Review

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

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

[1]Department of Sport Science, School of Science and Technology, Nottingham Trent University, Nottingham, United Kingdom

[2]Department of Computing & Advanced Wellbeing Research Centre, Sheffield Hallam University, Sheffield, United Kingdom

[3]Wolfson School of Mechanical, Electrical and Manufacturing Engineering, Loughborough University, Loughborough, United Kingdom

[4]Department of Psychology, Nottingham Trent University, Nottingham, United Kingdom

Corresponding Author:
Daniele Magistro, PhD
Department of Sport Science
School of Science and Technology
Nottingham Trent University
College Drive
Clifton
Nottingham, NG11 8NS
United Kingdom
Phone: 44 115 848 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.

Objective: This study aims to provide a systematic review of the instruments available for assessing the psychological aspects of the relationship 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 (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines across different databases: Scopus, PubMed, and IEEE Xplore. The search strategy encompassed studies meeting the following inclusion criteria: (1) the instrument could assess psychological dimensions related to social and domestic robots, including attitudes, beliefs, opinions, feelings, and perceptions; (2) the study focused on validating the instrument; (3) the study evaluated the psychometric properties of the instrument; (4) the study underwent peer review; and (5) the study was in English. Studies focusing on industrial robots, rescue robots, or robotic arms or those primarily concerned with technology validation or measuring anthropomorphism were excluded. Independent reviewers extracted instrument properties and the methodological quality of their evidence following the Consensus-Based Standards for the Selection of Health Measurement Instruments guidelines.

Results: From 3828 identified records, the search strategy yielded 34 (0.89%) 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 (24/27, 89%) and internal consistency (26/27, 96%), consideration of other psychometric properties was frequently inconsistent or absent. No instrument evaluated measurement error and responsiveness despite their significance in the clinical context. Most of the instruments (17/27, 63%) were targeted at both adults and older adults (aged ≥18 years). There was a limited number of instruments specifically designed for children, older adults, and health care 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 consider a broader range of psychometric properties. This

XSL•FO
RenderX

is essential to ensure the creation of reliable and valid measures for assessing people's psychological dimensions regarding social and domestic robots. Among its limitations, this review included instruments applicable to both social and domestic robots while excluding those for other specific types of robots (eg, industrial robots).

KEYWORDS

Introduction

Background

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 cover widespread 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 people with autism spectrum disorder [7] or dementia [8].

While many studies have explored users' opinions and requirements to design and develop this technology to meet their needs in a participatory design framework [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 the perspectives and preferences of people regarding robots represents a crucial point for their development and acceptance [15-17]. How a robot is perceived plays a major role in the human-robot relationship [18]. Existing literature has identified several factors linked to individuals' predispositions toward robots and how robots are used [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 focusing on different dimensions related to human-robot interactions with social robots reportedly indicate that most of the assessments are made using 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 use of instruments in research, it is crucial to possess an understanding of instrument properties and 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 is meant to assess 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.

Objectives

The aim of this research was 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, and interviews. This review assessed both the instrument psychometric properties and the quality of evidence linked to each property with a view to (1) provide 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 (2) establish indications for the future development and validation of such instruments.

Methods

Search Strategy and Eligibility Criteria

A systematic review was conducted following the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines [34]. The search was conducted from June 2023 to July 2023 in the following computerized databases: Scopus, PubMed, and IEEE Xplore. This was done as searching at least 2 databases is recommended for the best coverage of the topic and to decrease chances of inappropriate conclusions [35]. The search strategy aimed to find literature related to the validation and assessment of the psychometric properties of instruments designed to evaluate individuals' psychological dimensions in relation to robots. To accomplish this, the following index terms—"robot*," "social," "home," "domestic," "questionnaire," "survey," "assessment," "measure*," "psychom*," "valid," and "reliab*"—were used. The search strategy is provided in Multimedia Appendix 1.

The search strategy aimed to incorporate instruments suitable for use with social and domestic robots. Therefore, those

specifically designed for other types of robots were excluded. The search strategy targeted studies meeting the following inclusion criteria: (1) the instrument could assess psychological dimensions related to social and domestic robots, including attitudes, beliefs, opinions, feelings, and perceptions; (2) the study focused on validating the instrument; (3) the study evaluated the psychometric properties of the instrument; (4) the study underwent peer review; and (5) the study was in English. Given the focus of our research, studies centered on industrial robots, rescue robots, or robotic arms or those primarily validating technology or measuring anthropomorphism were excluded. In total, 3 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 the established criteria. Those that passed the screening were then evaluated through full-text reading. At the end of each step (title and abstract screening and full-text screening), interrater agreement among the 3 reviewers was evaluated, indicating good agreement (Fleiss κ=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 were 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 (mean, SD, and range). A total of 2 reviewers extracted these 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 the COSMIN guidelines include content validity (which assesses item 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 across all the studies for that given measure. While "sufficient" or "insufficient" clearly indicate whether the criteria were met, the procedure considered studies as "indeterminate" when they addressed the relevant aspect but failed to provide sufficient information to determine whether the criteria had been met

[36,38]. Furthermore, the "inconsistent" category encompassed 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 using the reviewed instruments, considering this study 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 using the identified measures, each study was independently evaluated based on the following suggested generic hypothesis [36]: when the instruments measure related but dissimilar constructs, correlations should fall within the range of 0.30 to 0.50, and when they measure similar constructs, the correlations should be ≥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 "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 3 factors: risk of bias, inconsistency of findings (less relevant for content validity), and imprecision (low sample sizes). It is worth noting that "indirectness" was not evaluated because the review lacked a defined target population.

In total, 2 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

Overview

The search strategy resulted in a total of 3828 articles. After the removal of 14.52% (556/3828) of duplicates, a further 83.49% (3196/3828) of the articles were excluded during title and abstract screening. Then, of the remaining 76 articles, 42 (55%) were excluded during the full-text evaluation for not meeting the research criteria.

The overall search 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 provided in Figure 1. The information extracted for each measure is reported in Table 1, whereas ratings and quality of evidence are reported in Table 2. In the following sections, we discuss the results per instrument, grouping them according to the included population.

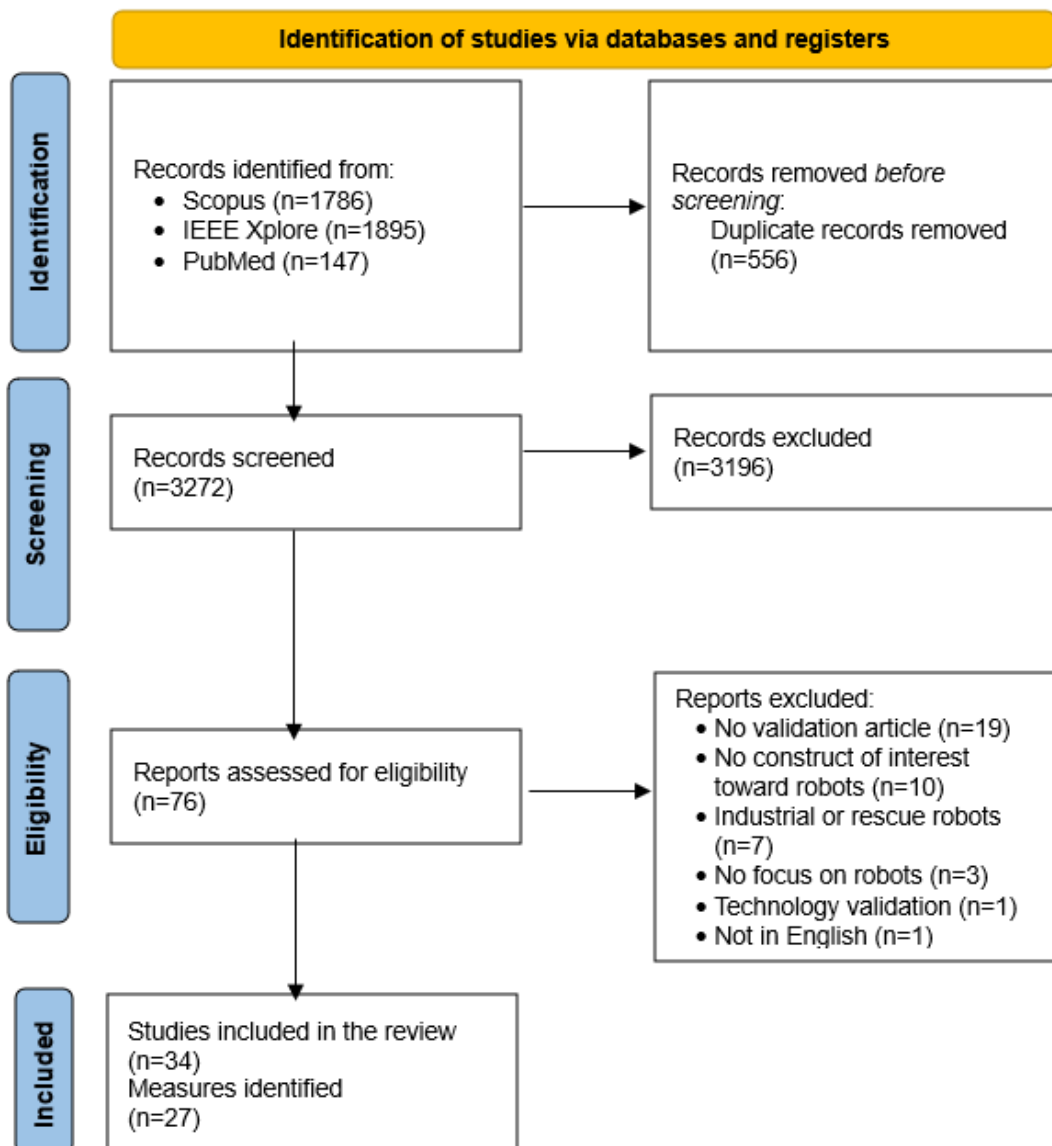**Figure 1.** Flowchart of the search and study selection.

