Instruments for Measuring Psychological Dimensions in Human-Robot Interaction: A Systematic Review of Psychometric Properties.

Roberto Vagnetti, Nicola Camp, Matthew Story, Khaoula Ait-Belaid, Suvo Mitra, Massimiliano Zecca, Alessandro Di Nuovo, Daniele Magistro

Submitted to: Journal of Medical Internet Research
on: December 18, 2023

Table of Contents

Instruments for Measuring Psychological Dimensions in Human-Robot Interaction: A Systematic Review of Psychometric Properties.

Roberto Vagnetti[1]; Nicola Camp[1]; Matthew Story[2]; Khaoula Ait-Belaid[3]; Suvo Mitra[4]; Massimiliano Zecca[3]; Alessandro Di Nuovo[2]; Daniele Magistro[1]

[1]Department of Sport Science School of Science and Technology Nottingham Trent University Nottingham GB
[2]Department of Computing & Advanced Wellbeing Research Centre Sheffield Hallam University Sheffield GB
[3]Wolfson School of Mechanical, Electrical and Manufacturing Engineering Loughborough University Loughborough GB
[4]Department of Psychology Nottingham Trent University Nottingham GB

Corresponding Author:
Daniele Magistro
Department of Sport Science
School of Science and Technology
Nottingham Trent University
College Drive
Clifton
Nottingham
GB

Abstract

Background: Numerous user-related psychological dimensions can significantly influence the dynamics between humans and robots. For developers and researchers, it is crucial to have a comprehensive understanding of the psychometric properties of the available instruments used to assess these dimensions, as they indicate the reliability and validity of the assessment.

Objective: This study aims to provide a review of the instruments available for assessing the psychological aspects between people and social and domestic robots, offering a summary of their psychometric properties and the quality of the evidence.

Methods: A systematic review was conducted following the PRISMA guidelines across different databases including Scopus, PubMed and IEEEXplore. The research strategy included studies that reported the development and psychometric validation of instruments designed to assess individuals' psychological dimensions related to robots. Studies concentrating on industrial robots, rescue robots, robotic arms, or those primarily concerned with technology validation or measuring anthropomorphism were excluded. Independent reviewers extracted instruments properties and methodological quality of their evidence following the COnsensus-based Standards for the selection of health Measurement INstruments (COSMIN) guidelines.

Results: From 3,828 identified records, the research strategy yielded 34 articles that validated and examined the psychometric properties of 27 instruments designed to assess individuals' psychological dimensions in relation to social and domestic robots. These instruments encompass a broad spectrum of psychological dimensions. While most studies predominantly focused on structural validity and internal consistency, consideration of other psychometric properties was frequently inconsistent or absent. Most of the instruments were targeted at both adults and older adults (18 years old and above). There is a limited number of instruments specifically designed for children, older adults, and healthcare contexts.

Conclusions: Given the strong interest in assessing psychological dimensions in the human-robot relationship, there is a need to develop new instruments using more rigorous methodologies and to consider a broader range of psychometric properties. This is essential to ensure the creation of reliable and valid measures for assessing people's psychological dimensions toward social and domestic robots. Among the limitations, the review included instruments applicable to both social robots and domestic robots, while excluding those for some specific types of robots (e.g., industrial robots). Clinical Trial: This work was supported by the EPSRC and NIHR (grant number EP/W031809/1, IMACTIVE). The review was not registered with any relevant database.

(JMIR Preprints 18/12/2023:55597)
DOI: https://doi.org/10.2196/preprints.55597

Preprint Settings

1) Would you like to publish your submitted manuscript as preprint?
Please make my preprint PDF available to anyone at any time (recommended).
Please make my preprint PDF available only to logged-in users; I understand that my title and abstract will remain visible to all users.
Only make the preprint title and abstract visible.
✓ **No, I do not wish to publish my submitted manuscript as a preprint.**
2) If accepted for publication in a JMIR journal, would you like the PDF to be visible to the public?
✓ **Yes, please make my accepted manuscript PDF available to anyone at any time (Recommended).**
Yes, but please make my accepted manuscript PDF available only to logged-in users; I understand that the title and abstract will remain v
Yes, but only make the title and abstract visible (see Important note, above). I understand that if I later pay to participate in <a href="http

Original Manuscript

Instruments for Measuring Psychological Dimensions in Human-Robot Interaction: A Systematic Review of Psychometric Properties.

Roberto Vagnetti[1], Nicola Camp[1], Matthew Story[2], Khaoula Ait-Belaid[3], Suvo Mitra[4,] Massimiliano Zecca[3], Alessandro Di Nuovo[2], and Daniele Magistro[1*]

[1] Department of Sport Science, School of Science and Technology, Nottingham Trent University
[2] Department of Computing & Advanced Wellbeing Research Centre, Sheffield Hallam University
[3] Wolfson School of Mechanical, Electrical and Manufacturing Engineering, Loughborough University
[4] Department of Psychology, Nottingham Trent University

*Corresponding author:
Daniele Magistro, Ph.D.
Department of Sport Science
School of Science and Technology
Nottingham Trent University
College Drive
Clifton
Nottingham
NG11 8NS
United Kingdom
Phone: +44 (0)115 848 3522
Email: daniele.magistro@ntu.ac.uk

Abstract

Background: Numerous user-related psychological dimensions can significantly influence the dynamics between humans and robots. For developers and researchers, it is crucial to have a comprehensive understanding of the psychometric properties of the available instruments used to assess these dimensions, as they indicate the reliability and validity of the assessment.

Objectives: This study aims to provide a systematic review of the instruments available for assessing the psychological aspects between people and social and domestic robots, offering a summary of their psychometric properties and the quality of the evidence.

Methods: A systematic review was conducted following the PRISMA guidelines across different databases including Scopus, PubMed and IEEEXplore. The search strategy encompassed studies meeting the following inclusion criteria: a) the instrument could assess psychological dimensions related to social and domestic robots, including attitudes, beliefs, opinions, feelings, and perceptions; b) the study focused on validating the instrument; c) it evaluated the psychometric properties of the instrument; d) it underwent peer review; e) it was in English. Studies focusing on industrial robots, rescue robots, robotic arms, or those primarily concerned with technology validation or measuring anthropomorphism were excluded. Independent reviewers extracted instruments properties and methodological quality of their evidence following the COnsensus-based Standards for the selection of health Measurement INstruments (COSMIN) guidelines.

Results: From 3,828 identified records, the research strategy yielded 34 articles that validated and examined the psychometric properties of 27 instruments designed to assess individuals' psychological dimensions in relation to social and domestic robots. These instruments encompass a broad spectrum of psychological dimensions. While most studies predominantly focused on structural validity and internal consistency, consideration of other psychometric properties was frequently inconsistent or absent. Despite their significance in the clinical context, no instrument has evaluated measurement error and responsiveness. Most of the instruments were targeted at both adults and older adults (18 years old and above). There is a limited number of instruments specifically designed for children, older adults, and healthcare contexts.

Conclusions: Given the strong interest in assessing psychological dimensions in the human-robot relationship, there is a need to develop new instruments using more rigorous methodologies and to consider a broader range of psychometric properties. This is essential to ensure the creation of reliable and valid measures for assessing people's psychological dimensions toward social and domestic robots. Among the limitations, the review included instruments applicable to both social robots and domestic robots, while excluding those for some specific types of robots (e.g., industrial robots).

The review was not registered with any relevant database.

Keywords: Psychometric; Human-Robot Interaction; Psychological Dimensions; Robot; Assessment; Systematic Review

INTRODUCTION

There is a growing interest in the field of social robotics when it comes to creating robots that can cater to people's needs. This is evidenced by the increasing number of publications covering various aspects of robotics [1]. This interest stems from the desire to develop robots that can engage in social interaction with humans, serving as collaborators, companions, tutors, and partners in various applications. Applications of social and domestic robots covers a widespread of areas, for instance, they have been proposed for educational purposes [2], for mental health and well-being [3], to support older adults in their homes [4–6] or to support different clinical populations such as autism spectrum disorder [7] or people with dementia [8].

While many studies have explored users' opinions and requirements to design and develop this technology in order to meet their needs in a participatory design framework (for instance, [9–13]), a major challenge for the success of social robots is the fact that their mere presence in everyday life does not automatically increase their chances of being accepted or users' willingness to interact with them [14]. Thus, understanding perspectives and preferences of people toward robots represents a crucial point for their development and acceptance [15–17]. How a robot is perceived plays a major role within the human-robot relationship [18]. Existing literature has identified several factors linked to individuals' predispositions towards robots and how robots are utilised [19]. Peoples' robot acceptance is influenced by attitudes and intentions to use robots [20]. According to the Unified Theory of Acceptance and Use of Technology [21], factors in the intention to use robots include attitude, perceived usefulness, perceived ease of use, enjoyment, trust and anxiety. However, many other psychological dimensions have been investigated within the human-robot relations, such as: beliefs [22], adaptability, control, companionship, sociability [19], attractiveness [23], social presence [24], intentionality [25], and expectations [26]. Thus, numerous user-related psychological dimensions can significantly influence the dynamics of the human-robot relationship. Systematic reviews focussing on different dimensions related to human-robot interactions with social robots reportedly indicate that most of the assessments are made by self-report measurements, raising concerns about their suitability [27–29].

For developers and researchers, it is crucial to have a comprehensive understanding of the psychometric properties of the available instruments. To make a reasoned decision regarding the utilisation of instruments in research, it is crucial to possess an understanding of instrument properties and to make comparisons between them [30]. Indeed, psychometric properties encompass attributes of an instrument that serve as indicators of its reliability and validity [31]. They help ascertain whether the measure accurately assesses what it's meant to and consistently gauges the intended dimension. In this context, systematic reviews of instrument psychometric properties can assist practitioners and researchers in choosing the most suitable measurement instrument tailored to their specific needs [32]. These reviews are valuable because they consider both the instrument psychometric properties and the methodological quality of the studies conducted to assess them [33]. This knowledge is essential for making informed decisions and effectively evaluating the performance and impact of robots in various applications.

The aim of the present research is to conduct a systematic review of the instruments documented in the literature for assessing individuals' psychological dimensions in relation to social and domestic robots, such as attitudes, beliefs, perceptions, opinions, and emotions. In this review, 'instrument' refers to a specific tool used for data collection and measurement, such as questionnaires, scales, interviews, etc. This review will assess both the instrument psychometric properties and the quality of evidence linked to each property with view to a) providing practitioners and researchers with a comprehensive guide to the available instruments and their psychometric properties, enabling them to make informed choices based on their specific requirements, and b) establishing indications for the future development and validation of such instruments.

METHOD

Search strategy and eligibility criteria

A systematic review was performed following the Preferred Reporting Items for Systematic Reviews and Meta-Analyses guidelines (PRISMA; [34]). The searching strategy was performed from June to July 2023 searching in the following computerised databases: Scopus, PubMed, and IEEEXplore. This was done as searching at least two databases is recommended for the best coverage of the topic and to decrease chances of inappropriate conclusions [35]. The research strategy aimed to find literature related to the validation and the assessment of psychometric properties of instruments designed to evaluate individuals' psychological dimensions in relation to robots. To accomplish this, the index terms: "robot*", "social", "home", "domestic", "questionnaire", "survey", "assessment", "measure*", "psychom*", "valid", and "reliab*" were used. The research strategy has been provided as Supplementary Material (Appendix A).

The research strategy aimed to incorporate instruments suitable for use with social and domestic robots. Nevertheless, those specifically designed for other types of robots were excluded. The search strategy comprised studies meeting the following inclusion criteria: a) the instrument could assess psychological dimensions related to social and domestic robots, including attitudes, beliefs, opinions, feelings, and perceptions; b) the study focused on validating the instrument; c) it evaluated the psychometric properties of the instrument; d) it underwent peer review; e) it was in English. Given the focus of our research, studies centred on industrial robots, rescue robots, robotic arms, or those primarily validating technology or measuring anthropomorphism were excluded. Three members of the research group independently assessed the eligibility of the articles after establishing the criteria with the research team. Initially, the titles and abstracts of the articles resulting from the search were screened based on established criteria. Those that passed the screening were then evaluated through full-text reading. At the end of each step (title/abstract screening and full text screening), inter-rater agreement among the three reviewers was evaluated, indicating good agreement (Fleiss Kappa = 0.83 and 0.92, respectively), and any disagreements were resolved through discussion. The systematic review and protocol were not registered with any relevant database.

