., Reynolds, A., & Thomas, C. M. (2018). Technology Design for Frailty. In D. Chau & T. F. Osborne (Eds.), *Using Technology to Improve Care of Older Adults* (pp. 103-133). New York: Springer Publishing Company.
- Vaidyam, A. N., Wisniewski, H., Halamka, J. D., Kashavan, M. S., & Torous, J. B. (2019). Chatbots and conversational agents in mental health: a review of the psychiatric landscape. *The Canadian Journal of Psychiatry*, *64*(7), 456-464. <https://doi.org/10.1177/0706743719828977> .
- Valtorta, N. K., Moore, D. C., Barron, L., Stow, D., & Hanratty, B. (2018). Older adults' social relationships and health care utilization: A systematic review. *American Journal of Public Health*, *108*(4), e1-e10. <https://doi.org/10.2105/AJPH.2017.304256>.
- Victor, C. R., & Yang, K. (2012). The prevalence of loneliness among adults: a case study of the United Kingdom. *The Journal of Psychology*, *146*(1-2), 85-104. <https://doi.org/10.1080/00223980.2011.613875>.
- Weigold, A., Weigold, I. K., Drakeford, N. M. & Dykema, S. S. (2016). Equivalence of paper-and-pencil and computerized self-report surveys in older adults. *Computers in Human Behavior*, *54*, 407–413. <https://doi.org/10.1016/j.chb.2015.08.033>
- Weigold, A., Weigold, I. K., & Natera, S. N. (2018). Mean scores for self-report surveys completed using paper-and-pencil and computers: A meta-analytic test of equivalence. *Computers in Human Behavior*, *86*, 153–164. <https://doi.org/10.1016/j.chb.2018.04.038>.
- Weiss, R. S. (1973). *Loneliness: The experience of Emotional and Social Isolation*. Cambridge, MA: MIT Press.

White, M. K., Maher, S. M., Rizio, A. A., & Bjorner, J. B. (2018). A meta-analytic review of measurement equivalence study findings of the SF-36® and SF-12® Health Surveys across electronic modes compared to paper administration. *Quality of Life Research*, 27(7), 1757-1767. <https://doi.org/10.1007/s11136-018-1851-2>.

World Health Organization (WHO) (2016). What are integrated people-centred health services? Retrieved from <https://www.who.int/service-delivery-safety/areas/people-centred-care/en/>. Accessed 4 November 2019.