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

Instrument name	Study	Total items	Type of items	Construct or constructs	Subscales (number of items)	Type of administration	Target population	Validation population nationality and number of participants	Age of the validation population (y)
COIRS[a]	Robert and van den Bergh [39]	12	Scale from 0 to 4	Assess openness to robot interaction	Intrinsic interest in interacting with a robot (3); openness to socioemotional interactions with a robot (5); openness to utilitarian interaction with a robot (4)	Self-report	Children	American	Range 8-11
Robot gratification questionnaire[b]	de Jong et al [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 gratification—sought (4); experiential gratification—obtained (4)	Observational	Children	Dutch—24 children (15 male; 9 female)	Mean 9.31 (SD 1.15; range 7-11)
Robot social presence scale[b]	Chen et al [41]	17	5-point Likert scale	Assess robot social presence	Perceived presence (4); interaction behavior 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 (174 male; 320 female)	Range 18-60
REI[c]	Bevilacqua et al [42]	41	5-point Likert scale	Assess the acceptability of SARs[d] in the older adult population	Perceived robot personality (11); human-robot interaction (10); perceived benefit (6); easiness of use (6); perceived usefulness (7)	Self-report	Older adults	Not reported—21 (13 male; 8 female)	Mean 62.9 (SD 3.9)
UNRAQ[e]	Tobis et al [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 (179 male; 541 female)	Mean 52.0 (SD 37.0; range 19-91)
A revision of the TSES-R[f]	Zhang et al [45]	20	5-point Likert scale	Parental expectations regarding robots for health care	Capabilities dimension (5); fictional view dimension (5); social or emotional dimension (4); playful distraction dimension (3); assistive role dimension (3)	Self-report	Parents	Not reported	Not reported

XSL•FO
RenderX

Instrument name	Study	Total items	Type of items	Construct or constructs	Subscales (number of items)	Type of administration	Target population	Validation population nationality and number of participants	Age of the validation population (y)
GAToRS[g]	Koverola et al [46]	20	7-point Likert scale	People's attitudes 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 (192 male; 283 female)	Mean 40.23 (SD 13.51); all participants were aged >18 years
SPRS[h]	Mandl et al [47]	18	Semantic differential scale	Measure social perception of robots	Anthropomorphism (6); morality or sociability (6); activity or cooperation (4)	Self-report	General population	German, Austrian, Swiss, and from other countries—1032 (538 male; 480 female; 14 non-binary)	Mean 33.83 (SD 12.66)
SSRIT[i]	Chi et al [48]	50	5-point Likert scale	Assess trust in interactions with AI[j] 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); effort expectancy (4); perceived service risk (5); robot service fit (3); facilitating robot use condition (3)	Self-report	Consumers	Ethnicities were reported—sample 1: 452 (38.9% male; 60.8% female; 0.2% other); sample 2: 362 (38.7% male; 61% female; 0.3% other)	Range >18->65
ATTUSR-C[k]	Chen et al [49]	15	5-point Likert scale	Assess ATTUSR-C questionnaire for use with Taiwanese health personnel	Unidimensional	Self-report	Health personnel	Taiwanese—95 (95% female)	Mean 44.5 (SD 11.9; range 25-63)
Intentional Acceptance of Social Robots	de Jong et al [50]	4	5-point bar scale	Assess children's intention to use social robots	Unidimensional	Self-report	Children	Dutch—87 (39 male; 48 female)	Mean 9.17 (SD 0.85; range 7-11)
ERAS[l]	Sisman et al [51]	17	5-point Likert scale	Measure the attitudes of school students toward the use of humanoid robots in educational settings	Engagement (5); enjoyment (4); anxiety (4); intention (4)	Self-report	Children	Not reported—232 (128 male; 104 female)	Range 10-13
SE-HRI[m]	der Pütten and Bock [52]	18	6-point Likert scale	Measure people's perceived self-efficacy in dealing with robots	Unidimensional	Self-report	General population	German—450 (288 female; 4 unknown); American—209 (104 male; 105 female)	German: mean 25.15 (SD 6.66; range 18-59); American: mean 26.48 (SD 9.11; range 16-69)

XSL·FO
RenderX

Instrument name	Study	Total items	Type of items	Construct or constructs	Subscales (number of items)	Type of administration	Target population	Validation population nationality and number of participants	Age of the validation population (y)
SE-HRI short version	der Pütten and Bock [52]	10	6-point Likert scale	Measure people's perceived self-efficacy in dealing with robots	Unidimensional	Self-report	General population	English-speaking and German-speaking sample—196 (101 male; 95 female)	Mean 36.91 (SD 13.97; range 18-82)
MCRS[n]	Nomura et al [53]	21	7-point Likert scale	Measure moral concern for robots	Basic moral concern (12); concern for psychological harm (9)	Self-report	General population	Japanese—group 1: 121 (66 male; 55 female); group 2: 200 (100 male; 100 female)	Mean 20.1 (SD 1.6) for group 1; range 20s-60s for group 2
RoSAS[o]	Carpinella et al [54] and Oliveira et al [55]	18 (note: 10 for Portuguese version)	9-point Likert scale	Measure social perception of robots	Warmth (6); competence (6); discomfort (6)	Self-report	General population	Not reported—210 (105 male; 104 female; 1 not identified); Portuguese—185 (45% male; 55% female)	Not reported—Portuguese sample: mean 23.40 (SD 5.21; range 18-35)
HEXACO-60 for HRI[b,p]	Siri et al [56]	60	5-point Likert scale	Evaluate how people perceive the personality traits of robots	Empathy, altruism, or sociability; integrity; dependability; self-confidence	Self-report	General population	Italian—133 (not reported)	Mean 34.46 (SD 14.17; range 19-65)
Sense of safety and security for robots in elder care[b]	Akalin et al [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 (47 male; 53 female)	Mean 35.48 (SD 10.58; range 14-62)
PERNOD[q]	Kamide et al [58]	33	7-point Likert scale	Evaluate humanoid robots	Familiarity (12); utility (7); motion (4); controllability (5); toughness (5)	Self-report	General population	Japanese—380 university students (140 male; 239 female; 1 unknown)	Mean 20.31 (SD 2.89)
Multidimensional Robot Attitude Scale	Ninomiya et al [59]	49	7-point Likert scale	Assess attitudes toward domestic robots	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% male); Taiwanese—130 (46.9% male); Chinese—40.5% male	Japanese: mean 22.3 (SD 1.9); Chinese: mean 23.6 (SD 1.6); Taiwanese: mean 24.2 (SD 5.0)
RUSH[r]	Turja et al [60]	6	5-point Likert scale	Measure robot use self-efficacy in health care work	Unidimensional	Self-report	Health care workers	Finnish—3 samples: 200 home care workers (93.5% female), 1889 nurses (89.8% female), and 1554 nurses and physiotherapists (95% female)	Sample 1: mean 43.2 (SD 11.8; range 19-65); sample 2: mean 45.5 (SD 12.1; range 17-68); sample 3: mean 47.5 (SD 10.4; range 19-70)

Instrument name	Study	Total items	Type of items	Construct or constructs	Subscales (number of items)	Type of administration	Target population	Validation population nationality and number of participants	Age of the validation population (y)
RAS[s]	Nomura et al [61] and Cai et al [62]	11	6-point Likert scale	Measuring the anxiety that prevents individuals from interacting with robots that have 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 (197 male; 199 female; 4 unknown); Chinese: sample 1 composed of 305 adults (138 male; 167 female) and sample 2 composed of 740 adults (319 male; 421 female)	Japanese: mean 21.4; Chinese: sample 1 range 18-≥60 and sample 2 range 18-60
RERS[t]	Nomura and Kanda [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—2s0 university students (not reported)	Not reported
Child-robot relationship formation[b]	Straten et al [64]	13	5-point bar scales	Assess child-robot relationship formation	Closeness (5); trust (4); perceived social support (5)	Self-reported	Children in middle childhood	Dutch—87 children (39 male; 48 female)	Mean 9.17 (SD 0.85; range 7-11)
Almere model	Heerink et al [21] and He et al [65]	41 (30 for the Mandarin version)	5-point Likert scale	Acceptance of assistive social agents by older adults	Anxiety (4); attitude toward 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 (18 male; 22 female), experiment 2: 88 participants (28 male; 60 female), experiment 3: 30 older adults (8 male; 22 female), and experiment 4: 30 older adults (16 male; 14 female); Chinese—317 (55.5% female)	Dutch—experiment 1: range 65-89, experiment 2: NA[u], experiment 3: range 65-94, and experiment 4: range 65-89; Chinese—mean 70.3 (SD 7.5)
Frankenstein Syndrome Questionnaire	Nomura et al [66,67]	30	7-point Likert scale	Measure acceptance of humanoid robots, including expectations and anxieties regarding this technology in the public	General anxiety toward humanoid 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)	Self-reported	General population	Japanese (the questionnaire is also available in English)—1000 (500 male; 500 female)	Range 20s-60s