Data Extraction

From the included studies, the following data have been extracted for each instrument: the name of the scale, references identified during the systematic review process, the total number of items, a description of the type of items, a description of the construct measured, a description of the subscales if any, the number of items in each subscale, the administration of the instrument, the target population of the instrument, and the characteristics of the population used to validate the instrument, including nationality, sample sizes, and age (range, mean, and standard deviation). Two independent reviewers extracted this data independently, and any disagreements were resolved through consensus with a third reviewer.

Assessment of the instrument properties and methodological quality

The COnsensus-based Standards for the selection of health Measurement INstruments (COSMIN) guidelines were adapted to evaluate instrument properties and the methodological quality of the evidence obtained from the identified studies [36–38]. The instrument properties defined and considered by COSMIN guidelines include content validity (which assesses items relevance, comprehensiveness, and comprehensibility), structural validity, internal consistency, cross-cultural validity, measurement invariance, reliability, measurement error, criterion validity, construct validity, and responsiveness.

The overall rating of each psychometric property per instrument could be sufficient (+), insufficient (-), indeterminate (?) or inconsistent (±) depending on the scores obtained through all the studies for that given measure. While "sufficient" or "insufficient" clearly indicate whether the criteria have been met, the procedure considered studies as "indeterminate" when they addressed the relevant aspect but failed to provide sufficient information to determine if the criteria had been met [36,38]. Furthermore, the "inconsistent" category would signify a combination of both "sufficient" and "insufficient" results. We chose not to resolve inconsistent results but instead to provide this process to individuals interested in utilising the reviewed instruments, considering this study as a valuable summary of the instrument properties available to date. In the context of content validity evaluations, 'insufficient' was assigned to each subcategory (relevance, comprehensiveness, and comprehensibility) when these aspects were not evaluated during the development or validation of the measure. In terms of hypothesis testing for construct validity and considering the multiple dimensions assessed by the identified measures, each study was independently evaluated based on the suggested generic hypothesis [36]: when the instruments measure related but dissimilar constructs, correlations should fall within the range of 0.30-0.50, and when they measure similar constructs, the correlations should be equal to or greater than 0.50. We considered group differences when hypotheses were clearly stated, supported by the literature, and used specifically to assess instrument properties. An important aspect to consider is that due to the various constructs that the instrument could assess and the relatively recent development of these scales, it is challenging to establish a reliable gold standard. Consequently, this study did not assess criterion validity among the indicators used to validate the instruments.

The methodological quality of each instrument property in each study was assessed as either "High," "Moderate," "Low," or "Very Low", following the COSMIN guidelines [36–38]. Subsequently, the overall quality of the body of evidence for each psychometric property could be downgraded based on three factors: risk of bias, inconsistency of findings (less relevant for content validity), and imprecision (low sample sizes). It's worth noting that "indirectness" was not evaluated because the review lacked a defined target population.

Two independent raters extracted data from each record included in the research process and assessed the risk of bias for each psychometric property in each study. Any disagreements that arose were resolved through consensus with a third reviewer.

RESULTS

The search strategy resulted in a total of 3828 articles. After the removal of 556 duplicates, a further 3196 articles were excluded during the title and abstract screening, then of the remaining 76 articles, 42 were excluded during the full text evaluation for not meeting the research criteria.

The overall research strategy resulted in the inclusion of a total of 34 articles evaluating 27 measures aimed at assessing people’s attitudes toward social robots. A summary of the research process is reported in Figure 1. The information extracted for each measure is reported in Table 1, while ratings and quality of evidence are reported in Table 2. Below we discuss the results per instrument, grouping them according to the included population.

PLEASE INSERT FIGURE 1 ABOUT HERE

Table 1. Summary of the data extracted for the instruments identified through the systematic review.

Instrument name	Reference(s)	Total items	Type of items	Construct(s)	Subscales (#items)	Type of administration	Target population	Validation population nationality – N. of participants (M; F)	Age-range; and/or mean (SD) of the validation population
Children's Openness to Interacting with a Robot Scale (COIRS)	[39]	12	0-4 scale	Assess openness to robot interaction	Intrinsic Interest in Interacting with a robot (3); Openness to socioemotional interactions with robot (5); Openness to utilitarian interaction with a robot (4).	Self-report	Children	U.S.-American	Range = 8-11 years
Robot gratification questionnaire*	[40]	26	5-point Likert scale	Assess gratification (sought and obtained)	Hedonic gratification-sought (3); Hedonic gratification-obtained (3); Informative gratification sought (3); Informative gratification obtained (3); Social gratification-obtained (3); Social gratification-sought (3); Experiential gratifications-sought (4); Experiential gratifications-obtained (4).	Observational	Children	Dutch – 24 children (15M; 9F)	Range = 7-11 years Mean = 9.31 (SD= 1.15)
Robot social presence scale*	[41]	17	5-point Likert scale	Assess robot social presence	Perceived presence (4); Interaction behaviour perception (4); Interactive expression and information understanding (4); Perceived emotional interdependence (4); Attention allocation (4); Emotional understanding and expressiveness (4)	Self-report	General population	Chinese – 494 (174M; 320F)	Range = 18-60 years
Robot-Era Inventory (REI)	[42]	41	5-poin	Assess the acceptabilit	Perceived robot personality (11);	Self-report	Older adults	Not reported –	Mean = 62.9 years

			t Likert scale	y of SARs in the older population,	Human Robot Interaction (10); Perceived benefit (6); Easiness of use (6); Perceived usefulness (7)			21 (13M; 8 F)	(SD= 3.9)
Users' Needs, Requirements and Abilities Questionnaire (UNRAQ)	[43,44]	34	5-point Likert scale	Assess older persons' needs and requirements regarding the properties and functions of a robot	Interaction with the robot and technical issues (10); Assistive role of the robot (13); Social aspects of using the robot (6); Ethical issues (5)	Self-report	General population	Not reported – 720 (179M; 541F)	Range= 19-91 years; Mean= 52.0 (SD=37.0)
A revision of the Technology-Specific Expectation Scale (TSES-R)	[45]	20	5-point Likert scale	Parental expectations towards robots for healthcare	Capabilities dimension (5); Fictional view dimension (5); Social/Emotional dimension (4); Playful distraction dimension (3); Assistive role dimension (3)	Self-report	Parents	Not reported	Not reported
General Attitudes Towards Robots Scale (GAToRS)	[46]	20	7-point Likert scale	People's attitude toward robots	Personal Level Positive Attitude (5); Personal Level Negative Attitude (5); Societal Level Positive Attitude (5); Societal Level Negative Attitude (5)	Self-report	General population	Finnish – 477 (192M; 283F)	Mean= 40.23 years(SD= 13.51) all participants were above 18 years old.
Social Perception of Robots Scale (SPRS)	[47]	18	Semantic differential scale	Measure social perception of robots	Anthropomorphism (6); Morality/ Sociability (6); Activity/ cooperation (4)	Self-report	General population	German, Austrian, Swiss and from other countries – 1032 (538M; 480F; 14NB)	Mean=33.83 years SD(12.66)
Social Service Robot Interaction Trust (SSRIT)	[48]	50	5-point Likert scale	Assess trust in interactions with AI social robots in service delivery	Familiarity (4); Robot Use Self-Efficacy (5); Social Influence (4); Technology Attachment (3); Trust Stance in Technology (3); Anthropomorphism (7); Robot Performance (9);	Self-report	Consumers	Ethnicities are reported - Sample 1: 452 (38.9%M; 60.8%F; 0.2%Other) Sample 2: 362 (38.7%M; 61.0%F;	Range=18-over 65 years

					Effort Expectancy (4); Perceived Service Risk (5); Robot-Service Fit (3); Facilitating Robot-use Condition (3)			0.3Other)	
Attitudes towards the use of social robot (ATTUSR-C)	[49]	15	5-point Likert scale	Assess attitudes towards the use of social robot (ATTUSR-C) questionnaire for use with Taiwanese health personnel	Unidimensional	Self-report	Health personnel	Taiwanese – 95 (95%F)	Range=25-63 years; Mean = 44.5 (SD=11.9)
Intentional acceptance of Social Robots	[50]	4	5-point bar scale	Assess children intention to use social robots	Unidimensional	Self-report	Children	Dutch – 87 (39M; 48F)	Range=7-11 years; M= 9.17 (SD=0.85)
Educational robot attitude scale (ERAS)	[51]	17	5-point Likert	Measure the attitudes school students towards the use of humanoid robots in educational settings.	Engagement (5); Enjoyment (4); Anxiety (4); Intention (4)	Self-report	Children	Not reported-232 (128M; 104F)	Range= 13-13 years
Self-Efficacy in Human-Robot-Interaction Scale (SE-HRI)	[52]	18	6-point Likert scale	Measure people's perceived self-efficacy in dealing with robots	Unidimensional	Self-report	General population	German - 450 (288F; 4 unknown) U.S.-American 209 (104M; 105F)	German: range= 18-59 years : Mean =25.15 (SD=6.66) American: range= 16-69 years; Mean= 26.48 (SD=9.11)
Self-Efficacy in Human-Robot-Interaction Scale (SE-HRI) – Short version	[52]	10	6-point Likert scale	Measure people's perceived self-efficacy in dealing with robots	Unidimensional	Self-report	General population	English speaking sample and German speaking sample - 196 (101M; 95F)	Range 18-82 Mean = 36.91 (SD=13.97)
Moral Concern for Robots Scale (MCRS)	[53]	21	7-point Likert	Measure moral concern for robots	Basic moral concern (12); Concern for psychological harm (9)	Self-report	General population	Japanese – group 1: 121 (66M; 55F) and	Mean = 20.1 years (SD= 1.6) for group 1;

			scale					group 2: 200 (100M; 100F)	Range=20's-60's for group 2
Robotic Social Attributes Scale (RoSAS)	[54,55]	18 (note: 10 for Portuguese version)	9-point Likert	Measure social perception of robots	Warmth (6); Competence (6); Discomfort (6)	Self-report	General population	Not reported - 210 (105M; 104F, 1 not identified) Portuguese - 185 (45%M; 55%F)	Not reported Portuguese sample: Range = 18-35 years Mean = 23.40 (SD= 5.21)
HEXACO-60 for HRI*	[56]	60	5-point Likert scale	Evaluate how people perceive the personality traits of robots	Empathy/ Altruism/ Sociability; Integrity; Dependability; Self-confidence	Self-report	General population	Italian – 133 (not reported)	Range 19-65; Mean= 34.46 (SD=14.17)
Sense of Safety and security for robots in eldercare*	[57]	12	Semantic difference scales	Measure sense of safety and security for robots in eldercare	Sense of Safety (6); Sense of Security (6);	Self-report	General population	Not reported – 100 (47M, 53F)	Range 14-62 years; Mean= 35.48 (SD=10.58)
PERNOD (PERception to humaNOiD)	[58]	33	7-point Likert scale	Evaluate humanoid robot	Familiarity (12); Utility (7); Motion (4); Controllability (5); Toughness (5)	Self-report	General population	Japanese – 380 university students (140M; 239F; 1 unknown)	Mean=20.31 years (SD=2.89)
Multi-dimensional Robot Attitude Scale	[59]	49	7-point Likert scale	Assess attitudes toward domestic robot	Familiarity (6); Interest (7); Negative Attitude (5); Self-efficacy (4); Appearance (7); Utility (5); Cost (3); Variety (3)	Self-report	Adults	Japanese – 175 (77.8%M) Taiwanese - 130 (46.9%M) Chinese - (40.5%M)	Japanese: Mean =22.3 years (SD=1.9) Chinese: Mean =23.6 years (SD=1.6) Taiwanese: Mean = 24.2 years (SD=5.0)
Robot use self-efficacy in healthcare work (RUSH)	[60]	6	5-point Likert scale	Measure robot use self-efficacy in healthcare work	Unidimensional	Self-report	Health care workers	Finnish- 3 samples: 200 homecare workers (93.5%F), 1889 Nurses (89.8%F), 1554 nurses and Physiotherapists	Sample1: range 19-65 years, Mean=43.2 (SD=11.8); Sample2: range 17-68 years, Mean=45.5 (SD=12.1); Sample3:

								(95%F)	range=19-70 years, Mean=47.5 (SD=10.4)
Robot Anxiety Scale (RAS)	[61] [62]	11	6-point Likert scale	Measuring the anxiety that prevents individuals from interaction with robots having functions of communication in daily life.	Anxiety toward Communication Capability of Robots (3); Anxiety toward Behavioral Characteristics of Robots (4); Anxiety toward Discourse with Robots (4)	Self-report	General population	Japanese: 400 university students (197M; 199F; 4 unknown) Chinese: sample 1 composed of 305 adults (138M; 167F) and sample 2 composed of 740 adults (319M; 421F)	Japanese: mean =21.4 years; Chinese: sample 1 range= 18-60 years and above; Sample 2 range= 18-60 years
Rapport-Expectation with a Robot Scale (RERS)	[63]	18	7-point Likert scale	Measure people's expectations for rapport	Expectation as a conversation partner (11); Expectation for togetherness (7)	Self-report	General population	Not reported – 20 University students (not reported)	Not reported
Child robot relationship formation*	[64]	13	5-points bar scales	Assess child-robot relationship formation	Closeness (5); Trust (4); Perceived social support (5)	Self-reported	Children middle childhood	Dutch – 87 children (39M; 48F)	Range= 7-11 years, Mean =9.17 (SD=0.85)
Almere model	[21,65]	41 (30 for the Mandarin version)	5-point Likert scale	Acceptance of assistive social agents by older adults.	Anxiety (4); Attitude towards the assistive social agent (3); Facilitating conditions (2); Intention to use (3); Perceived Adaptiveness (3); Perceived enjoyment (5); Perceived ease of use (5); Perceived Sociability (4); Perceived usefulness (3); Social Influence (2); Social presence (5); Trust (2)	Self-reported	Older adults	Dutch – Experiment 1: 40 older adults (18M; 22F) Experiment 2: 88 participants (28M, 60F) Experiment 3: 30 older adults (8M, 22F) Experiment 4: 30 older adults (16M, 14F) Chinese – 317 (55.5%F)	Dutch - Experiment 1: Range= 65-89 years Experiment 2:NA Experiment 3: range= 65-94 years Experiment 4: range 65-89 years Chinese – Mean=70.3 years (SD=7.5)
Frankenstein Syndrome	[66,67]	30	7-poin	Measure acceptance	General anxiety toward humanoid	Self-reporte	General	Japanese (the	Range= 20's-60's

Questionnaire			t Likert scale	of humanoid robots including expectations and anxieties toward this technology in the general public	robots (13); Apprehension toward social risks of humanoid robots (5); Trustworthiness for developers of humanoid robots (4); Expectation for humanoid robots in daily life (5)	d	population	questionnaire is also available in English) – 1000 (500M, 500F)	
Negative Attitude Towards Robot (NARS)	[68–72]	14 (Portuguese and Polish version have 12 items, English version has 11 items)	5-point Likert scale (7-point Likert scale for Portuguese and Polish version)	Measure humans' negative attitudes toward robots	Negative Attitudes toward Situations of Interaction with Robots (6); Negative Attitudes toward the Social Influence of Robots (5); Negative Attitudes toward Emotions in Interaction with Robots (3) Note: Portuguese and Polish version have two dimensions, NARHT: negative attitudes towards robots with human traits; NATIR: negative attitudes towards interactions with robots. English version has three dimension measuring different construct	Self-reported	General population	Japanese – 240 university students (146M, 92F, 2 unknown); U.S.-American – 54 undergraduate students (13M, 41F); Portuguese – four studies total sample of 997 (401M, 598F, 3 not reported) Polish – 213 (80M, 91F, 42Not reported); English– 28 university students and staff (14M,14F)	Japanese - Mean= 22.0 years (not reported) American – range= 18-25 years; Portuguese range= 18-71 years; Polish – Mean= 29.36 years (SD= 10.15); English: range 18-55 years

Notes: * the name of the instrument is not provided in the original article.

Table 2. Table 2. Results obtained from the overall rating (OR): sufficient (+), insufficient (-), indeterminate (?), inconsistent (±) and the Quality of Evidence (QoE): "High," "Moderate," "Low," or "Very Low".

nstrument	Content Validity								Structural Validity		Internal consistency		Cross-cultural validity		Measurement invariance		Reliability		Construct validity	
	Overall		*Relevance*		*Comprehensiveness*		*Comprehensibility*													
	OR	QoE	OR	QoE	OR	QoE	OR	QoE	OR	QoE	OR	QoE	OR	QoE	OR	QoE	OR	QoE	OR	QoE
COIRS	±	Low	-	Low	+	Low	+	Low	+	High	+	High							?	Low
Robot gratificat on question naire*	±	Low	+	Low	-	Low	-	Low	?	Low	±	Low							±	Very low
Robot Social Presence scale*	+	Moderate	+	Moderate	+	Moderate	+	Moderate	+	Moderate	+	High								
REI	-	Very low	±	Low	-	Very low	-	Very low	?	Very low	-	Very low							?	Very low
UNRAQ	-	Very low	-	Very Low	-	Very low	-	Very low			±	High					+	Moderate		
TSES-R	-	Very low	-	Very low	-	Very low	-	Very low	?	Very low	+	moderate								
GAToRS	±	Very low	+	Low	-	Very low	-	Very low	+	High	+	High							±	high
SPRS	-	Very low	-	Low	-	Very low	-	Very low	-	High	±	High							±	High
SSRIT	+	Moderate	+	Moderate	+	Moderate	+	Low	+	High	+	High							+	High
ATTUSR-C	±	Very low	+	Very low	-	Very low	+	Very low	?	Moderate	+	Moderate								
ntentional acceptance of social robots	-	Very low	-	Very low	-	Very low	-	low	+	Moderate	+	Moderate			+	Low			±	Very low
ERAS	-	Moderate	-	Moderate	-	Moderate	+	Moderate	?	High	+	High								
SE-HRI	±	Very low	+	Very low	-	Very low	+	Very low	+	High	+	High							+	High
SE-HRI (short version											+	High							+	High
MCRS	-	Vey low	-	Very low	-	Very low	-	Very low	?	Moderate	+	High							?	Very low
RoSAS	-	Very low	-	Very low	-	Very low	+	Low	+	High	+	High					±	Moderate	±	Moderate
HEXACO-60 for HRI*	-	Very low	±	Very low	-	Very low	-	Very low	?	Very low										
Sense of Safety and security for robots in eldercare*	-	Very low	-	Very low	-	Very low	-	Very low	?	Very low	+	Moderate								
PERNOD	-	Very low	-	Low	-	Very low	-	Very low	?	High	+	High								
Multi-dimensional robot attitude scale	-	Very low	-	Very low	-	Low	-	Very low			±	High								
RUSH	-	Very low	-	Very low	-	Very low	-	Very low	?	High	+	High							-	Very Low
RAS	+	Mode	+	Mode	+	Moder	?	Moder	+	High	+	High							±	High

Instrument	Content Validity								Structural Validity		Internal consistency		Cross-cultural validity		Measurement invarianc		Reliability		Construct validity	
	Overall		*Relevance*		*Comprehensiveness*		*Comprehensibility*													
	OR	QoE	OR	QoE	OR	QoE	OR	QoE	OR	QoE	OR	QoE	OR	QoE	OR	QoE	OR	QoE	OR	QoE
		rate		rate		ate		ate												
RERS	-	Very low	±	Very low	-	Very low	-	Very low	?	Very low	+	Moderate							?	Low
Child robot relationship formation*	±	Very low	±	Very low	-	Very low	+	Low	+	Moderate	+	Moderate							±	Very low
Almere model	+	Moderate	+	Moderate	+	Moderate	?	Moderate	+	High	+	High					+	High	?	Low
Frankenstein Syndrome Questionnaire	-	Moderate	-	Moderate	-	Moderate	-	Moderate	?	Moderate	+	High	-	Very low						
NARS	-	Low	-	Low	-	Low	-	Very low	-	High	+	High	-	Very low			±	Low	±	High

Notes: *the name of the instrument is not reported in the original article. Measurement error and Responsiveness are absent from the table because no article assessed these properties, and Criterion validity is not reported in accordance with the explanation given in the method. Blank cells represent psychometric properties that have not been evaluated for the instrument.

Instruments to assess children’s psychological dimensions toward robots

The *Children's Openness to Interacting with a Robot Scale* (COIRS, [39]) measures openness to new experiences and psychological boundaries related to robot interactions. The scale was developed through focus groups with parents, teachers, and researchers and underwent cognitive pre-testing with colleagues and researchers. During the validation, Exploratory Factor Analysis (EFA) revealed a three-dimensional structure with good internal consistency (α ranging from 0.72 to 0.78) and sufficient structural validity (RMSEA=0.07, CFI=0.93, RMSR=0.07) for the three dimensions. Construct validity was assessed by correlating the average COIRS score with other scales. However, correlations were not performed with the subscales, making the construct validity for each subscale unclear. A comparison by age and gender found no significant differences, though the purpose of the comparison was not reported.

The questionnaire developed by Jong et al. [40] aims to assess children's uses and gratifications of social robots based on the literature on children's media gratifications. After a brief interaction with a social robot, 88 Dutch children were interviewed. Through coding of their responses to an open-ended question, categories of gratifications were identified, and a questionnaire was developed to measure four types of gratifications. The items were derived from previous questionnaires and children's answers. The gratification types were subsequently categorised into sought and obtained, although the theoretical rationale for this choice was not provided. During the validation, two subscales did not reach sufficient internal consistency. The EFA results did not provide information about the goodness of the 8-dimension solution. Some of the subscales did not provide sufficient evidence for the expected hypothesis tested by the authors (Pearson’s correlations ranging from 0.12 to 0.78).

The *Intentional acceptance of social robots* [50] is a unidimensional instrument developed to assess children’s intentional acceptance of social robots defined as children's intention to use a social robot repeatedly and/or for a long term in their daily life. The researchers reviewed existing measures and focused on Heerink et al.'s (2010) scale. They adjusted and refined items by referencing specific activities and adapting the language for children, through discussions, and with suggestions, from primary school teachers. The items were also translated to Dutch. Pilot testing with four children led to further adjustments. The Confirmatory Factor Analysis (CFA) revealed a good fit of the data: $\chi 2$ (2, N = 87) = 3.56, p = .16, CFI = 0.97, SRMR = 0.04. Measurement invariance was assessed between boys and girls, showing sufficient results. Internal consistency showed sufficient results for the overall sample (ranges 0.72-0.85). According to hypothesis testing, the scale showed enough correlation with the enjoyment measure (r = 0.49) but low correlation with other measures, i.e. social presence (r=0.24) and social anxiety (r=-0.20).

The *Educational Robot Attitude Scale* (ERAS; [51]) was developed through a process involving the creation of an item pool based on existing literature. The scale was reviewed by experts for content and face validity. A pilot test with twenty school children was conducted to assess item comprehension. The scale showed a 4-dimension solution according to EFA, fitting indexes are not reported. The reliability of the scale was satisfactory (Cronbach's α ranged from 0.81 to 0.85).

Straten et al. [64] developed a measure to assess child-robot relationships by three self-report scales of closeness, trust, and perceived social support in which constructs are derived from theories of interpersonal relationships. The researchers developed the scales by reviewing existing measures and refining item content, translated into Dutch. Comprehensibility was assessed through teachers, and pilot studies. The measure's validation demonstrated a good model fit based on CFA results: $\chi2$ (62, N = 87) = 62.277, p = .466, CFI = .999, SRMR = .052. Hypothesis testing with concurrently measured variables, which were significantly shortened, yielded mixed results.

Instruments to assess psychological dimensions of adults (aged 18 years and over) towards robots

Chen and colleagues [41] proposed a 6-dimension questionnaire to assess robots' social presence. Researchers retrieved papers related to social presence and identified questions for a human-robot interaction scale, divided into theoretical dimensions, following expert evaluation and translation. Three experts in artificial intelligence, psychology, and sociology respectively assessed the proposed definition and model, tested face validity, and reviewed content and discriminant validity for each dimension of the scale. Then 5 respondents experienced in using social robots were invited for structured interviews to assess the clarity, precision, repetition, conflict, and understandability of the questionnaire. Validation results indicated good fit from EFA ($\chi2$/df = 2.160, RMSEA = 0.048, TLI = 0.928, NFI = 0.939, AGFI = 0.926, SRMR = 0.052, CFI = 0.966, GFI = 0.950.) and Cronbach's alpha were above 0.70.

The *Users' Needs Requirements and Abilities Questionnaire* (UNRAQ; [43,44]) was developed through a process that involved a literature review and collaboration with the ENRICHME project partners. It is an instrument that can be used to collect data about the use of social robots in the care of older people. The validation sample consisted of 720 older adult participants, 125 of them repeated the assessment two weeks apart. Evaluation of psychometric properties indicated good Cronbach's α for each dimension (all above 0.70) and test-retest reliability for each subscale measured by ICC (range: 0.81-0.93).

