Instrument Development for R-Service Quality: A Literature Review

Yanqing Lin
Aalto University School of Business
yanqing.lin@aalto.fi

Abstract
Motivated by a paucity of knowledge on the measurement of robotic service (r-service) quality, the current study strives to review the existing literature on r-service quality, with a focus on the potential methodological issues of developing measurement instruments and identifying the dimensionality of r-service quality. With a content analysis of 53 articles, this study identifies several methodological limitations of existing studies in developing measurement scales of r-service quality. This review reveals that dimensions of r-service quality are prone to be contingent on specific contexts of service industry and service type. Several common dimensions regarding evaluating r-service are identified, including tangibility, responsiveness, reliability, empathy, assurance, ease of use/usability, usefulness, anthropomorphism, perceived intelligence, and social presence. This study is the first systematic literature review on r-service quality dimensionality.

1. Introduction
Robotics and artificial intelligence (AI) have emerged in service sectors in recent years, resulting in a rapid rise of r-service. R-service refers to the service delivered by a robot [1]. Service robots are defined as "system-based autonomous and adaptable interfaces that interact, communicate and deliver service to an organization’s customers" [2, p. 909]. Service robots can be virtual or with a physical presence [3]. Generally, virtual robots, e.g., chatbots, are used in e-service, whereas robots with a physical presence are deployed in offline service contexts. The service robot market is snowballing and projected to grow at a compound annual growth rate of approximately a quarter and reach 102.5 billion USD by 2025 [4]. Such service industries as hotel [5], [6], tourism [7], [8], education [9], and restaurant [10], [11] are the early adopters of service robots. In particular, the Covid-19 pandemic has made robotics unprecedentedly relevant to service sectors, particularly hospitality, for deploying robots can keep social distance and decrease human touch [12], [13].

Robotics is predicted to profoundly change service sectors and add to an essential and integral part of future consumer experience [14], [15]. The majority of the existing literature focuses on the antecedents that contribute to consumer satisfaction and intention to use r-service based on the theories like SERVQUAL [11], [13], Technology Acceptance Model (TAM) [16], [17], or Social Presence Theory [18], [19]. Service robots are embedded with AI, allowing them to enter humans-preserving domains, such as contextual and bilateral interactions between robots (as regular staff) and consumers [20], [21]. Compared to conventional digital services (e.g., self-service technology), humanlike interaction and emotional elements may affect consumer responses to r-services, resulting in differentiated facilitators and barriers to tackle r-services [3]. The importance is being further emphasized to develop systematic scales concerning dimensions affecting consumers to adopt and evaluate r-services. Unfortunately, there is a lack of consensus on the factors affecting r-service quality. By instrumenting r-service quality, this study strives to bridge this gap.

R-service quality can be conceptualized as the extent to which a service robot facilitates efficient and effective service delivery, involving from pre- to post-delivery of r-service [22]. R-service quality plays a vital role in numerous aspects of service commerce, e.g., consumer attitudes towards the r-service[6], [9], [23], consumer satisfaction and loyalty [24], [25], willingness to use [3], [26], [27], intention to (re)use r-service [5], [7], [28], [29], recommendation intention [11], etc. In light of the apparent importance of r-service quality in service encounters, the achievement of superior r-service quality has been identified as a crucial strategy for service practitioners [13], [30]. With the advent of the AI era, r-service quality has been increasingly important for service success, helping service organizations sustain competitive advantage in volatile environments [13], [31]. However, the conceptualization and measurement development of r-service quality is at its embryonic stage [11], [30], [31].