Instrument name	Study	Total items	Type of items	Construct or constructs	Subscales (number of items)	Type of administration	Target population	Validation population nationality and number of participants	Age of the validation population (y)
NARS[v]	Nomura et al [68], Piçarra et al [69], Pochwatko et al [70], Syrdal et al [71], and Xia and Le-Tendre [72]	14 (the Portuguese and Polish versions have 12 items, and the English version has 11 items)	5-point Likert scale (7-point Likert scale for the Portuguese and Polish versions)	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: the Portuguese and Polish versions have 2 dimensions, NARHT[w] and NATIR[x], and the English version has 3 dimensions measuring different constructs)	Self-reported	General population	Japanese—240 university students (146 male; 92 female; 2 unknown); American—54 undergraduate students (13 male; 41 female); Portuguese—4 studies with a total sample of 997 (401 male; 598 female; 3 not reported); Polish—213 (80 male; 91 female; 42 not reported); English—28 university students and staff (14 male; 14 female)	Japanese—mean 22.0 (SD not reported); American—range 18-25; Portuguese—range 18-71; Polish—mean 29.36 (SD 10.15); English—range 18-55

[a]COIRS: Children's Openness to Interacting With a Robot Scale.

[b]The name of the instrument was not provided in the original article.

[c]REI: Robot-Era Inventory.

[d]SAR: socially assistive robot.

[e]UNRAQ: Users' Needs, Requirements, and Abilities Questionnaire.

[f]TSES-R: Technology-Specific Expectation Scale–R.

[g]GAToRS: General Attitudes Toward Robots Scale.

[h]SPRS: Social Perception of Robots Scale.

[i]SSRIT: Social Service Robot Interaction Trust.

[j]AI: artificial intelligence.

[k]ATTUSR-C: Chinese version of the Attitudes Toward The Use of Social Robots.

[l]ERAS: Educational Robot Attitude Scale.

[m]SE-HRI: Self-Efficacy in Human-Robot Interaction Scale.

[n]MCRS: Moral Concern for Robots Scale.

[o]RoSAS: Robotic Social Attributes Scale.

[p]HRI: human-robot interaction.

[q]PERNOD: Perception to Humanoid scale.

[r]RUSH: Robot Use Self-Efficacy in Healthcare Work.

[s]RAS: Robot Anxiety Scale.

[t]RERS: Rapport-Expectation With a Robot Scale.

[u]Not available.

[v]NARS: Negative Attitudes Toward Robots Scale.

[w]NARHT: negative attitudes toward robots with human traits.

[x]NATIR: negative attitudes toward interactions with robots.

Table 2. Results obtained from the overall rating (OR) quality of evidence (QoE)[a].

Instrument	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[b]	±[c]	Low	_[d]	Low	+[e]	Low	+	Low	+	High	+	High	__[f]	—	—	—	—	—	?[g]	Low
Robot gratification questionnaire[h]	±	Low	+	Low	−	Low	−	Low	?	Low	±	Low	—	—	—	—	—	—	±	Very low
Robot social presence scale[h]	+	Moderate	+	Moderate	+	Moderate	+	Moderate	+	Moderate	+	High	—	—	—	—	—	—	—	—
REI[i]	−	Very low	±	Low	−	Very low	−	Very low	?	Very low	−	Very low	—	—	—	—	—	—	?	Very low
UNRAQ[j]	−	Very low	−	Very low	−	Very low	−	Very low	—	—	±	High	—	—	—	—	+	Moderate	—	—
TSES-R[k]	−	Very low	−	Very low	−	Very low	−	Very low	?	Very low	+	Moderate	—	—	—	—	—	—	—	—
GAToRS[l]	±	Very low	+	Low	−	Very low	−	Very low	+	High	+	High	—	—	—	—	—	—	±	High
SPRS[m]	−	Very low	−	Low	−	Very low	−	Very low	−	High	±	High	—	—	—	—	—	—	±	High
SSRIT[n]	+	Moderate	+	Moderate	+	Moderate	+	Low	+	High	+	High	—	—	—	—	—	—	+	High
ATTUSR-C[o]	±	Very low	+	Very low	−	Very low	+	Very low	?	Moderate	+	Moderate	—	—	—	—	—	—	—	—
Intentional Acceptance of Social Robots	−	Very low	−	Very low	−	Very low	−	Low	+	Moderate	+	Moderate	—	—	+	Low	—	—	±	Very low
ERAS[p]	−	Moderate	−	Moderate	−	Moderate	+	Moderate	?	High	+	High	—	—	—	—	—	—	—	—
SE-HRI[q]	±	Very low	+	Very low	−	Very low	+	Very low	+	High	+	High	—	—	—	—	—	—	+	High
SE-HRI (short version)	—	—	—	—	—	—	—	—	—	—	+	High	—	—	—	—	—	—	+	High
MCRS[r]	−	Very low	−	Very low	−	Very low	−	Very low	?	Moderate	+	High	—	—	—	—	—	—	?	Very low
RoSAS[s]	−	Very low	−	Very low	−	Very low	+	Low	+	High	+	High	—	—	—	—	±	Moderate	±	Moderate
HEXACO-60 for HRI[d,t]	−	Very low	±	Very low	−	Very low	−	Very low	?	Very low	—	—	—	—	—	—	—	—	—	—
Sense of safety and security for robots in elder care[h]	−	Very low	−	Very low	−	Very low	−	Very low	?	Very low	+	Moderate	—	—	—	—	—	—	—	—
PERNOD[u]	−	Very low	−	Low	−	Very low	−	Very low	?	High	+	High	—	—	—	—	—	—	—	—

XSL•FO
RenderX

Instrument	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
Multidimensional Robot Attitude Scale	–	Very low	–	Very low	–	Low	–	Very low	—	—	±	High	—	—	—	—	—	—	—	—
RUSH[v]	–	Very low	–	Very low	–	Very low	–	Very low	?	High	+	High	—	—	—	—	—	—	–	Very low
RAS[w]	+	Moderate	+	Moderate	+	Moderate	?	Moderate	+	High	+	High	—	—	—	—	—	—	±	High
RERS[x]	–	Very low	±	Very low	–	Very low	–	Very low	?	Very low	+	Moderate	—	—	—	—	—	—	?	Low
Child-robot relationship formation[h]	±	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[y]	–	Low	–	Low	–	Low	–	Very low	–	High	+	High	–	Very low	—	—	±	Low	±	High

[a]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 *Methods* section.

[b]COIRS: Children's Openness to Interacting With a Robot Scale.

[c]Inconsistent rating.

[d]Insufficient rating.

[e]Sufficient rating.

[f]Blank cells represent psychometric properties that were not evaluated for that instrument.

[g]Indeterminate rating.

[h]The name of the instrument was not reported in the original article.

[i]REI: Robot-Era Inventory.

[j]UNRAQ: Users' Needs, Requirements, and Abilities Questionnaire.

[k]TSES-R: Technology-Specific Expectation Scale–R.

[l]GAToRS: General Attitudes Toward Robots Scale.

[m]SPRS: Social Perception of Robots Scale.

[n]SSRIT: Social Service Robot Interaction Trust.

[o]ATTUSR-C: Chinese version of the Attitudes Toward the Use of Social Robots.

[p]ERAS: Educational Robot Attitude Scale.

[q]SE-HRI: Self-Efficacy in Human-Robot Interaction Scale.

[r]MCRS: Moral Concern for Robots Scale.

[s]RoSAS: Robotic Social Attributes Scale.

[t]HRI: human-robot interaction.

[u]PERNOD: Perception to Humanoid scale.

[v]RUSH: Robot Use Self-Efficacy in Healthcare Work.

[w]RAS: Robot Anxiety Scale.

[x]RERS: Rapport-Expectation With a Robot Scale.

[y]NARS: Negative Attitudes Toward Robots Scale.

Instruments to Assess Children's Psychological Dimensions Toward Robots

The *Children's Openness to Interacting With a Robot Scale* [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 pretesting with colleagues and researchers. During the validation, exploratory factor analysis (EFA) revealed a 3D structure with good internal consistency (Cronbach α ranging from 0.72 to 0.78) and sufficient structural validity (root mean square error of approximation [RMSEA]=0.07; comparative fit index [CFI]=0.93; root mean square residual=0.07) for the 3 dimensions. Construct validity was assessed by correlating the average Children's Openness to Interacting With a Robot Scale score with those of 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, although the purpose of the comparison was not reported.