The *General Attitudes Towards Robots Scale* (GAToRS; [46]) was developed to assess attitude as a pre-disposition to respond favourably or unfavourably to objects in the world and makes a distinction between personal and societal levels of attitudes towards robots differentiating them between positive and negative. In the pilot study, the authors only report that the measure was developed partly based on other instruments. The 4-dimension factor was considered adequate as it fell between the suggested factors of different evaluation methods. Only two of the dimensions had a Cronbach's α above 0.70. The authors developed a revised version of the questionnaire by conducting a pilot study, extracting items from other instruments, and collecting new items through open questions posted in science fiction fandom and robotics-oriented Facebook groups. The authors further refined these items through various EFAs. The final version of the questionnaire consisted of 20 items along with 4 criterion items. A CFA indicated good fit ($\chi2(164) = 429.98$, $p < 0.001$, CFI = 0.91, TLI = 0.896, RMSEA = 0.058, 90% CI = [0.052, 0.064], SRMR = 0.057) and Cronbach's α for each subscale were above 0.70. Correlations with the Negative attitude towards robot scale indicated mixed results (range 0.2 to 0.8), however, authors did not report a specific hypothesis and, given the GAToRS measures attitude towards robots, we would have expected strong correlations of > 0.50.

The *Social Perception of Robots Scale* (SPRS; [47]) was developed as a short scale for measuring social perceptions of robots that comprises of sociability, competence, morality, and anthropomorphism that can be applied to different robots in diverse research settings. Though a definition for each scale is provided, authors did not describe a theoretical background for the social perception dimension and for its subcomponents. The authors composed items based on three different instruments to address the three main dimensions of social perception. The EFA results indicated a 3-dimension factor (anthropomorphism, morality/sociability, activity/cooperation), and a subsequent CFA did not indicate good fit ($\chi2$ (115) = 508.12, $p = .000$; RMSEA = .101; CFI = .796; TLI = .759; SRMR = .096). Regarding internal consistency, the third dimension resulted in a low index (α = 0.64) while the first and second were sufficient (α = 0.82 and α = 0.85, respectively). Regarding hypothesis testing, only some of the expected correlations were confirmed (r range: 0.08-0.96), indicating mixed results.

The *Self-Efficacy in Human-Robot-Interaction Scale* (SE-HRI; [52]) was developed to create a German and an English version of a valid and reliable instrument for measuring people's perceived self-efficacy in dealing with robots. The first version of the SE-HRI Scale consisted of items that were either adapted from different questionnaires or theoretically generated. An EFA indicated a two-dimension solution (namely, Self-Efficacy and Loss of control) which showed good internal consistency (α = 0.945 and α = 0.864, respectively). A CFA was conducted with the German version of the measure and a different sample, however it did not reach sufficient structural validity (χ 2/df-ratio of 5.21 and poor values for the other fit-indices: RMSEA = .097, CFI = .84, and SRMR = .055) and a subsequent analysis with reduced items indicated a 1-factor solution and a good model fit (χ 2/ df-ratio of 2.98, RMSEA = .066, CFI = .95, and SRMR = .029). This result was replicated for the English version. Comprehensibility of the German version was assessed with six older adults. Hypothesis testing performed with correlations indicated sufficient values (r > 0.30), however, we should note that with only one scale, a general self-efficacy measure, was close to this value (r= 0.271 for the German sample and r = 0.298 for the English sample). The authors also proposed a short version based results from EFA. A CFA indicated good fit of the short version for both the German and the English sample; also, hypothesis testing indicated correlations above 0.30, however, this was true also for one scale that the authors used as a discriminant measure.

Nomura and colleagues [53] develop the *Moral Concern for Robots Scale* (MCRS). The definition or a theoretical background of moral concern is not clearly provided. The MCRS was obtained by adopting items from existing questionnaires. Additionally, they created items based on human moral treatment and on scenes of possible robot abuse. Through a questionnaire-based survey, the collected data were analysed using factor analysis, resulting in a two-factor structure. No fitting statistics are reported. Each dimension indicated good internal consistency (α = 0.912 and α = 0.876, respectively). Most of the correlations conducted by the author for construct validity were below 0.30, and two dimensions indicated high correlations with the developed measure namely “Mental state” and “Social Partner” however, they were not assessed with validated measures so these results could not properly be considered as an evidence of construct validity.

The *HEXACO-60 for Robots* [56] is based on the HEXACO model of personality and proposes individuals are characterised by six domains. The authors adapted the items of the HEXACO-60 original questionnaire [73] addressing “a robot” as the subject of each original item. Even if the construct was clearly described and had a theoretical background, a representative population has not been involved in the elicitation of relevant items, thus relevance was considered as indeterminate. The authors performed an EFA which indicated a 4-dimension solution; fitting statistics were not reported.

Akalin and colleagues [57] developed a scale to evaluate the sense of safety and security of robots for older adults care. The authors developed the items after videos of different type of robot interactions were shown to participants; three items were based on the Godspeed questionnaire [74]. Definitions of safety and security used to construct items was not clearly reported. The authors calculated Cronbach’s α for the two dimensions and for each video scenario presented to the participants. All of the Cronbach α reported were above 0.70, indicating good internal consistency of the scales. Factor analysis was performed to identify the most important item associated with the two dimensions.

The *Robotic Social Attributes Scale* (RoSAS) developed by Carpinella and colleagues [54] assesses warmth, competence, and discomfort perceived in robots. While the first two dimensions are drawn from social psychology, they lack a clear definition, making it challenging to assess the content of items related to these dimensions. The development of this scale involved four studies. In the first study, an EFA was conducted on the Goodspeed questionnaire [74], resulting in three factors reflecting anthropomorphism, perceived intelligence, and likeability. In the second study, participants were presented with the Godspeed items, a list of attributes from the Stereotype Content Model, and the Bem Sex Role Inventory [75,76]. Participants indicated whether each item was associated with robots. EFA reduced items and suggested three dimensions: warmth (α = 0.91), competence (α = 0.84), and discomfort (α = 0.82). The third study trialled the developed RoSAS, presenting participants with familiar and unfamiliar animals and human linguistic categories to demonstrate that the dimension of "discomfort" emerges when individuals are evaluating robots. In the fourth study, the questionnaire was validated by comparing different types of robots to assess if participants' perceptions varied based on the scale. However, references to support the hypotheses were not provided. In a separate study, Oliveira and colleagues [55] performed a Portuguese translation of the RoSAS and assessed the comprehension of its items. A CFA suggested that the three dimensions were a good solution (CFI = 0.98, RMSEA = 0.05, SRMSR = 0.06), leading to a reduction in the number of items. Correlations with other measures for construct validity and reliability assessments yielded conflicting results.

The *Rapport-Expectation with a Robot Scale* (RERS; [77]) is designed to measure people's expectations for rapport with robots. To create this scale, students watched science fiction movie clips featuring robots and were asked about their feelings toward interacting with robots, distinguishing between fictional and real robots. Items were developed based on participant responses and from prior research. Subsequently, an EFA was conducted with a small sample, revealing a two-dimensional solution (α = 0.919 and α = 0.848). Unfortunately, no fit indices were reported. To assess construct validity, the same participants were used, with the assumption that there would be variations in their responses based on the different video clips they had viewed. Differences in scores were indeed found, but it is difficult to interpret these results, as there is no provided evidence to support the formulated hypothesis. Subsequently, an experimental task was carried out to evaluate predictive validity. However, the results were inconsistent, as only one out of the two hypotheses were confirmed.

The *Robot Anxiety Scale* (RAS, [61]) was developed to measure anxiety that inhibits people from interacting with robots. The items for this scale were generated through a pilot survey and content validity was assessed. A subsequent EFA revealed a three-dimensional factor solution. Following this, a CFA indicated a good fit (GFI = 0.949, AGFI = 0.917, RMSEA = 0.066) for each scale (α = 0.840, α = 0.844, and α = 0.796). Construct validity was evaluated by comparing the RAS with two other anxiety measures, all showing correlations below 0.30. Cai and colleagues (2023) translated the scale from Japanese to Chinese and assessed its comprehensibility and item content validity. Their study included a CFA that confirmed the RAS's structural validity ($\chi 2/df = 3.26$, SRMR = 0.02, CFI = 0.99, GFI = 0.96, TLI = 0.98, RMSEA = 0.06), correlations for construct validity indicated and good construct validity (absolute values of r ranged from 0.42 to 0.81). Overall, the correlations between the two studies yielded mixed results.

The *Frankenstein Syndrome Questionnaire* (FSQ), developed by Nomura and colleagues [67], is a questionnaire to gauge people's acceptance of humanoid robots. To develop this questionnaire, a survey was conducted to gather opinions, attitudes, and feelings toward humanoid robots from students in both Japan and the UK. A group of experts later reviewed the extracted items for content validity. The questionnaire was administered online, and a factor analysis revealed a four-dimensional solution (α range: 0.693-0.909). Goodness-of-fit indexes were not reported. In a subsequent study, the cross-cultural validity of the FSQ was examined [66], revealing differences in responses between Japanese and UK populations.

Nomura and colleagues [68] developed the *Negative Attitude toward Robots Scale* (NARS) to assess the predispositions in behaviour or reactions toward robots. They initially gathered opinions through a pilot survey, extracting 13 sentences, and obtained an additional 20 sentences from two other measures. The content validity was confirmed through expert discussions. During the validation, an EFA revealed a four-factor structure and. a CFA indicated a good fit (GFI = 0.900, AGFI = 0.856, RMSEA = 0.080), with α coefficients ranging from 0.648 to 0.782. Construct validity was assessed using Pearson's correlation with a measure of anxiety, but all coefficients were below 0.30. Test-retest reliability, assessed with Pearson's correlation, showed mixed results: two subscales had good reliability (r = 0.706 and r = 0.740), but not the 'Negative Attitudes toward Emotions in Interaction with Robots' (r = 0.538). Syrdal et al. [71] assessed the NARS in the English population after translating it. They removed three items, although Cronbach's alpha was not reported for each subscale. They conducted a PCA to assess item loadings on each dimension. Construct validity was assessed with 12 personality traits [78], yielding mixed results. A Portuguese validation of the measure was conducted by Piçarra et al. [69] resulting in a two-dimensional solution and a good model fit (CFI = 0.93, TLI = 0.90, RMSEA = 0.065). Each subscale displayed good internal consistency (α = 0.73 and α = 0.75), although construct validity was not evaluated with other standardised measures. The Polish version of the measure, as conducted by Pochwatko and colleagues (2015), resulted in a two-dimensional solution with two items removed. Both subscales exhibited good internal consistency (α = 0.84 and α = 0.79), but the study did not provide sufficient information to assess construct validity. Xia and LeTendre [72] conducted a cross-cultural validation of the questionnaire, recruiting American and international background students. A CFA confirmed the three-factor structure (CFI = 0.93, TLI = 0.91, RMSEA = 0.08, and SRMR = 0.08), and internal consistency was also confirmed (α ranged from 0.773 to 0.818). The study revealed differences between the two groups of students. It is important to note that the structural validity of the NARS yielded conflicting results, with some studies suggesting a three-dimensional solution while others proposed a two-dimensional solution.

The *PERception to umanoid* (PERNOD) scale developed by Kamide and colleagues [58] is designed to assess people's perspectives when evaluating humanoids. University students were required to describe their impressions after viewing a video recording of a humanoid robot. The responses were categorised into groups and adapted into items. An EFA indicated a five-dimensional solution, with each dimension demonstrating good internal reliability (α ranging from 0.79 to 0.86). However, no CFA or goodness-of-fit indexes were reported.

To assess attitude toward domestic robots, Ninomiya et al. [59] developed the *Multi-dimensional Robot Attitude Scale*. The authors did not provide a specific definition of 'attitude,' and they generated scale items based on descriptions provided by study participants. EFA was conducted, revealing a 12-dimensional structure. Subsequently, two to seven items were selected for each factor based on their loadings, with the aim of ensuring sufficient differentiation among them. The Cronbach's α values for most dimensions exceeded 0.70, except for the 'control' dimension, which had an α of 0.643.

The *Social Service Robot Interaction Trust* (SSRIT; [48]) assesses consumers' trust in interactions with AI social robots. The scale's items were generated through a literature review process and interviews, subsequently evaluated through a focus group. An EFA revealed an 11-factor solution. The Cronbach's α values ranged from 0.82 to 0.94, and a CFA indicated a good model fit (RMSEA = 0.03, CFI = 0.96, TLI = 0.96, SRMR = 0.05). Concurrent validity was assessed by comparing the SSRIT with the Interpersonal Trust Scale [79] and the Technology Artifact Scale [80], revealing high correlations (r = 0.78 and r = 0.84, respectively).

Instruments to assess older adults' psychological dimensions towards robots

The *Robot Era Inventory* (REI; [42]) is designed to measure older adults' acceptance of social robots across five dimensions based on the Robot-Era Model proposed by the author. The inventory items were derived from existing scales found in the literature. A preliminary validation of the questionnaire was conducted. The internal consistency analysis yielded mixed results, with two of the proposed subscales showing insufficient Cronbach's α values (0.67 and 0.69). Construct validity was assessed by examining the correlations between the Robot Era Inventory and the UTAUT, although a clear hypothesis was not reported. The associations between the dimensions showed mixed results.

Additional constructs were considered, and items were adapted from questionnaires present in the literature. A path analysis was used to test hypothesised relations among the dimensions. In an experiment comparing responses to a robot in more social versus less social conditions, differences were found on four subscales. The study showed sufficient internal consistency of the instrument. He et al. [65] translated the Almere Technology Acceptance Questionnaire (ATAQ) into Mandarin Chinese and evaluated its psychometric properties among older adults in China. They performed a content analysis with six experts. EFA followed by CFA revealed a 9-dimension solution ($\chi^2/df = 2.006$, RMSEA = 0.069, RMR = 0.059, GFI = 0.816, IFI = 0.913, TLI = 0.896, CFI = 0.912). Cronbach's alpha coefficients indicated mixed results, ranging from 0.664 to 0.891, indicating varied internal consistency across dimensions. The test-retest reliability coefficient was satisfactory with an overall value of 0.980, and domain-specific values ranging from 0.918 to 0.986 [65].

Instruments to assess psychological dimensions of healthcare professionals towards robots

The *Robot Use Self-Efficacy in Healthcare* (RUSH; [60]) is a measure developed and validated with healthcare workers to assess the self-efficacy of healthcare workers in using robots in their work. There is no reported information regarding the items' development. The validation of the measure indicated sufficient internal consistency ($\alpha = 0.90$), only factor loadings of the factor analysis performed are reported. Regarding correlations performed for construct validity, results obtained for the 6-items version of the measure are not reported. Instead, authors performed a correlation analysis for the short version of the measure (3-item version, RUSH-3). Most correlations with other measures were insufficient ($r < 0.30$), except for one ($r = 0.33$), which was measured with only one item, and its validation is not reported.

The *Chinese version of Attitudes Towards The Use of Social Robot* (ATTUSR-C; [49]) questionnaire is a modified and translated version of the questionnaire proposed by Costescu and David [81]. Although the original version provides a clear definition of “attitude”, the study was not aimed at validating the questionnaire, and there is no evidence of concept elicitation or literature search in item generation. In this version, a panel of five expert academic nursing professors assessed the content validity of the ATTUSR-C questionnaire, rating item clarity and appropriateness. Items with an Item Content Validity Index (I-CVI) below 70% were eliminated. Additionally, 10 clinical instructors assessed the instrument for face validity by evaluating the clarity of each item questionnaire. This process indicated sufficient evidence for relevance and comprehensibility of the items, however, professionals were not asked about comprehensiveness of items, thus it has been evaluated as insufficient. During validation, the EFA interpretation led to a one-dimension solution, with no reported fit indexes. Cronbach α was sufficient (α = 0.84).

Among these measures, we can highlight the *Technology-Specific Expectation of Robots – R* (TSES-R; [45]), which was developed to assess parents' expectations in health-related robot interactions. The scale consists of items adapted from the work of Alves-Oliveira et al. [82], and additional items were created by the authors, organised into three dimensions. PCA was used to determine the item loadings for each dimension; however, fit indices are not reported. Each subscale demonstrated good internal consistency (α = 0.869, α = 0.839, and α = 0.800, respectively). Details regarding the sample used for this analysis are not provided.

DISCUSSION

The use of social robots has generated substantial research interest, and it is unsurprising that numerous studies have explored the variables that influence the human-robot relationship. This exploration is essential for understanding people's attitudes towards these emerging technological tools. The current review to provide both practitioners and researchers with an up-to-date framework of psychometrically validated instruments for assessing the psychological dimensions relevant to the interaction with social and domestic robots. The systematic literature review identified a total of 27 validated measures across 34 articles. These findings suggest a growing interest in psychological constructs related to understanding human-robot relationships, indicating their increasing importance and relevance. Indeed, as detailed in Table 1, the dimensions assessed through the validated scales encompass different constructs.

Although it indicates validated instruments to assess different dimensions, this review also highlights important limitations in terms of psychometric properties. To enhance the quality and accuracy of the available instruments, these limitations should be considered in future development or revisions of instruments for assessing people's psychological dimensions relating to robots. The majority of the instruments (89%) have primarily concentrated on assessing the structural validity (48% of evidence) and internal consistency (69% high quality of evidence) of instruments. Construct validity was considered in the 63% of the instruments (41% high quality of evidence). Cross-cultural validity was evaluated for only two instruments (7%, both of which exhibited low quality of evidence), and measurement invariance for only one instrument (3%, low quality of evidence). Notably, the aspects of measurement error and responsiveness have been disregarded across all the instruments. Content validity was identified in most of the studies (96%); however, none of them exhibited an overall high quality of evidence. Moreover, there is a noticeable scarcity of tools specifically tailored for children, older adults, and healthcare contexts. This highlights the necessity for the development and validation of instruments encompassing a more comprehensive range of psychometric properties. Such an approach is vital for the advancement of this growing area of research, ensuring that assessments are not only thorough but also tailored to the unique characteristics and needs of diverse populations and contexts.

In terms of content validity, many studies have inadequately assessed this property, often demonstrating very low methodological quality and neglecting aspects such as item relevance, comprehensiveness, and comprehensibility. This result aligns with previous findings from other reviews, which have shown that studies often offer unclear definitions of constructs or fail to provide any definition at all [28,29]. Thus, given the interest in validating instruments, it is important to further stress the importance of considering this aspect. Content validity reflects how the content of the scale adequately reflects the construct the instrument is intended to measure [83] and it is considered to be the most important measurement property [37]. Relevance and comprehensiveness refer to how well the items are aligned with the construct of interest, ensuring that all key aspects of the construct are thoroughly evaluated. [38]. Comprehensibility considers how well items are interpreted, which can have an impact on the quality of responses and measurement accuracy [84]. Therefore, careful consideration of the construct's definition and its theoretical basis should be undertaken during the development of this instrument to enhance methodological rigor and improve the quality of the assessment.

Measurement error and responsiveness have not been addressed in any of the studies identified. Measurement error indicates the amount of error, systematic and random, that could not be attributed to a true change in the construct measured [83]. It could be assessed through Minimally Important Change, which indicates if a change in the measurement is considered important [85], or the Smallest Detectable Change which indicates if the change in score is of sufficient magnitude that it has low probability of being a random error [86]. Responsiveness indicates how the instrument could detect change over time in the measured construct [87], which is considered to reflect longitudinal validity [88]. However, even if they are important properties, we should also note that these two properties have a strong emphasis within the clinical context [88,89]. Thus, we suggest considering these two properties with caution and within the context and aim for which the instrument is used.

Most studies have primarily focused on structural validity, typically through exploratory factor analysis (EFA) or confirmatory factor analysis (CFA), as well as internal consistency, as measured by Cronbach’s alpha. However, it is important to note that several studies had “indeterminate” findings on structural validity and did not report goodness-of-fit statistics for their models or provide sufficient information to assess the appropriateness of their structural models. This result expands upon what Naneva et al. [29] previously reported. Thus, the results suggest improvement of the structural assessment of the instruments. It could be suggested to report the goodness of fit in exploratory analysis [90] and to further conduct confirmatory analyses to evaluate model fit in relation to this psychometric property. Indeed, while EFA is an exploratory approach to determine the appropriate number of factors, CFA requires a strong empirical foundation and is typically employed in later phases based on empirical and theoretical grounds [91].

In terms of cross-cultural validity, only two instruments, the NARS [72] and the Frankenstein Syndrome Questionnaire [66], have assessed this aspect. However, in both cases, there is insufficient evidence to demonstrate cross-cultural validity, and the quality of the methodology is low. Despite the challenges associated with considering this property, cross-cultural validity offers valuable instruments for diverse cultures [92]. Its importance is evident from the multinational studies conducted on the topic of human-robot interactions [93,94], which is also highlighted by the diverse nationalities of the validation samples revealed in the present review. This aspect should be given further consideration in the context of psychological measurements for human-robot relations, with particular attention to the methods employed.

Similarly, measurement invariance has been examined in only one study [50]. This is particularly concerning since it would indicate that differences between groups evaluated through most of these instruments could be due to group specific characteristics rather than to true differences in the dimensions assessed by them [95]. Thus, they should be interpreted with some caution.

Reliability, which refers to the proportion of the overall variance in the measure that can be attributed to true differences between individuals [36], or in other words, how the variability observed between individuals is not influenced by errors [96], has been largely overlooked. Only four instruments have provided evidence for the assessment of this property: the UNRAQ, the RoSAS, the Almere model, and the NARS scale [44,55,65,68]. Consequently, many of the identified instruments did not demonstrate reliability, which represents a significant limitation for most available instruments.

When assessing construct validity, a significant proportion of studies have employed correlations with other instruments. Nevertheless, these correlations often yielded inconsistent results. A problem faced in this evaluation was that the majority of studies did not establish hypotheses regarding expected correlations beforehand. In the validation of these measures, it is recommended to formulate valuable and clear hypotheses that address the construct under investigation.

Most of the instruments reviewed have a target population of young adults to older adults; however, we should note that they did not consider measurement invariance due to age-related differential item functioning, thus it could not be established if certain items could favour individuals from different age groups with different backgrounds, or, for instance, due to specific response formats [97]. In this regard, this aspect should be taken into consideration when developing these instruments. Only two instruments, REI [42] and the Almere Model [21], were designed specifically for the older adult population. However, they exhibit limited psychometric properties, indicating the need to develop instruments for this specific demographic group.

Only three instruments considered the clinical context. The RUSH [60] and ATTUSR-C [49] were designed for healthcare professionals, while the TSES-R focused on parents' expectations [45]. However, these three instruments only demonstrated sufficient internal consistency, indicating that there is still a need to develop psychometrically valid and reliable instruments in the healthcare context. This is particularly important given the literature's emphasis on the use of social robots in healthcare settings [98].

The review indicated that only five instruments in the literature are validated for children. Most of them only demonstrated sufficient structural validity and internal consistency, suggesting that the available measures to assess psychological dimensions of children toward robots have important limitations. Consequently, there is a need for the development of improved instruments for children.

Finally, it is worth noting that the majority of the reviewed studies did not effectively utilise item response theory (IRT). While there is some debate regarding the best approach [99], considering the conditions of the validation study (e.g. [100]), authors should also take this framework into consideration.

Despite the review indicating a strong interest in developing instruments to assess the psychological facets of the human-robot relationship, it also highlights that only some psychometric properties are systematically considered, while other important ones tend to be overlooked. Psychometric properties indicate whether the instrument utilised is a valid and reliable form to assess the dimension of interest [31]. Poorly or incompletely validated instruments have limited use for specific conditions, populations and countries [101]. Limitations in these properties may raise concerns regarding the accuracy of reported outcomes in research and in making informed decisions

[102].

A significant limitation of the available instruments is the absence of consideration or clear description of context for the robot's use during development and validation. A precise delineation of the usage context is an integral aspect of instrument development and content validity evaluation [38], which serves to indicate the relevance of the developed items composing the instrument. This is important as preliminary evidence suggests that the context in which the robot is presented or utilised may impact the components of human-robot interaction [29]. This information is critical for practitioners and clinicians, as it indicates the appropriate use of these instruments for specific purposes and clinical population.

This suggests a need to develop instruments with a broader range of psychometric properties through studies with higher methodological quality of evidence. The analysis suggests that, in addition to the commonly assessed psychometric properties, particular attention should be given to content validity, cross-cultural validity, reliability, measurement invariance, and construct validity through rigorous methodologies. Particular attention should be dedicated to targeted groups and the potential application of the instruments in different contexts. Measurement error and responsiveness remain important properties, and their assessment should be guided by the rationale of the developed instrument. Researchers should consider that these properties have significant weight in a clinical context.