The questionnaire developed by de Jong et al [40] aimed to assess children's uses of and gratifications regarding 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 4 types of gratification. The items were derived from previous questionnaires and children's answers. The gratification types were subsequently categorized into sought and obtained, although the theoretical rationale for this choice was not provided. During the validation, 2 subscales did not reach sufficient internal consistency. The EFA results did not provide information about the goodness of the 8D solution. Some of the subscales did not provide sufficient evidence for the expected hypothesis tested by the authors (Pearson 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 or for a long period in their daily life. The researchers reviewed existing measures and focused on the scale by Heerink et al [21]. 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 into Dutch. Pilot-testing with 4 children led to further adjustments. The confirmatory factor analysis (CFA) revealed a good fit of the data (N=87, χ^2_2=3.6, *P*=.16; CFI=0.97; standardized root mean square residual=0.04). Measurement invariance was assessed between boys and girls, showing sufficient results. Internal consistency showed sufficient results for the overall sample (range 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 (ie, social presence [*r*=0.24] and social anxiety [*r*=−0.20]).

The *Educational Robot Attitude Scale* [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 20 schoolchildren was conducted to assess item comprehension. The scale showed a 4D solution according to EFA (fitting indexes were not reported). The reliability of the scale was satisfactory (Cronbach α ranged from 0.81 to 0.85).

Straten et al [64] developed a measure to assess child-robot relationships using 3 self-report scales of closeness, trust, and perceived social support in which constructs were 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 (N=87, χ^2_{62}=62.3, *P*=.47; CFI=0.999; standardized root mean square error [SRMR]=0.052). Hypothesis testing with concurrently measured variables, which were significantly shortened, yielded mixed results.

Instruments to Assess Psychological Dimensions of Adults (Aged ≥18 Years) Toward Robots

Chen et al [41] proposed a 6D 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. A total of 3 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 (chi-square–to–df ratio=2.160; RMSEA=0.048; Tucker-Lewis index [TLI]=0.928; normed fit index=0.939; Adjusted Goodness of Fit Index [AGFI]=0.926; SRMR=0.052; CFI=0.966; goodness-of-fit index [GFI]=0.950), and the Cronbach α values were of >0.70.

The *Users' Needs, Requirements, and Abilities Questionnaire* [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 on the use of social robots in the care of older people. The validation sample consisted of 720 older adult participants, 125 of whom repeated the assessment 2 weeks apart. Evaluation of psychometric properties indicated good Cronbach α values for each dimension (all >0.70) and test-retest reliability for each subscale measured using the intraclass correlation coefficient (range 0.81-0.93).

The *General Attitudes Toward Robots Scale* [46] was developed to assess attitude as a predisposition to respond favorably or unfavorably to objects in the world and makes a distinction between personal and societal levels of attitudes toward robots, differentiating them between positive and negative. In the pilot study, the authors only reported that the measure was developed

partly based on other instruments. The 4D factor was considered adequate as it fell between the suggested factors of different evaluation methods. Only 2 of the dimensions had a Cronbach α value of >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 forums 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 (χ^2_{164}=430.0, P<.001; CFI=0.91; TLI=0.896; RMSEA=0.058, 90% CI 0.052-0.064; SRMR=0.057), and the Cronbach α values for each subscale were of >0.70. Correlations with the Negative Attitudes Toward Robots Scale (NARS) indicated mixed results (range 0.2-0.8); however, the authors did not report a specific hypothesis and, given that the General Attitudes Toward Robots Scale measures attitude toward robots, we would have expected strong correlations of >0.50.

The *Social Perception of Robots Scale* [47] was developed as a short scale for measuring social perceptions of robots that comprises sociability, competence, morality, and anthropomorphism that can be applied to different robots in diverse research settings. Although a definition for each scale was provided, the authors did not describe a theoretical background for the social perception dimension and for its subcomponents. The authors composed items based on 3 different instruments to address the 3 main dimensions of social perception. The EFA results indicated a 3D factor (anthropomorphism, morality or sociability, and activity or cooperation), and a subsequent CFA did not indicate good fit (χ^2_{115}=508.1, P<.001; RMSEA=0.101; CFI=0.796; TLI=0.759; SRMR=0.096). Regarding internal consistency, the third dimension resulted in a low index (Cronbach α=0.64), whereas the first and second dimensions had sufficient indexes (Cronbach α=0.82 and 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* [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 Self-Efficacy in Human-Robot Interaction Scale consisted of items that were either adapted from different questionnaires or theoretically generated. An EFA indicated a 2D solution (namely, self-efficacy and loss of control), which showed good internal consistency (Cronbach α=0.945 and 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 (chi-square–to–df ratio of 5.21 and poor values for the other fit indexes: RMSEA=0.097; CFI=0.84; SRMR=0.055), and a subsequent analysis with reduced items indicated a 1-factor solution and a good model fit (chi-square–to–df ratio=2.98; RMSEA=0.066; CFI=0.95; SRMR=0.029). This result was replicated for the English version. The comprehensibility of the German version was assessed with 6 older adults. Hypothesis testing performed with correlations indicated sufficient values (r>0.30); however, we

should note that, with only 1 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 on results from the EFA. A CFA indicated good fit of the short version for both the German and English samples. In addition, hypothesis testing indicated correlations of >0.30; however, this was true also for 1 scale that the authors used as a discriminant measure.

Nomura et al [53] developed the *Moral Concern for Robots Scale*. A definition or a theoretical background of moral concern was not clearly provided. The Moral Concern for Robots Scale was obtained by adopting items from existing questionnaires. In addition, they created items based on human moral treatment and scenes of possible robot abuse. Through a questionnaire-based survey, the collected data were analyzed using factor analysis, resulting in a 2-factor structure. No fitting statistics were reported. Each dimension indicated good internal consistency (Cronbach α=0.912 and 0.876). Most of the correlations conducted by the authors for construct validity were of <0.30, and 2 dimensions indicated high correlations with the developed measure, namely, "Mental state" and "Social partner"; however, they were not assessed using validated measures, so these results could not properly be considered as evidence of construct validity.

The *HEXACO-60 for Robots* [56] is based on the HEXACO model of personality and proposes that individuals are characterized by 6 domains. The authors adapted the items of the HEXACO-60 original questionnaire [73] addressing "a robot" as the subject of each original item. Even though the construct was clearly described and had a theoretical background, a representative population was not involved in the elicitation of relevant items; thus, relevance was considered indeterminate. The authors performed an EFA that indicated a 4D solution; fitting statistics were not reported.

Akalin et al [57] developed a scale to evaluate the sense of safety and security of robots for older adult care. The authors developed the items after videos of different types of robot interactions were shown to participants; 3 items were based on the Godspeed Questionnaire Series [74]. Definitions of safety and security used to construct items were not clearly reported. The authors calculated the Cronbach α for the 2 dimensions and for each video scenario presented to the participants. All the Cronbach α values reported were of >0.70, indicating good internal consistency of the scales. Factor analysis was performed to identify the most important item associated with the 2 dimensions.

The *Robotic Social Attributes Scale* (RoSAS) developed by Carpinella et al [54] assesses warmth, competence, and discomfort perceived in robots. While the first 2 dimensions were drawn from social psychology, they lacked a clear definition, making it challenging to assess the content of items related to these dimensions. The development of this scale involved 4 studies. In the first study, an EFA was conducted on the Goodspeed Questionnaire Series [74], resulting in 3 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 the number of items and suggested 3 dimensions: warmth (Cronbach α=0.91), competence (Cronbach α=0.84), and discomfort (Cronbach α=0.82). The third study trialed 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 whether participants' perceptions varied based on the scale. However, references to support the hypotheses were not provided. In a separate study, Oliveira et al [55] performed a Portuguese translation of the RoSAS and assessed the comprehension of its items. A CFA suggested that the 3 dimensions were a good solution (CFI=0.98; RMSEA=0.05; SRMR=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* [63] was designed to measure people's expectations regarding 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 previous research. Subsequently, an EFA was conducted with a small sample, revealing a 2D solution (Cronbach α=0.919 and 0.848). Unfortunately, no fit indexes 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 was 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 1 of the 2 hypotheses was 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 3D factor solution. Following this, a CFA indicated a good fit (GFI=0.949; AGFI=0.917; RMSEA=0.066) for each scale (Cronbach α=0.840, 0.844, and 0.796). Construct validity was evaluated by comparing the RAS with 2 other anxiety measures, all showing correlations of <0.30. Cai et al [62] 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 square–to–df ratio=3.26; SRMR=0.02; CFI=0.99; GFI=0.96; TLI=0.98; RMSEA=0.06), and correlations for construct validity indicated good construct validity (absolute values of *r* ranged from 0.42 to 0.81). Overall, the correlations between the 2 studies yielded mixed results.