While the current study yielded intriguing results, it is important to acknowledge its limitations. The review and analysis in this study primarily focus on questionnaires suitable for assessing social and domestic robots. However, it is crucial to note that questionnaires tailored for other specific types of robots, such as industrial robots, exist and warrant evaluation to offer valuable insights in that domain as well. Additionally, there are alternative measures within the existing literature for assessing psychological constructs in the context of human-robot interaction. In this regard, it is worth mentioning the Godspeed questionnaire [74]. The research strategy in this review focussed on studies dedicated to the validation of instruments, considering eligible those that addressed this aspect as one of their primary objectives. In the context of the relatively novel field of instrument development for human-robot interactions, it was not feasible to identify gold standards for assessing criterion validity in this review. Nevertheless, it's worth emphasizing that the findings from this study may contribute to the identification of gold standards for other instruments in the future. Indeed, there is significant variability in their usage, with some scales being rarely employed, while others are more commonly utilised (e.g., the Almere model) in the literature. The cause of this variability cannot be definitively determined. Notably, the NARS, the Almere Model, and the RAS are the oldest scales identified in the systematic literature, potentially contributing to their continued use. This raises the possibility that certain important constructs may be systematically overlooked, or the psychometric properties of the instruments might be disregarded. The primary objective of this review is to offer a comprehensive overview of the instruments available to measure various dimensions and to conduct a critical selection based on the currently available psychometric properties of these instruments. As another aspect to consider, individuals may have distinct preferences regarding various physical characteristics of robots, and these preferences are likely influenced by personal factors. However, the extent to which appearance can impact or enhance the human-robot relationship remains a topic that requires more comprehensive exploration. Indeed, determining the ideal form that an agent, like a robot, should take is particularly challenging [104].

CONCLUSIONS

Numerous psychometrically validated instruments exist for assessing various psychological constructs within the realm of human-robot relations, applicable to both social robots and domestic robots. This review aimed to provide a comprehensive overview of these instruments, offering insights into their psychometric properties. While there is a notable interest in developing and validating such instruments, this review also puts forth guidelines and considerations for both the creation of new and the review of existing ones. The review indicates the necessity to develop and validate new instruments for human-robot interactions, encompassing more methodologically rigorous approaches and a broader spectrum of psychometric properties. Researchers should carefully consider the targeted populations and the context of use during the development.

Funding Statement

This work was supported by the EPSRC and NIHR (grant number EP/W031809/1, IMACTIVE). The founding sponsors had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the article, and in the decision to publish the results.

Data Availability

Not applicable.

Conflicts of Interest

None declared.

Author Contributions

Conceptualisation, R.V., M.Z., A.D.N., and D.M.; Methodology, R.V., S.M., and D.M.; Formal Analysis, R.V., N.C., and D.M.; Investigation, R.V., N.C., M.S., K.A., and D.M.; Resources, M.Z., A.D.N., and D.M.; Data curation, R.V., and N.C.; Writing - Original Draft, R.V., N.C., M.S., and

K.A.; Visualization, R.V.; Supervision S.M., M.Z., A.D.N. and D.M.; Project administration, S.M. and D.M.; Funding acquisition S.M., M.Z., A.D.N. and D.M.; all the authors contributed to Writing - Review & Editing.

Abbreviations
CFA: confirmatory factor analysis
CFI: comparative fit index
EFA: exploratory factor analysis
GFI: goodness of fit index
RMSEA: root mean square error of approximation
SRMR: standardized root mean squared error
TLI: Tucker-Lewis index

References

1. Mejia C, Kajikawa Y. Bibliometric Analysis of Social Robotics Research: Identifying Research Trends and Knowledgebase. Applied Sciences Multidisciplinary Digital Publishing Institute; 2017 Dec;7(12):1316. doi: 10.3390/app7121316

2. Belpaeme T, Kennedy J, Ramachandran A, Scassellati B, Tanaka F. Social robots for education: A review. Science Robotics American Association for the Advancement of Science; 2018 Aug 15;3(21):eaat5954. doi: 10.1126/scirobotics.aat5954

3. Scoglio AA, Reilly ED, Gorman JA, Drebing CE. Use of Social Robots in Mental Health and Well-Being Research: Systematic Review. Journal of Medical Internet Research 2019 Jul 24;21(7):e13322. doi: 10.2196/13322

4. Camp N, Lewis M, Hunter K, Johnston J, Zecca M, Di Nuovo A, Magistro D. Technology Used to Recognize Activities of Daily Living in Community-Dwelling Older Adults. Int J Environ Res Public Health 2020 Dec 28;18(1):163. PMID:33379319

5. Di Nuovo A, Broz F, Wang N, Belpaeme T, Cangelosi A, Jones R, Esposito R, Cavallo F, Dario P. The multi-modal interface of Robot-Era multi-robot services tailored for the elderly. Intel Serv Robotics 2018 Jan 1;11(1):109–126. doi: 10.1007/s11370-017-0237-6

6. Robinson H, MacDonald B, Broadbent E. The Role of Healthcare Robots for Older People at Home: A Review. Int J of Soc Robotics 2014 Nov 1;6(4):575–591. doi: 10.1007/s12369-014-0242-2

7. Conti D, Trubia G, Buono S, Di Nuovo A, Di Nuovo S. Brief Review of Robotics in Low-Functioning Autism Therapy. 2020.

8. Yu C, Sommerlad A, Sakure L, Livingston G. Socially assistive robots for people with dementia: Systematic review and meta-analysis of feasibility, acceptability and the effect on cognition, neuropsychiatric symptoms and quality of life. Ageing Res Rev 2022 Jun;78:101633. PMID:35462001

9. Camp N, Di Nuovo A, Hunter K, Johnston J, Zecca M, Lewis M, Magistro D. Perceptions of Socially Assistive Robots Among Community-Dwelling Older Adults. In: Cavallo F, Cabibihan J-J, Fiorini L, Sorrentino A, He H, Liu X, Matsumoto Y, Ge SS, editors. Social Robotics Cham: Springer Nature Switzerland; 2022. p. 540–549. doi: 10.1007/978-3-031-24670-8_48

10. Fosch-Villaronga E, Lutz C, Tamò-Larrieux A. Gathering Expert Opinions for Social Robots' Ethical, Legal, and Societal Concerns: Findings from Four International Workshops. Int J of Soc Robotics 2020 May 1;12(2):441–458. doi: 10.1007/s12369-019-00605-z

11. Ramírez-Duque AA, Aycardi LF, Villa A, Munera M, Bastos T, Belpaeme T, Frizera-Neto A, Cifuentes CA. Collaborative and Inclusive Process with the Autism Community: A Case Study in Colombia About Social Robot Design. Int J of Soc Robotics 2021 Apr 1;13(2):153–167. doi: 10.1007/s12369-020-00627-y

12. Vagnetti R, Camp N, Story M, Ait-Belaid K, Bamforth J, Zecca M, Di Nuovo A, Mitra S, Magistro D. Robot Companions and Sensors for Better Living: Defining Needs to Empower Low Socio-economic Older Adults at Home. Qatar; 2023.

13. Winkle K, Senft E, Lemaignan S. LEADOR: A Method for End-To-End Participatory Design of Autonomous Social Robots. Frontiers in Robotics and AI 2021;8. Available from: https://www.frontiersin.org/articles/10.3389/frobt.2021.704119 [accessed Oct 13, 2023]

14. Bartneck C, Nomura T, Kanda T, Suzuki T, Kennsuke K. A cross-cultural study on attitudes towards robots. 2005. doi: 10.13140/RG.2.2.35929.11367

15. Cavallo F, Esposito R, Limosani R, Manzi A, Bevilacqua R, Felici E, Nuovo AD, Cangelosi A, Lattanzio F, Dario P. Robotic Services Acceptance in Smart Environments With Older Adults: User Satisfaction and Acceptability Study. Journal of Medical Internet Research 2018 Sep 21;20(9):e9460. doi: 10.2196/jmir.9460

16. Conti D, Di Nuovo S, Buono S, Di Nuovo A. Robots in Education and Care of Children with Developmental Disabilities: A Study on Acceptance by Experienced and Future Professionals. Int J of Soc Robotics 2017 Jan 1;9(1):51–62. doi: 10.1007/s12369-016-0359-6

17. de Graaf MMA, Ben Allouch S, van Dijk JAGM. Why Would I Use This in My Home? A Model of Domestic Social Robot Acceptance. Human–Computer Interaction Taylor & Francis; 2019 Mar 4;34(2):115–173. doi: 10.1080/07370024.2017.1312406

18. Caruana N, Moffat R, Miguel-Blanco A, Cross ES. Perceptions of intelligence & sentience shape children’s interactions with robot reading companions. Sci Rep Nature Publishing Group; 2023 May 5;13(1):7341. doi: 10.1038/s41598-023-32104-7

19. de Graaf MMA, Ben Allouch S. Exploring influencing variables for the acceptance of social robots. Robotics and Autonomous Systems 2013 Dec 1;61(12):1476–1486. doi: 10.1016/j.robot.2013.07.007

20. Shin D-H, Choo H. Modeling the acceptance of socially interactive robotics: Social presence in human–robot interaction. Interaction Studies John Benjamins; 2011 Jan 1;12(3):430–460. doi: 10.1075/is.12.3.04shi

21. Heerink M, Kröse B, Evers V, Wielinga B. Assessing Acceptance of Assistive Social Agent Technology by Older Adults: the Almere Model. Int J of Soc Robotics 2010 Dec 1;2(4):361–375. doi: 10.1007/s12369-010-0068-5

22. Chatzoglou PD, Lazaraki V, Apostolidis SD, Gasteratos AC. Factors Affecting Acceptance of Social Robots Among Prospective Users. Int J of Soc Robotics 2023 Jul 6; doi: 10.1007/s12369-023-01024-x

23. Eckel C, Wilson R. Judging a Book by Its Cover: Beauty and Expectations in a Trust Game. Political Research Quarterly 2006 Jun 1;59. doi: 10.1177/106591290605900202

24. Chen Y-C, Yeh S-L, Lin W, Yueh H-P, Fu L-C. The Effects of Social Presence and Familiarity on Children–Robot Interactions. Sensors Multidisciplinary Digital Publishing Institute; 2023 Jan;23(9):4231. doi: 10.3390/s23094231

25. Ziemke T. Understanding Social Robots: Attribution of Intentional Agency to Artificial and Biological Bodies. Artificial Life 2023 Aug 1;29(3):351–366. doi: 10.1162/artl_a_00404

26. Dou X, Wu C-F, Wang X, Niu J. User Expectations of Social Robots in Different Applications:

An Online User Study. In: Stephanidis C, Kurosu M, Degen H, Reinerman-Jones L, editors. HCI International 2020 - Late Breaking Papers: Multimodality and Intelligence Cham: Springer International Publishing; 2020. p. 64–72. doi: 10.1007/978-3-030-60117-1_5

27. David D, Thérouanne P, Milhabet I. The acceptability of social robots: A scoping review of the recent literature. Computers in Human Behavior 2022 Dec;137:107419. doi: 10.1016/j.chb.2022.107419

28. Krägeloh CU, Bharatharaj J, Sasthan Kutty SK, Nirmala PR, Huang L. Questionnaires to Measure Acceptability of Social Robots: A Critical Review. Robotics 2019 Oct 21;8(4):88. doi: 10.3390/robotics8040088

29. Naneva S, Sarda Gou M, Webb TL, Prescott TJ. A Systematic Review of Attitudes, Anxiety, Acceptance, and Trust Towards Social Robots. Int J of Soc Robotics 2020 Dec 1;12(6):1179–1201. doi: 10.1007/s12369-020-00659-4

30. Streiner DL, Norman GR, Cairney J. Health Measurement Scales: A Practical Guide to Their Development and Use. Oxford University Press; 2015. ISBN:978-0-19-968521-9

31. Ginty AT. Psychometric Properties. In: Gellman MD, Turner JR, editors. Encyclopedia of Behavioral Medicine New York, NY: Springer; 2013. p. 1563–1564. doi: 10.1007/978-1-4419-1005-9_480

32. Mokkink LB, Terwee CB, Stratford PW, Alonso J, Patrick DL, Riphagen I, Knol DL, Bouter LM, de Vet HCW. Evaluation of the methodological quality of systematic reviews of health status measurement instruments. Qual Life Res 2009 Apr 1;18(3):313–333. doi: 10.1007/s11136-009-9451-9

33. Terwee CB, Mokkink LB, Knol DL, Ostelo RWJG, Bouter LM, de Vet HCW. Rating the methodological quality in systematic reviews of studies on measurement properties: a scoring system for the COSMIN checklist. Qual Life Res 2012 May 1;21(4):651–657. doi: 10.1007/s11136-011-9960-1

34. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, Shamseer L, Tetzlaff JM, Akl EA, Brennan SE, Chou R, Glanville J, Grimshaw JM, Hróbjartsson A, Lalu MM, Li T, Loder EW, Mayo-Wilson E, McDonald S, McGuinness LA, Stewart LA, Thomas J, Tricco AC, Welch VA, Whiting P, Moher D. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ British Medical Journal Publishing Group; 2021 Mar 29;372:n71. PMID:33782057

35. Ewald H, Klerings I, Wagner G, Heise TL, Stratil JM, Lhachimi SK, Hemkens LG, Gartlehner G, Armijo-Olivo S, Nussbaumer-Streit B. Searching two or more databases decreased the risk of missing relevant studies: a metaresearch study. Journal of Clinical Epidemiology 2022 Sep 1;149:154–164. doi: 10.1016/j.jclinepi.2022.05.022

36. Mokkink LB, Prinsen C, Patrick DL, Alonso J, Bouter LM, De Vet HC, Terwee CB, Mokkink L. COSMIN methodology for systematic reviews of patient-reported outcome measures (PROMs). User manual 2018;78(1):6–3.