The *Frankenstein Syndrome Questionnaire* (FSQ) developed by Nomura et al [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 regarding humanoid robots from students in both Japan and the United Kingdom. A group of experts later reviewed the extracted items for content validity. The questionnaire was administered on the web, and a factor analysis revealed a 4D solution (Cronbach α range 0.693-0.909). GFIs 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 et al [68] developed the NARS to assess the predispositions in behavior or reactions toward robots. They initially gathered opinions through a pilot survey extracting 13 sentences and obtained an additional 20 sentences from 2 other measures. The content validity was confirmed through expert discussions. During the validation, an EFA revealed a 4-factor structure, and a CFA indicated a good fit (GFI=0.900; AGFI=0.856; RMSEA=0.080), with Cronbach α coefficients ranging from 0.648 to 0.782. Construct validity was assessed using Pearson correlation with a measure of anxiety, but all coefficients were of <0.30. Test-retest reliability, assessed using Pearson correlation, showed mixed results: 2 subscales had good reliability (*r*=0.706 and *r*=0.740), but the "Negative attitudes toward emotions in interaction with robots" subscale did not (*r*=0.538). Syrdal et al [71] assessed the NARS in the English population after translating it. They removed 3 items, although the Cronbach α was not reported for each subscale. They conducted a principal component analysis to assess item loadings on each dimension. Construct validity was assessed using 12 personality traits [77], yielding mixed results. A Portuguese validation of the measure was conducted by Piçarra et al [69], resulting in a 2D solution and a good model fit (CFI=0.93; TLI=0.90; RMSEA=0.065). Each subscale showed good internal consistency (Cronbach α=0.73 and 0.75), although construct validity was not evaluated using other standardized measures. The Polish version of the measure, as conducted by Pochwatko et al [70], resulted in a 2D solution with 2 items removed. Both subscales exhibited good internal consistency (Cronbach α=0.84 and 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 3-factor structure (CFI=0.93; TLI=0.91; RMSEA=0.08; SRMR=0.08), and internal consistency was also confirmed (Cronbach α ranged from 0.773 to 0.818). The study revealed differences between the 2 groups of students. It is important to note that the structural validity of the NARS yielded conflicting results, with some studies suggesting a 3D solution whereas others proposed a 2D solution.

The *Perception to Humanoid* scale developed by Kamide et al [58] was designed to assess people's perspectives when evaluating humanoid robots. University students were required to describe their impressions after viewing a video recording of a humanoid robot. The responses were categorized into groups and adapted into items. An EFA indicated a 5D solution, with each dimension demonstrating good internal reliability (Cronbach α ranging from 0.79 to 0.86). However, no CFA or GFIs were reported.

To assess attitudes toward domestic robots, Ninomiya et al [59] developed the *Multidimensional 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 12D structure. Subsequently, 2 to 7 items were selected for each factor based on their loadings with the aim of ensuring sufficient differentiation among them. The Cronbach α values for most dimensions exceeded 0.70 except for the value for the "control" dimension, which was 0.643.

The *Social Service Robot Interaction Trust* [48] assesses consumers' trust in interactions with artificial intelligence 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 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 Social Service Robot Interaction Trust with the Interpersonal Trust Scale [78] and the Technology Artifact Scale [79], revealing high correlations (*r*=0.78 and *r*=0.84, respectively).

Instruments to Assess Older Adults' Psychological Dimensions Toward Robots

The *Robot-Era Inventory* [42] was designed to measure older adults' acceptance of social robots across 5 dimensions based on the Robot-Era Model proposed by the authors. 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 2 of the proposed subscales showing insufficient Cronbach α values (0.67 and 0.69). Construct validity was assessed by examining the correlations between the Robot-Era Inventory and the Unified Theory of Acceptance and Use of Technology, although a clear hypothesis was not reported. The associations between the dimensions showed mixed results.

The *Almere model* was developed to test the acceptance of assistive social agents by older adults [21]. The questionnaire items were adapted from the UTAUT to fit the context of assistive robot and screen agent technology, specifically addressing older adult users in a care home environment. Additional constructs were considered, and items were adapted from questionnaires in the literature. A path analysis was used to test hypothesized relationships among the dimensions. In an experiment comparing responses to a robot in more social versus less social conditions, differences were found on 4 subscales. The study showed sufficient internal consistency of the instrument. He et al [65] translated the Almere Technology Acceptance Questionnaire into Mandarin Chinese and evaluated its psychometric properties among older adults in China. They performed a content analysis with 6 experts. EFA followed by CFA revealed a 9D solution (chi square–to–df ratio=2.006; RMSEA=0.069; root mean-square residual=0.059; GFI=0.816; incremental fit index=0.913; TLI=0.896; CFI=0.912). Cronbach α 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 Health Care Professionals Toward Robots

The *Robot Use Self-Efficacy in Healthcare Work* [60] is a measure developed and validated with health care workers to assess their self-efficacy in using robots in their work. There was no reported information regarding the items' development. The validation of the measure indicated sufficient internal consistency (Cronbach α=0.90); only factor loadings of the factor analysis performed were reported. Regarding correlations performed for construct validity, results obtained for the 6-item version of the measure were not reported. Instead, the authors performed a correlation analysis for the short version of the measure (3-item version). Most correlations with other measures were insufficient (*r*<0.30) except for one (*r*=0.33), which was measured with only 1 item, and its validation was not reported.

The Chinese version of the *Attitudes Toward the Use of Social Robots* (ATTUSR-C) [49] questionnaire is a modified and translated version of the questionnaire proposed by Costescu and David [80]. Although the original version provides a clear definition of "attitude," the study was not aimed at validating the questionnaire, and there was no evidence of concept elicitation or literature search in item generation. In this version, a panel of 5 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 of <70% were eliminated. In addition, 10 clinical instructors assessed the instrument for face validity by evaluating the clarity of each questionnaire item. This process indicated sufficient evidence for relevance and comprehensibility of the items; however, the professionals were not asked about the comprehensiveness of the items, and thus, it was evaluated as insufficient. During validation, the EFA interpretation led to a 1D solution, with no reported fit indexes. The Cronbach α was sufficient (0.84).

Among the instruments designed for health care, we can highlight the *Technology-Specific Expectation Scale–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 [81], and additional items were created by the authors, organized into 3 dimensions. Principal component analysis was used to determine the item loadings for each dimension; however, fit indexes were not reported. Each subscale demonstrated good internal consistency (Cronbach α=0.869, 0.839, and 0.800). Details regarding the sample used for this analysis were not provided.

Discussion

Principal Findings

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 toward these emerging technological tools. This review aimed 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. This 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 using 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 related to robots. Most of the studies (24/27, 89%) primarily concentrated on assessing the structural validity (12/24, 50% of evidence) and internal consistency (26/27, 96%, of which 18/26, 69% had a high quality of evidence) of the instruments. Construct validity was considered for 63% (17/27) of the instruments (7/17, 41% with a high quality of evidence). Cross-cultural validity was evaluated for only 7% (2/27) of the instruments (both of which exhibited low quality of evidence), and measurement invariance was evaluated for only 4% (1/27) of the instruments (low quality of evidence). Notably, the measurement error and responsiveness aspects were disregarded across all the instruments. Content validity was identified in most of the studies (26/27, 96%); however, none of them exhibited an overall high quality of evidence. Moreover, there was a noticeable scarcity of tools specifically tailored for children, older adults, and health care 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.

Regarding content validity, many studies (17/26, 65%) 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, the importance of considering this aspect should be further stressed. Content validity refers to how the content of the scale adequately reflects the construct the instrument is intended to measure [82], 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 [83]. Therefore, careful consideration of the construct's definition and its theoretical basis should be taken during the development of the instrument to enhance methodological rigor and improve the quality of the assessment.

Measurement error and responsiveness were not 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 [82]. It could be assessed through minimally important change, which indicates whether a change in the measurement is considered important [84], or the smallest detectable change, which indicates whether the change in score is of sufficient magnitude to have a low probability of being a random error [85]. Responsiveness indicates how the instrument could detect change over time in the measured construct [86], which is considered to reflect longitudinal validity [87]. However, even if they are important, we should also note that these 2 properties place a strong emphasis in the clinical context [87,88]. Thus, we suggest considering these 2 properties with caution and within the context and aim for which the instrument is used.

Most studies (24/27, 89%) primarily focused on structural validity, typically through EFA or CFA, as well as internal consistency as measured using the Cronbach α. However, it is important to note that several studies (12/24, 50%) 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 [89] and 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 used in later phases on empirical and theoretical grounds [90].

Regarding cross-cultural validity, only 7% (2/27) of the instruments, the NARS [72] and the FSQ [66], assessed this aspect. However, in both cases, there was insufficient evidence to demonstrate cross-cultural validity, and the quality of the methodology was low. Despite the challenges associated with considering this property, cross-cultural validity offers valuable instruments for diverse cultures [91]. Its importance is evident from the multinational studies conducted on the topic of human-robot interactions [92,93] and is also highlighted by the diverse nationalities of the validation samples included in this review. This aspect should be given further consideration in the context of psychological measurements for human-robot relations, with particular attention to the methods used.

Similarly, measurement invariance was examined in only 3% (1/34) of the studies [50]. This is particularly concerning as it would indicate that differences between groups evaluated using most of these instruments could be due to group-specific characteristics rather than to true differences in the dimensions assessed by them [94]. 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 [95], was largely overlooked. Only 15% (4/27) of the instruments provided evidence for the assessment of this property: the Users' Needs,

Requirements, and Abilities Questionnaire; the RoSAS; the Almere model; and the NARS [44,55,65,68]. Consequently, many of the identified instruments did not demonstrate reliability, which represents a significant limitation of most available instruments.

When assessing construct validity, a significant proportion of studies (14/17, 82%) used correlations with other instruments. Nevertheless, these correlations often yielded inconsistent results. A problem faced in this evaluation was that many studies (7/14, 50%) 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 (17/27, 63%) reviewed had 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, and thus, it could not be established whether certain items could favor individuals from different age groups with different backgrounds or due to specific response formats [96]. In this regard, this aspect should be considered when developing these instruments. Only 7% (2/27) of the instruments, the Robot-Era Inventory [42] and the Almere model [21], were designed specifically for the older adult population. However, they exhibited limited psychometric properties, indicating the need to develop instruments for this specific demographic group.

Only 11% (3/27) of the instruments considered the clinical context. The Robot Use Self-Efficacy in Healthcare Work [60] and ATTUSR-C [49] were designed for health care professionals, whereas the Technology-Specific Expectation Scale–R focused on parents' expectations [45]. However, these 3 instruments only demonstrated sufficient internal consistency, indicating that there is still a need to develop psychometrically valid and reliable instruments in the health care context. This is particularly important given the literature's emphasis on the use of social robots in health care settings [97].

This review indicated that only 19% (5/27) of the instruments in the literature were validated for children. Most of them (3/5, 60%) only demonstrated sufficient structural validity and internal consistency, suggesting that the available measures to assess the 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 most of the reviewed studies (26/27, 96%) did not effectively use item response theory. While there is some debate regarding the best approach [98], given the conditions of the validation study [99], authors should also take this framework into consideration.

Despite this 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, whereas other important ones tend to be overlooked. Psychometric properties indicate whether the instrument used 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 [100]. Limitations in these properties may raise concerns regarding the accuracy of reported outcomes in research and in making informed decisions [101].

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 use 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 there is preliminary evidence suggesting that the context in which the robot is presented or used 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 populations.

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 paid to content validity, cross-cultural validity, reliability, measurement invariance, and construct validity through rigorous methodologies. Particular attention should be paid 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 this study yielded intriguing results, it is important to acknowledge its limitations. The review and analysis in this study primarily focused 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 into those domains as well. In addition, there are alternative measures in the existing literature for assessing psychological constructs in the context of human-robot interaction. In this regard, it is worth mentioning the Godspeed Questionnaire Series [74]. The search strategy in this review focused 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 is worth emphasizing that the findings of this study may contribute to the identification of gold standards for other instruments in the future. Indeed, there is significant variability in their use, with some scales being rarely used whereas others are more commonly used (eg, 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 was to offer a comprehensive overview of the instruments available to measure various dimensions and 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, such as a robot, should take is particularly challenging [102].

human-robot relationships applicable to both social 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 instruments and the review of existing ones. This 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 development.

Conclusions

Numerous psychometrically validated instruments exist for assessing various psychological constructs within the realm of

Data Availability

Data sharing is not applicable to this paper as no data sets were generated or analyzed during this study.

Authors' Contributions

RV, MZ, ADN, and DM contributed to conceptualization. Methodology was designed by RV, SM, and DM. Formal analysis was conducted by RV, NC, and DM. Investigation was carried out by RV, NC, MS, KAB, and DM. Resources were provided by MZ, ADN, and DM. Data curation was managed by RV and NC. The original draft of the manuscript was written by RV, NC, MS, and KAB. RV contributed to visualization. Supervision was provided by SM, MZ, ADN, and DM. Project administration was handled by SM and DM. Funding acquisition was secured by SM, MZ, ADN, and DM. All the authors contributed to writing—review and editing.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Search strategy for each database.
[DOCX File , 13 KB-Multimedia Appendix 1]

Multimedia Appendix 2

PRISMA checklist.
[PDF File (Adobe PDF File), 126 KB-Multimedia Appendix 2]

References

1. Mejia C, Kajikawa Y. Bibliometric analysis of social robotics research: identifying research trends and knowledgebase. Appl Sci. Dec 18, 2017;7(12):1316. [doi: 10.3390/app7121316]
2. Belpaeme T, Kennedy J, Ramachandran A, Scassellati B, Tanaka F. Social robots for education: a review. Sci Robot. Aug 15, 2018;3(21):eaat5954. [doi: 10.1126/scirobotics.aat5954] [Medline: 33141719]
3. Scoglio AA, Reilly ED, Gorman JA, Drebing CE. Use of social robots in mental health and well-being research: systematic review. J Med Internet Res. Jul 24, 2019;21(7):e13322. [FREE Full text] [doi: 10.2196/13322] [Medline: 31342908]
4. Camp N, Lewis M, Hunter K, Johnston J, Zecca M, Di Nuovo A, et al. Technology used to recognize activities of daily living in community-dwelling older adults. Int J Environ Res Public Health. Dec 28, 2020;18(1):163. [FREE Full text] [doi: 10.3390/ijerph18010163] [Medline: 33379319]

5.   Di Nuovo A, Broz F, Wang N, Belpaeme T, Cangelosi A, Jones R, et al. The multi-modal interface of Robot-Era multi-robot services tailored for the elderly. Intel Serv Robotics. Sep 2, 2017;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 Soc Robotics. Jul 3, 2014;6(4):575-591. [doi: 10.1007/s12369-014-0242-2]

7.   Di Nuovo A, Conti D, Trubia G, Buono S, Di Nuovo SF. Brief review of robotics in low-functioning autism therapy. In: Proceedings of the Second Symposium on Psychology-Based Technologies. 2020. Presented at: Second Symposium on Psychology-Based Technologies; September 28-29, 2020; Naples, Italy.

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. Jun 2022;78:101633. [FREE Full text] [doi: 10.1016/j.arr.2022.101633] [Medline: 35462001]

9.   Camp N, Di Nuovo A, Hunter K, Johnston J, Zecca M, Lewis M, et al. Perceptions of socially assistive robots among community-dwelling older adults. In: Proceedings of the 14th International Conference on Social Robotics. 2022. Presented at: ICSR 2022; December 13-16, 2022; Florence, Italy. [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 Soc Robotics. Nov 20, 2019;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, et al. Collaborative and inclusive process with the autism community: a case study in Colombia about social robot design. Int J Soc Robotics. Feb 07, 2020;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, et al. Robot companions and sensors for better living: defining needs to empower low socio-economic older adults at home. In: Proceedings of the 15th International Conference on Social Robotics. 2023. Presented at: ICSR 2023; December 3-7, 2023; Doha, Qatar. [doi: 10.1007/978-981-99-8715-3_31]

13.   Winkle K, Senft E, Lemaignan S. LEADOR: a method for end-to-end participatory design of autonomous social robots. Front Robot AI. 2021;8:704119. [FREE Full text] [doi: 10.3389/frobt.2021.704119] [Medline: 34926589]

14.   Bartneck C, Nomura T, Kanda T, Suzuki T, Kennsuke K. A cross-cultural study on attitudes towards robots. In: Proceedings of the International Conference on Human-Computer Interaction. 2005. Presented at: HCI 2005; June 20-23, 2005; Las Vegas, NV. [doi: 10.13140/RG.2.2.35929.11367]

15.   Cavallo F, Esposito R, Limosani R, Manzi A, Bevilacqua R, Felici E, et al. Robotic services acceptance in smart environments with older adults: user satisfaction and acceptability study. J Med Internet Res. Sep 21, 2018;20(9):e264. [FREE Full text] [doi: 10.2196/jmir.9460] [Medline: 30249588]

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 Soc Robotics. Jun 1, 2016;9(1):51-62. [doi: 10.1007/s12369-016-0359-6]

17.   de Graaf MM, Ben Allouch S, van Dijk JA. Why would I use this in my home? A model of domestic social robot acceptance. Hum Comput Interact. Jul 21, 2017;34(2):115-173. [doi: 10.1080/07370024.2017.1312406]

18.   Caruana N, Moffat R, Miguel-Blanco A, Cross ES. Perceptions of intelligence and sentience shape children's interactions with robot reading companions. Sci Rep. May 05, 2023;13(1):7341. [FREE Full text] [doi: 10.1038/s41598-023-32104-7] [Medline: 37147422]

19.   de Graaf MM, Ben Allouch S. Exploring influencing variables for the acceptance of social robots. Robot Auton Syst. Dec 2013;61(12):1476-1486. [doi: 10.1016/j.robot.2013.07.007]