37. Prinsen CAC, Mokkink LB, Bouter LM, Alonso J, Patrick DL, de Vet HCW, Terwee CB. COSMIN guideline for systematic reviews of patient-reported outcome measures. Qual Life

Res 2018 May 1;27(5):1147–1157. doi: 10.1007/s11136-018-1798-3

38. Terwee CB, Prinsen C, Chiarotto A, De Vet H, Bouter LM, Alonso J, Westerman MJ, Patrick DL, Mokkink LB. COSMIN methodology for assessing the content validity of PROMs–user manual. Amsterdam: VU University Medical Center 2018; Available from: https://cosmin.nl/wp-content/uploads/COSMIN-methodology-for-content-validity-user-manual-v1.pdf [accessed Oct 12, 2023]

39. Robert D, van den Bergh V. Children's Openness to Interacting with a Robot Scale (COIRS). The 23rd IEEE International Symposium on Robot and Human Interactive Communication 2014. p. 930–935. doi: 10.1109/ROMAN.2014.6926372

40. Jong C de, Kühne R, Peter J, Straten CLV, Barco A. What Do Children Want from a Social Robot? Toward Gratifications Measures for Child-Robot Interaction. 2019 28th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN) 2019. p. 1–8. doi: 10.1109/RO-MAN46459.2019.8956319

41. Chen N, Liu X, Zhai Y, Hu X. Development and validation of a robot social presence measurement dimension scale. Sci Rep Nature Publishing Group; 2023 Feb 19;13(1):2911. doi: 10.1038/s41598-023-28817-4

42. Bevilacqua R, Di Rosa M, Riccardi GR, Pelliccioni G, Lattanzio F, Felici E, Margaritini A, Amabili G, Maranesi E. Design and Development of a Scale for Evaluating the Acceptance of Social Robotics for Older People: The Robot Era Inventory. Frontiers in Neurorobotics 2022;16. Available from: https://www.frontiersin.org/articles/10.3389/fnbot.2022.883106 [accessed Jun 9, 2023]

43. Tobis S, Cylkowska-Nowak M, Wieczorowska-Tobis K, Pawlaczyk M, Suwalska A. Occupational Therapy Students' Perceptions of the Role of Robots in the Care for Older People Living in the Community. Occup Ther Int 2017 Feb 7;2017:9592405. PMID:29097983

44. Tobis S, Neumann-Podczaska A, Kropinska S, Suwalska A. UNRAQ—A Questionnaire for the Use of a Social Robot in Care for Older Persons. A Multi-Stakeholder Study and Psychometric Properties. International Journal of Environmental Research and Public Health 2021 Jun 7;18:6157. doi: 10.3390/ijerph18116157

45. Zhang F, Broz F, Ferrari O, Barakova E. TSES-R: An Extended Scale for Measuring Parental Expectations toward Robots for Children in Healthcare. 2023. p. 262. doi: 10.1145/3568294.3580084

46. Koverola M, Kunnari A, Sundvall J, Laakasuo M. General Attitudes Towards Robots Scale (GAToRS): A New Instrument for Social Surveys. Int J of Soc Robotics 2022 Sep 1;14(7):1559–1581. doi: 10.1007/s12369-022-00880-3

47. Mandl S, Bretschneider M, Asbrock F, Meyer B, Strobel A. The Social Perception of Robots Scale (SPRS): Developing and Testing a Scale for Successful Interaction Between Humans and Robots. In: Camarinha-Matos LM, Ortiz A, Boucher X, Osório AL, editors. Collaborative Networks in Digitalization and Society 50 Cham: Springer International Publishing; 2022. p. 321–334. doi: 10.1007/978-3-031-14844-6_26

48. Chi OH, Jia S, Li Y, Gursoy D. Developing a formative scale to measure consumers' trust

toward interaction with artificially intelligent (AI) social robots in service delivery. Computers in Human Behavior 2021 May 1;118:106700. doi: 10.1016/j.chb.2021.106700

49. Chen S-C, Jones C, Moyle W. Health Professional and Workers Attitudes Towards the Use of Social Robots for Older Adults in Long-Term Care. Int J of Soc Robotics 2020 Nov 1;12(5):1135–1147. doi: 10.1007/s12369-019-00613-z

50. de Jong C, Kühne R, Peter J, van Straten CL, Barco A. Intentional acceptance of social robots: Development and validation of a self-report measure for children. International Journal of Human-Computer Studies 2020 Jul 1;139:102426. doi: 10.1016/j.ijhcs.2020.102426

51. Sisman B, Gunay D, Kucuk S. Development and validation of an educational robot attitude scale (ERAS) for secondary school students. Interactive Learning Environments Taylor & Francis; 2019;27(3):377–388.

52. Pütten AR-VD, Bock N. Development and Validation of the Self-Efficacy in Human-Robot-Interaction Scale (SE-HRI). J Hum-Robot Interact 2018 Dec 5;7(3):21:1-21:30. doi: 10.1145/3139352

53. Nomura T, Kanda T, Yamada S. Measurement of Moral Concern for Robots. 2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI) 2019. p. 540–541. doi: 10.1109/HRI.2019.8673095

54. Carpinella CM, Wyman AB, Perez MA, Stroessner SJ. The Robotic Social Attributes Scale (RoSAS): Development and Validation. 2017 12th ACM/IEEE International Conference on Human-Robot Interaction (HRI 2017. p. 254–262.

55. Oliveira R, Arriaga P, Stroessner SJ, Paiva A. Preliminary validation of the European Portuguese version of the Robotic Social Attributes Scale (RoSAS). Human Behavior and Emerging Technologies 2021;3(5):750–758. doi: 10.1002/hbe2.311

56. Siri G, Marchesi S, Wykowska A, Chiorri C. The Personality of a Robot. An Adaptation of the HEXACO – 60 as a Tool for HRI. In: Li H, Ge SS, Wu Y, Wykowska A, He H, Liu X, Li D, Perez-Osorio J, editors. Social Robotics Cham: Springer International Publishing; 2021. p. 707–717. doi: 10.1007/978-3-030-90525-5_62

57. Akalin N, Kiselev A, Kristoffersson A, Loutfi A. An Evaluation Tool of the Effect of Robots in Eldercare on the Sense of Safety and Security. In: Kheddar A, Yoshida E, Ge SS, Suzuki K, Cabibihan J-J, Eyssel F, He H, editors. Social Robotics Cham: Springer International Publishing; 2017. p. 628–637. doi: 10.1007/978-3-319-70022-9_62

58. Kamide H, Takubo T, Ohara K, Mae Y, Arai T. Impressions of Humanoids: The Development of a Measure for Evaluating a Humanoid. Int J of Soc Robotics 2014 Jan 1;6(1):33–44. doi: 10.1007/s12369-013-0187-x

59. Ninomiya T, Fujita A, Suzuki D, Umemuro H. Development of the Multi-dimensional Robot Attitude Scale: Constructs of People's Attitudes Towards Domestic Robots. 2015. p. 491. doi: 10.1007/978-3-319-25554-5_48ISBN:978-3-319-25553-8

60. Turja T, Rantanen T, Oksanen A. Robot use self-efficacy in healthcare work (RUSH): development and validation of a new measure. AI & Soc 2019 Mar 1;34(1):137–143. doi:

10.1007/s00146-017-0751-2

61. Nomura T, Suzuki T, Kanda T, Kato K. Measurement of Anxiety toward Robots. ROMAN 2006 - The 15th IEEE International Symposium on Robot and Human Interactive Communication 2006. p. 372–377. doi: 10.1109/ROMAN.2006.314462

62. Cai J, Sun Y, Niu C, Qi W, Fu X. Validity and Reliability of the Chinese Version of Robot Anxiety Scale in Chinese Adults. International Journal of Human–Computer Interaction Taylor & Francis; 2023 Mar 14;0(0):1–10. doi: 10.1080/10447318.2023.2188535

63. Nomura T, Kanda T. Rapport–expectation with a robot scale. International Journal of Social Robotics Springer; 2016;8:21–30.

64. Straten CLV, Kühne R, Peter J, De Jong C, Barco A. Closeness, trust, and perceived social support in child-robot relationship formation: Development and validation of three self-report scales. IS 2020 Jan 24;21(1):57–84. doi: 10.1075/is.18052.str

65. He Y, Liu Q, He Q, Li L. Translation, Adaptation, and Psychometric Testing of the Almere Technology Acceptance Questionnaire (ATAQ) Among Older Adults in China. Clin Interv Aging 2022 Sep 12;17:1353–1364. PMID:36117570

66. Nomura T, Syrdal DS, Dautenhahn K. Differences on Social Acceptance of Humanoid Robots between Japan and the UK. 2015;

67. Nomura T, Sugimoto K, Syrdal DS, Dautenhahn K. Social acceptance of humanoid robots in Japan: A survey for development of the frankenstein syndorome questionnaire. 2012 12th IEEE-RAS International Conference on Humanoid Robots (Humanoids 2012) Osaka, Japan: IEEE; 2012. p. 242–247. doi: 10.1109/HUMANOIDS.2012.6651527

68. Nomura T, Suzuki T, Kanda T, Kato K. Measurement of negative attitudes toward robots. IS 2006 Nov 13;7(3):437–454. doi: 10.1075/is.7.3.14nom

69. Piçarra N, Giger J-C, Pochwatko G, Gonçalves G. Validation of the Portuguese version of the Negative Attitudes towards Robots Scale. European Review of Applied Psychology 2015 Mar 1;65(2):93–104. doi: 10.1016/j.erap.2014.11.002

70. Pochwatko G, Giger J-C, Różańska-Walczuk M, Świdrak J, Kukielka K, Mozaryn J, Piçarra N. Polish Version of the Negative Attitude Toward Robots Scale (NARS-PL). Journal of Automation, Mobile Robotics and Intelligent Systems 2015 Jul 27;65–72. doi: 10.14313/JAMRIS_2-2015/25

71. Syrdal DS, Dautenhahn K, Koay KL, Walters ML. The Negative Attitudes Towards Robots Scale and reactions to robot behaviour in a live Human-Robot Interaction study. The Society for the Study of Artificial Intelligence and the Simulation of Behaviour (AISB); 2009 Apr 1; Available from: http://uhra.herts.ac.uk/handle/2299/9641 [accessed Jun 30, 2023]

72. Xia Y, LeTendre G. Robots for Future Classrooms: A Cross-Cultural Validation Study of “Negative Attitudes Toward Robots Scale” in the U.S. Context. Int J of Soc Robotics 2021 Jul 1;13(4):703–714. doi: 10.1007/s12369-020-00669-2

73. Ashton MC, Lee K. The HEXACO–60: A Short Measure of the Major Dimensions of

Personality. Journal of Personality Assessment Routledge; 2009 Jul 1;91(4):340–345. PMID:20017063

74. Bartneck C, Kulić D, Croft E, Zoghbi S. Measurement Instruments for the Anthropomorphism, Animacy, Likeability, Perceived Intelligence, and Perceived Safety of Robots. Int J of Soc Robotics 2009 Jan 1;1(1):71–81. doi: 10.1007/s12369-008-0001-3

75. Bem SL. The measurement of psychological androgyny. Journal of Consulting and Clinical Psychology US: American Psychological Association; 1974;42(2):155–162. doi: 10.1037/h0036215

76. Fiske ST, Cuddy AJC, Glick P. Universal dimensions of social cognition: warmth and competence. Trends in Cognitive Sciences 2007 Feb 1;11(2):77–83. doi: 10.1016/j.tics.2006.11.005

77. Nomura T, Kanda T. Rapport–Expectation with a Robot Scale. Int J of Soc Robotics 2016 Jan 1;8(1):21–30. doi: 10.1007/s12369-015-0293-z

78. Eysenck M. Personality and Individual Differences. Springer New York, NY; 1985. Available from: https://link.springer.com/book/9781461294702 [accessed Oct 5, 2023]

79. McKnight DH, Choudhury V, Kacmar C. Developing and Validating Trust Measures for e-Commerce: An Integrative Typology. Information Systems Research INFORMS; 2002 Sep;13(3):334–359. doi: 10.1287/isre.13.3.334.81

80. Mcknight DH, Carter M, Thatcher JB, Clay PF. Trust in a specific technology: An investigation of its components and measures. ACM Trans Manage Inf Syst 2011 Jul 1;2(2):12:1-12:25. doi: 10.1145/1985347.1985353

81. Costescu CA, David DO. Attitudes toward Using Social Robots in Psychotherapy. Erdelyi Pszichologiai Szemle = Transylvanian Journal of Psychology Cluj-Napoca, Romania: Babes-Bolyai University, Department of Applied Psychology; 2014;15(1):3–20.

82. Alves-Oliveira P, Ribeiro T, Petisca S, di Tullio E, Melo FS, Paiva A. An Empathic Robotic Tutor for School Classrooms: Considering Expectation and Satisfaction of Children as End-Users. In: Tapus A, André E, Martin J-C, Ferland F, Ammi M, editors. Social Robotics Cham: Springer International Publishing; 2015. p. 21–30. doi: 10.1007/978-3-319-25554-5_3

83. Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, Bouter LM, de Vet HCW. The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. J Clin Epidemiol 2010 Jul;63(7):737–745. PMID:20494804

84. Lenzner T. Effects of Survey Question Comprehensibility on Response Quality. Field Methods SAGE Publications Inc; 2012 Nov 1;24(4):409–428. doi: 10.1177/1525822X12448166

85. de Vet HCW, Ostelo RWJG, Terwee CB, van der Roer N, Knol DL, Beckerman H, Boers M, Bouter LM. Minimally important change determined by a visual method integratingan anchor-based and a distribution-based approach. Qual Life Res 2007 Feb 1;16(1):131–142. doi: 10.1007/s11136-006-9109-9

86. Polit DF, Yang FM. Measurement and the measurement of change: a primer for the health professions. Wolters Kluwer Philadelphia; 2016.

87. Guyatt GH, Deyo RA, Charlson M, Levine MN, Mitchell A. Responsiveness and validity in health status measurement: A clarification. Journal of Clinical Epidemiology 1989 Jan 1;42(5):403–408. doi: 10.1016/0895-4356(89)90128-5

88. Angst F. The new COSMIN guidelines confront traditional concepts of responsiveness. BMC Medical Research Methodology 2011 Nov 18;11(1):152. doi: 10.1186/1471-2288-11-152

89. Terwee CB, Peipert JD, Chapman R, Lai J-S, Terluin B, Cella D, Griffiths P, Mokkink LB. Minimal important change (MIC): a conceptual clarification and systematic review of MIC estimates of PROMIS measures. Qual Life Res 2021 Oct 1;30(10):2729–2754. doi: 10.1007/s11136-021-02925-y

90. Fabrigar LR, Wegener DT. Exploratory factor analysis. Oxford University Press; 2011. Available from: https://books.google.com/books?hl=it&lr=&id=DSppAgAAQBAJ&oi=fnd&pg=PP1&dq=Exploratory+Factor+Analysis+Leandre+R.+Fabrigar&ots=amOkVSe9PA&sig=EhHpYGZ1ktBb0pUjpxETKji-dGE [accessed Oct 18, 2023]

91. Hoyle RH. Handbook of structural equation modeling. Guilford press; 2012. Available from: https://books.google.com/books?hl=it&lr=&id=qC4aMfXL1JkC&oi=fnd&pg=PR1&dq=Handbook+of+structural+equation+modeling&ots=EAVnauwDmQ&sig=gqSjVVIM6daruC2jQBIVm-Z8TSE [accessed Oct 18, 2023]

92. Huang WY, Wong SH. Cross-Cultural Validation. In: Michalos AC, editor. Encyclopedia of Quality of Life and Well-Being Research Dordrecht: Springer Netherlands; 2014. p. 1369–1371. doi: 10.1007/978-94-007-0753-5_630

93. Conti D, Cattani A, Di Nuovo S, Di Nuovo A. Are Future Psychologists Willing to Accept and Use a Humanoid Robot in Their Practice? Italian and English Students' Perspective. Frontiers in Psychology 2019;10. Available from: https://www.frontiersin.org/articles/10.3389/fpsyg.2019.02138 [accessed Dec 4, 2023]

94. Papadopoulos I, Wright S, Koulouglioti C, Ali S, Lazzarino R, Martín-García Á, Oter-Quintana C, Kouta C, Rousou E, Papp K, Krepinska R, Tothova V, Malliarou M, Apostolara P, Lesińska-Sawicka M, Nagorska M, Liskova M, Nortvedt L, Alpers L-M, Biglete-Pangilinan S, Oconer-Rubiano MaF, Chaisetsampun W, Wichit N, Ghassemi A-E, Jafarjalal E, Zorba A, Kuckert-Wöstheinrich A, Malla R, Toda T, Akman Ö, Öztürk C, Puvimanasinghe T, Ziaian T, Eldar-Regev O, Nissim S. Socially assistive robots in health and social care: Acceptance and cultural factors. Results from an exploratory international online survey. Japan Journal of Nursing Science 2023;20(2):e12523. doi: 10.1111/jjns.12523

95. Gregorich SE. Do Self-Report Instruments Allow Meaningful Comparisons Across Diverse Population Groups? Testing Measurement Invariance Using the Confirmatory Factor Analysis Framework. Med Care 2006 Nov;44(11 Suppl 3):S78–S94. PMID:17060839

96. de Vet HCW, Terwee CB, Knol DL, Bouter LM. When to use agreement versus reliability measures. Journal of Clinical Epidemiology 2006 Oct 1;59(10):1033–1039. doi:

10.1016/j.jclinepi.2005.10.015

97. Ownby RL, Waldrop-Valverde D. Differential item functioning related to age in the reading subtest of the test of functional health literacy in adults. J Aging Res 2013;2013:654589. PMID:24089638

98. González-González CS, Violant-Holz V, Gil-Iranzo RM. Social Robots in Hospitals: A Systematic Review. Applied Sciences Multidisciplinary Digital Publishing Institute; 2021 Jan;11(13):5976. doi: 10.3390/app11135976

99. Sébille V, Hardouin J-B, Le Néel T, Kubis G, Boyer F, Guillemin F, Falissard B. Methodological issues regarding power of classical test theory (CTT) and item response theory (IRT)-based approaches for the comparison of patient-reported outcomes in two groups of patients - a simulation study. BMC Med Res Methodol 2010 Mar 25;10:24. PMID:20338031

100. Jabrayilov R, Emons WHM, Sijtsma K. Comparison of Classical Test Theory and Item Response Theory in Individual Change Assessment. Appl Psychol Meas 2016 Nov;40(8):559–572. PMID:29881070

101. O'Connor A, McGarr O, Cantillon P, McCurtin A, Clifford A. Clinical performance assessment tools in physiotherapy practice education: a systematic review. Physiotherapy 2018 Mar 1;104(1):46–53. doi: 10.1016/j.physio.2017.01.005

102. Friberg J. Considerations for test selection: How do validity and reliability impact diagnostic decisions? Child Language Teaching & Therapy - CHILD LANG TEACH THER 2010 Apr 8;26:77–92. doi: 10.1177/0265659009349972

103. Terwee CB, Prinsen CAC, Chiarotto A, Westerman MJ, Patrick DL, Alonso J, Bouter LM, de Vet HCW, Mokkink LB. COSMIN methodology for evaluating the content validity of patient-reported outcome measures: a Delphi study. Qual Life Res 2018 May 1;27(5):1159–1170. doi: 10.1007/s11136-018-1829-0

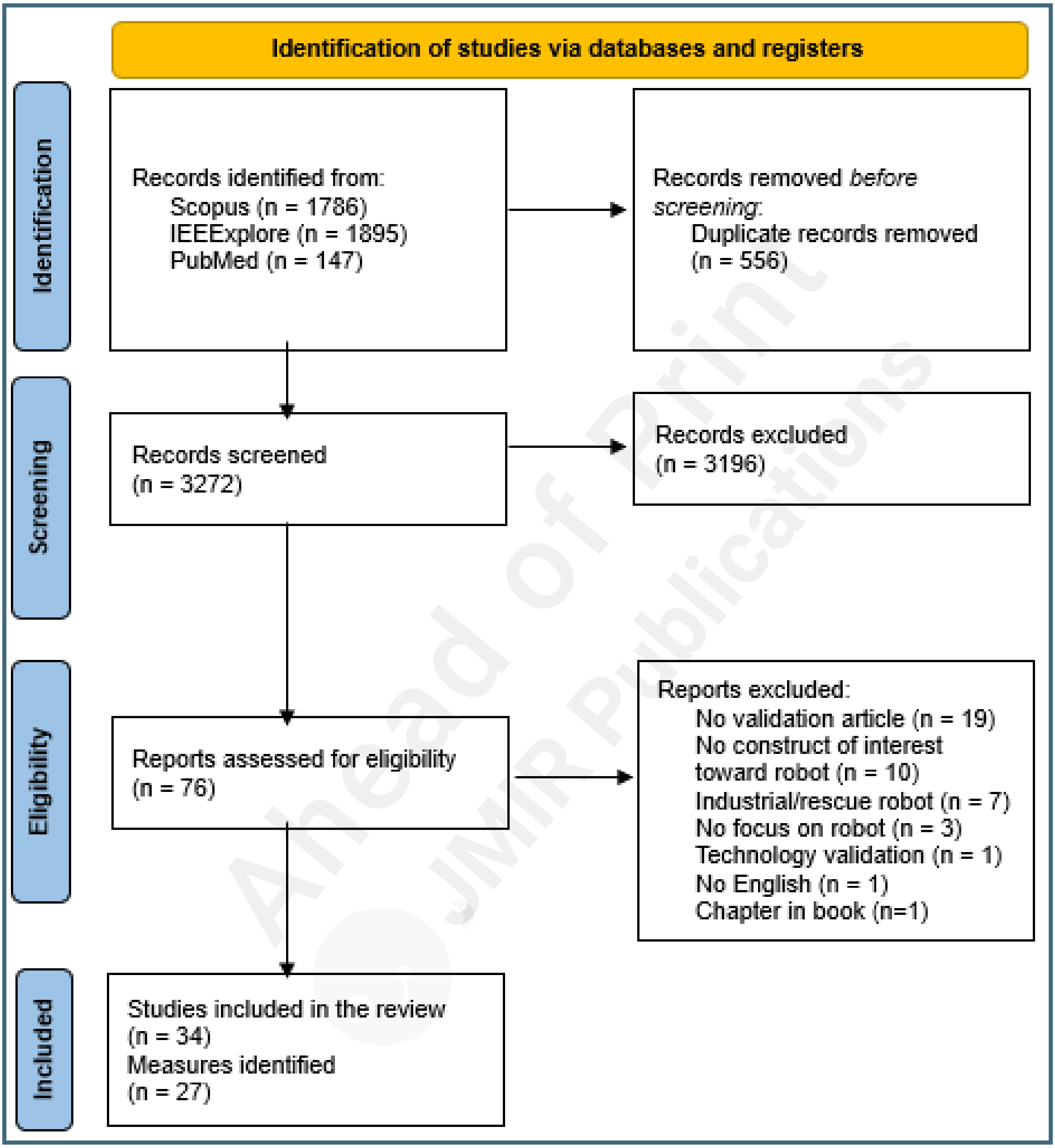
104. Seaborn K, Sekiguchi T, Tokunaga S, Miyake NP, Otake-Matsuura M. Voice Over Body? Older Adults' Reactions to Robot and Voice Assistant Facilitators of Group Conversation. Int J of Soc Robotics 2023 Feb 1;15(2):143–163. doi: 10.1007/s12369-022-00925-7

Supplementary Files

Manuscript revised.
URL: http://asset.jmir.pub/assets/4c5247b48ef4da65b79a242bdf2b52cf.docx

Figures

Flowchart search and selection.



Multimedia Appendixes

Research strategy for each database.
URL: http://asset.jmir.pub/assets/6649c6737a12f1d081277161f549580d.docx

CONSORT (or other) checklists

PRISMA checklist.
URL: http://asset.jmir.pub/assets/65d844b88592a6be4beb9457ccc2901f.pdf

PRISMA abstract checklist.
URL: http://asset.jmir.pub/assets/c8ec2a107bda03146a5326cc334eb264.pdf