20.   Shin DH, Choo H. Modeling the acceptance of socially interactive robotics. Interact Stud. Dec 6, 2011;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 Soc Robotics. Sep 4, 2010;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 Soc Robotics. Jul 06, 2023. [doi: 10.1007/s12369-023-01024-x]

23.   Wilson RK, Eckel CC. Judging a book by its cover: beauty and expectations in the trust game. Polit Res Q. Jul 02, 2016;59(2):189-202. [doi: 10.1177/106591290605900202]

24.   Chen YC, Yeh SL, Lin W, Yueh HP, Fu LC. The effects of social presence and familiarity on children-robot interactions. Sensors (Basel). Apr 24, 2023;23(9):4231. [FREE Full text] [doi: 10.3390/s23094231] [Medline: 37177434]

25.   Ziemke T. Understanding social robots: attribution of intentional agency to artificial and biological bodies. Artif Life. Aug 01, 2023;29(3):351-366. [FREE Full text] [doi: 10.1162/artl_a_00404] [Medline: 36943757]

26.   Dou X, Wu CF, Wang X, Niu J. User expectations of social robots in different applications: an online user study. In: Proceedings of the HCI International 2020 - Late Breaking Papers: Multimodality and Intelligence. 2020. Presented at: HCII 2020; July 19-24, 2020; Copenhagen, Denmark. [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. Comput Hum Behav. Dec 2022;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. Oct 21, 2019;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 Soc Robotics. Jun 03, 2020;12(6):1179-1201. [doi: 10.1007/s12369-020-00659-4]

30. Streiner DL, Norman GR. Health Measurement Scales: A Practical Guide to Their Development and Use. Oxford, UK. Oxford University Press; 2015.

31. Ginty AT. Psychometric properties. In: Gellman MD, Turner JR, editors. Encyclopedia of Behavioral Medicine. New York, NY. Springer; 2013.

32. Mokkink LB, Terwee CB, Stratford PW, Alonso J, Patrick DL, Riphagen I, et al. Evaluation of the methodological quality of systematic reviews of health status measurement instruments. Qual Life Res. Apr 2009;18(3):313-333. [FREE Full text] [doi: 10.1007/s11136-009-9451-9] [Medline: 19238586]

33. Terwee CB, Mokkink LB, Knol DL, Ostelo RW, Bouter LM, de Vet HC. Rating the methodological quality in systematic reviews of studies on measurement properties: a scoring system for the COSMIN checklist. Qual Life Res. May 2012;21(4):651-657. [FREE Full text] [doi: 10.1007/s11136-011-9960-1] [Medline: 21732199]

34. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ. Mar 29, 2021;372:n71. [FREE Full text] [doi: 10.1136/bmj.n71] [Medline: 33782057]

35. Ewald H, Klerings I, Wagner G, Heise TL, Stratil JM, Lhachimi SK, et al. Searching two or more databases decreased the risk of missing relevant studies: a metaresearch study. J Clin Epidemiol. Sep 2022;149:154-164. [FREE Full text] [doi: 10.1016/j.jclinepi.2022.05.022] [Medline: 35654269]

36. Mokkink LB, Prinsen CA, de Vet HC, Terwee CB. COSMIN methodology for conducting systematic reviews of patient-reported outcome measures (PROMs). In: Maggino F, editor. Encyclopedia of Quality of Life and Well-Being Research. Cham, Switzerland. Springer; 2023.

37. Prinsen CA, Mokkink LB, Bouter LM, Alonso J, Patrick DL, de Vet HC, et al. COSMIN guideline for systematic reviews of patient-reported outcome measures. Qual Life Res. May 2018;27(5):1147-1157. [FREE Full text] [doi: 10.1007/s11136-018-1798-3] [Medline: 29435801]

38. Terwee CB, Prinsen CA, Chiarotto A, de Vet HC, Bouter LM, Alonso J, et al. COSMIN methodology for assessing the content validity of PROMs: user manual. COSMIN. 2018. URL: https://cosmin.nl/wp-content/uploads/COSMIN-methodology-for-content-validity-user-manual-v1.pdf [accessed 2023-10-12]

39. Robert D, van den Bergh V. Children's Openness to Interacting with a Robot Scale (COIRS). In: Proceedings of the 23rd IEEE International Symposium on Robot and Human Interactive Communication. 2014. Presented at: ROMAN 2014; August 25-29, 2014; Edinburgh, UK. [doi: 10.1109/roman.2014.6926372]

40. de Jong C, Kühne R, Peter J, Straten CL, Barco A. What do children want from a social robot? Toward gratifications measures for child-robot interaction. In: Proceedings of the 28th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN). 2019. Presented at: RO-MAN 2019; October 14-18, 2019; New Delhi, India. [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. Feb 19, 2023;13(1):2911. [FREE Full text] [doi: 10.1038/s41598-023-28817-4] [Medline: 36807328]

42. Bevilacqua R, Di Rosa M, Riccardi GR, Pelliccioni G, Lattanzio F, Felici E, et al. Design and development of a scale for evaluating the acceptance of social robotics for older people: the robot era inventory. Front Neurorobot. Jul 07, 2022;16:883106. [FREE Full text] [doi: 10.3389/fnbot.2022.883106] [Medline: 35874107]

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. Feb 07, 2017;2017:9592405. [FREE Full text] [doi: 10.1155/2017/9592405] [Medline: 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. Int J Environ Res Public Health. Jun 07, 2021;18(11):6157. [FREE Full text] [doi: 10.3390/ijerph18116157] [Medline: 34200294]

45. Zhang F, Broz F, Ferrari O, Barakova E. TSES-R: an extended scale for measuring parental expectations toward robots for children in healthcare. In: Proceedings of the Companion of the 2023 ACM/IEEE International Conference on Human-Robot Interaction. 2023. Presented at: HRI '23; March 13-16, 2023; Stockholm, Sweden. [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 Robot. Jun 15, 2022;14:1559-1581. [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: Proceedings of the 23rd Working Conference on Virtual Enterprises. 2022. Presented at: PRO-VE 2022; September 19-21, 2022; Lisbon, Portugal. [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. Comput Hum Behav. May 2021;118:106700. [doi: 10.1016/j.chb.2021.106700]

49. Chen SC, Jones C, Moyle W. Health professional and workers attitudes towards the use of social robots for older adults in long-term care. Int J Soc Robot. Dec 07, 2019;12:1135-1147. [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. Int J Hum Comput Stud. Jul 2020;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. Interact Learn Environ. May 13, 2018;27(3):377-388. [doi: 10.1080/10494820.2018.1474234]

52.   Der Pütten AR, Bock N. Development and validation of the self-efficacy in human-robot-interaction scale (SE-HRI). ACM Trans Hum Robot Interact. Dec 05, 2018;7(3):1-30. [doi: 10.1145/3139352]

53.   Nomura T, Kanda T, Yamada S. Measurement of moral concern for robots. In: Proceedings of the 14th ACM/IEEE International Conference on Human-Robot Interaction. 2019. Presented at: HRI '19; March 11-14, 2019; Daegu, Korea (South). [doi: 10.1109/hri.2019.8673095]

54.   Carpinella CM, Wyman AB, Perez MA, Stroessner SJ. The robotic social attributes scale (RoSAS): development and validation. In: Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction. 2017. Presented at: HRI '17; March 6-9, 2017; Vienna, Austria. [doi: 10.1145/2909824.3020208]

55.   Oliveira R, Arriaga P, Stroessner SJ, Paiva A. Preliminary validation of the European Portuguese version of the robotic social attributes scale (RoSAS). Human Behav Emerg Technol. Dec 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: Proceedings of the 13th International Conference Social Robotics. 2021. Presented at: ICSR 2021; November 10-13, 2021; Singapore, Singapore. [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: Proceedings of the 9th International Conference Social Robotics. 2017. Presented at: ICSR 2017; November 22-24, 2017; Tsukuba, Japan. [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 Soc Robot. May 17, 2013;6:33-44. [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. In: Proceedings of the 7th International Conference Social Robotics. 2015. Presented at: ICSR 2015; October 26-30, 2015; Paris, France. [doi: 10.1007/978-3-319-25554-5_48]

60.   Turja T, Rantanen T, Oksanen A. Robot use self-efficacy in healthcare work (RUSH): development and validation of a new measure. AI Soc. Jul 29, 2017;34:137-143. [doi: 10.1007/s00146-017-0751-2]

61.   Nomura T, Suzuki T, Kanda T, Kato K. Measurement of anxiety toward robots. In: Proceedings of the 15th IEEE International Symposium on Robot and Human Interactive Communication. 2006. Presented at: ROMAN 2006; September 6-8, 2006; Hatfield, UK. [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. Int J Hum Comput Interact. Mar 14, 2023:1-10. [doi: 10.1080/10447318.2023.2188535]

63.   Nomura T, Kanda T. Rapport–expectation with a robot scale. Int J Soc Robot. Mar 13, 2015;8:21-30. [doi: 10.1007/s12369-015-0293-z]

64.   Straten CL, Kühne R, Peter J, de Jong C, Barco A. Closeness, trust, and perceived social support in child-robot relationship formation. Interact Stud. Jan 24, 2020;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. Sep 12, 2022;17:1353-1364. [FREE Full text] [doi: 10.2147/CIA.S367781] [Medline: 36117570]

66.   Nomura T, Syrdal DS, Dautenhahn K. Differences on social acceptance of humanoid robots between Japan and the UK. In: Proceedings of the 4th International Symposium on New Frontiers in Human-Robot Interaction. 2015. Presented at: HRI 2015; April 21-22, 2015; Canterbury, UK.

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. In: Proceedings of the 12th IEEE-RAS International Conference on Humanoid Robots. 2012. Presented at: HUMANOIDS 2012; November 29-December 1, 2012; Osaka, Japan. [doi: 10.1109/humanoids.2012.6651527]

68.   Nomura T, Suzuki T, Kanda T, Kato K. Measurement of negative attitudes toward robots. Interact Stud. Nov 15, 2006;7(3):437-454. [doi: 10.1075/is.7.3.14nom]

69.   Piçarra N, Giger JC, Pochwatko G, Gonçalves G. Validation of the Portuguese version of the negative attitudes towards robots scale. Eur Rev Appl Psychol. Mar 2015;65(2):93-104. [doi: 10.1016/j.erap.2014.11.002]

70.   Pochwatko G, Giger JC, Różańska-Walczuk M, Świdrak J, Kukiełka K, Możaryn J, et al. Polish version of the negative attitude toward robots scale (NARS-PL). J Autom Mob Robot Intell Syst. Jul 27, 2015;9(3):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. In: Proceedings of the Artificial Intelligence and the Simulation of Behaviour (AISB). 2009. Presented at: AISB'09; April 6-9, 2009; Edinburgh, UK.

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 Soc Robot. Jun 27, 2020;13:703-714. [doi: 10.1007/s12369-020-00669-2]

73. Ashton MC, Lee K. The HEXACO-60: a short measure of the major dimensions of personality. J Pers Assess. Jul 2009;91(4):340-345. [doi: 10.1080/00223890902935878] [Medline: 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 Soc Robot. Nov 20, 2008;1:71-81. [doi: 10.1007/s12369-008-0001-3]

75. Bem SL. The measurement of psychological androgyny. J Consult Clin Psychol. 1974;42(2):155-162. [doi: 10.1037/h0036215]

76. Fiske ST, Cuddy AJ, Glick P. Universal dimensions of social cognition: warmth and competence. Trends Cogn Sci. Feb 2007;11(2):77-83. [doi: 10.1016/j.tics.2006.11.005] [Medline: 17188552]

77. Eysenck HJ, Eysenck M. Personality and Individual Differences: A Natural Science Approach. New York, NY. Springer; 1985.

78. McKnight DH, Choudhury V, Kacmar C. Developing and validating trust measures for e-commerce: an integrative typology. Inf Syst Res. Sep 01, 2002;13(3):334-359. [doi: 10.1287/isre.13.3.334.81]

79. 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. Jul 01, 2011;2(2):1-25. [doi: 10.1145/1985347.1985353]

80. Costescu CA, David DO. Attitudes toward using social robots in psychotherapy. Transylvanian J Psychol. 2014;15(1):3. [FREE Full text]

81. 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: Proceedings of the 7th International Conference Social Robotics. 2015. Presented at: ICSR 2015; October 26-30, 2015; Paris, France. [doi: 10.1007/978-3-319-25554-5_3]

82. Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, et al. The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. J Clin Epidemiol. Jul 2010;63(7):737-745. [doi: 10.1016/j.jclinepi.2010.02.006] [Medline: 20494804]

83. Lenzner T. Effects of survey question comprehensibility on response quality. Field Methods. Sep 07, 2012;24(4):409-428. [doi: 10.1177/1525822X12448166]

84. de Vet HC, Ostelo RW, Terwee CB, van der Roer N, Knol DL, Beckerman H, et al. Minimally important change determined by a visual method integrating an anchor-based and a distribution-based approach. Qual Life Res. Feb 2007;16(1):131-142. [FREE Full text] [doi: 10.1007/s11136-006-9109-9] [Medline: 17033901]

85. Polit DF, Yang FM. Measurement and the Measurement of Change: A Primer for the Health Professions. Alphen aan den Rijn, The Netherlands. Wolters Kluwer; 2016.

86. Guyatt GH, Deyo RA, Charlson M, Levine MN, Mitchell A. Responsiveness and validity in health status measurement: a clarification. J Clin Epidemiol. 1989;42(5):403-408. [doi: 10.1016/0895-4356(89)90128-5] [Medline: 2659745]

87. Angst F. The new COSMIN guidelines confront traditional concepts of responsiveness. BMC Med Res Methodol. Nov 18, 2011;11:152; author reply 152. [FREE Full text] [doi: 10.1186/1471-2288-11-152] [Medline: 22099330]

88. Terwee CB, Peipert JD, Chapman R, Lai JS, Terluin B, Cella D, et al. Minimal important change (MIC): a conceptual clarification and systematic review of MIC estimates of PROMIS measures. Qual Life Res. Oct 2021;30(10):2729-2754. [FREE Full text] [doi: 10.1007/s11136-021-02925-y] [Medline: 34247326]

89. Fabrigar LR, Wegener DT. Exploratory Factor Analysis. Oxford, UK. Oxford University Press; 2011.

90. Hoyle RH. Handbook of Structural Equation Modeling. New York, NY. Guilford Publications; 2012.

91. Huang WY, Wong SH. Cross-cultural validation. In: Michalos AC, editor. Encyclopedia of Quality of Life and Well-Being Research. Dordrecht, The Netherlands. Springer; 2014.

92. 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. Front Psychol. Sep 18, 2019;10:2138. [FREE Full text] [doi: 10.3389/fpsyg.2019.02138] [Medline: 31620061]

93. Papadopoulos I, Wright S, Koulouglioti C, Ali S, Lazzarino R, Martín-García Á, et al. Socially assistive robots in health and social care: acceptance and cultural factors. Results from an exploratory international online survey. Jpn J Nurs Sci. Apr 2023;20(2):e12523. [doi: 10.1111/jjns.12523] [Medline: 36732396]

94. Gregorich SE. Do self-report instruments allow meaningful comparisons across diverse population groups? Testing measurement invariance using the confirmatory factor analysis framework. Med Care. Nov 2006;44(11 Suppl 3):S78-S94. [FREE Full text] [doi: 10.1097/01.mlr.0000245454.12228.8f] [Medline: 17060839]

95. de Vet HC, Terwee CB, Knol DL, Bouter LM. When to use agreement versus reliability measures. J Clin Epidemiol. Oct 2006;59(10):1033-1039. [doi: 10.1016/j.jclinepi.2005.10.015] [Medline: 16980142]

96. 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. [FREE Full text] [doi: 10.1155/2013/654589] [Medline: 24089638]

97. González-González CS, Violant-Holz V, Gil-Iranzo RM. Social robots in hospitals: a systematic review. Appl Sci. Jun 27, 2021;11(13):5976. [doi: 10.3390/app11135976]

98. Sébille V, Hardouin JB, Le Néel T, Kubis G, Boyer F, Guillemin F, et al. 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. Mar 25, 2010;10:24. [FREE Full text] [doi: 10.1186/1471-2288-10-24] [Medline: 20338031]

XSL•FO

RenderX

99. Jabrayilov R, Emons WH, Sijtsma K. Comparison of classical test theory and item response theory in individual change assessment. Appl Psychol Meas. Nov 2016;40(8):559-572. [FREE Full text] [doi: 10.1177/0146621616664046] [Medline: 29881070]

100. O'Connor A, McGarr O, Cantillon P, McCurtin A, Clifford A. Clinical performance assessment tools in physiotherapy practice education: a systematic review. Physiotherapy. Mar 2018;104(1):46-53. [doi: 10.1016/j.physio.2017.01.005] [Medline: 28844474]

101. Friberg JC. Considerations for test selection: how do validity and reliability impact diagnostic decisions? Child Lang Teach Ther. Apr 08, 2010;26(1):77-92. [doi: 10.1177/0265659009349972]

102. 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 Soc Robot. 2023;15(2):143-163. [FREE Full text] [doi: 10.1007/s12369-022-00925-7] [Medline: 36406778]

Abbreviations

AGFI: Adjusted Goodness of Fit Index
ATTUSR-C: Chinese version of the Attitudes Toward the Use of Social Robots
CFA: confirmatory factor analysis
CFI: comparative fit index
COSMIN: Consensus-Based Standards for the Selection of Health Measurement Instruments
EFA: exploratory factor analysis
FSQ: Frankenstein Syndrome Questionnaire
GFI: goodness-of-fit index
NARS: Negative Attitudes Toward Robots Scale
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
RAS: Robot Anxiety Scale
RMSEA: root mean square error of approximation
RoSAS: Robotic Social Attributes Scale
SRMR: standardized root mean square error
TLI: Tucker-Lewis index

XSL•FO
RenderX