Against this backdrop, the current study conducts a content analysis of the existing literature to examine determinants of r-service quality. To this end, this work reviews the existing studies on measurement models of r-service quality in the hope of discussing the dimensionality residing in a diversity of measurement factors. This study aims to offer insightful implications for developing instruments of r-service quality and for its application in commercial practice.
### Table 1. Critical Studies on Instrument Development for r-Service Quality

<table>
<thead>
<tr>
<th>Reference</th>
<th>Research context</th>
<th>Sampling</th>
<th>Type of Service robot</th>
<th>Data analysis procedure</th>
<th>Dependent variable</th>
<th>Dimensionality</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lin et al. [25]</td>
<td>Tourism guide cloud service quality</td>
<td>Survey (N = 536); Adult users (over 16 years old)</td>
<td>Tourism cloud services</td>
<td>SEM</td>
<td>Overall satisfaction, Loyalty</td>
<td>5 dimensions: Information Quality (6); Function quality (6); Real feedback (5); Multiple visual aids (2); Engagement (3)</td>
</tr>
<tr>
<td>Van et al. [27]</td>
<td>Service quality by human-machine interactive devices (HMI)</td>
<td>Open-ended interview (N = 5); senior managers or expert officials of the establishments</td>
<td>AI-enabled voice assistant robots</td>
<td>SEM</td>
<td>Value for Money Enhancers by use of HMI devices; Willingness to Use HMI</td>
<td>6 dimensions: Perceived Hygienic Usability (3); Perceived Safety for Usability (3); Assurance of Secure Service (2); Individualistic Involvement (3); Tangibility Associated with the Hygienic Service (3); Empathetic Secure Service and Update Information Sharing (3)</td>
</tr>
<tr>
<td>Chang and Tuin [13]</td>
<td>Service quality provided by robots in hotel setting</td>
<td>Survey (N = 201); guests of Chase Walker Hotel who used the robotic service</td>
<td>Hotel service robots</td>
<td>Importance performance analysis</td>
<td>Service quality</td>
<td>5 dimensions: Tangibles; Reliability; Responsiveness; Assurance; Empathy</td>
</tr>
<tr>
<td>de Kerven et al. [5]</td>
<td>Service quality provided by social robots in hospitality services</td>
<td>Semi-structured interview (N = 5); hospitality managers</td>
<td>Social robots in hospitality services</td>
<td>SEM</td>
<td>Intention to use social robot</td>
<td>7 dimensions: Empathy (3); Information Sharing (2); Perceived Usefulness (3); Perceived Ease of Use (3); Service Assurance (2); Personal Engagement (3); Tangibles (4)</td>
</tr>
<tr>
<td>Morita et al. [11]</td>
<td>Robotic service quality of a multi-robot cafe</td>
<td>Survey (N = 95); guests of the multi robot cafe at the 18th Yagami Festival of Keio University</td>
<td>Service robots in cafe</td>
<td>Bayesian network</td>
<td>Intention to revisit/recommend the robot cafe; Customer satisfaction</td>
<td>7 dimensions: Tangibles (3); Reliability (2); Responsiveness (1); Assurance (4); Empathy (1); Interactivity (2); Entertainment factor (8)</td>
</tr>
<tr>
<td>Chou et al. [31]</td>
<td>The service quality perceptions of human–robot interaction</td>
<td>Focus-group interview (N = 16); hotel managers</td>
<td>Service robots in hotel</td>
<td>EFA, one-way ANOVA</td>
<td>Perceived service quality</td>
<td>3 dimensions: Interaction quality (7); Outcome quality (6); Physical service environment (2)</td>
</tr>
<tr>
<td>Yu [32]</td>
<td>Service quality of hotel r-service</td>
<td>Experiment and survey (N = 233), hotel guests</td>
<td>Humanlike service robot</td>
<td>Three-way ANOVA</td>
<td>Service quality</td>
<td>4 dimensions: Responsiveness; Reliability; Assurance; Empathy</td>
</tr>
<tr>
<td>Park and Kwon [9]</td>
<td>Service quality provided by educational service robots</td>
<td>Survey (N = 609); Teachers in kindergartens and elementary schools, Parents, Researchers in the education field and robotic field, etc.</td>
<td>Teaching assistant (TA) robots</td>
<td>CFA, SEM</td>
<td>Intention to use</td>
<td>5 dimensions: Perceived enjoyment (4); Service quality (3); Perceived usefulness (4); Perceived ease of use (4); Attitudes (3)</td>
</tr>
<tr>
<td>Sohn and Kim [33]</td>
<td>Robot Utilization Expectation Index</td>
<td>Experiment (N = 102); majority are students with potential roles in robot utilization</td>
<td>Intelligent robot</td>
<td>CFA, SEM</td>
<td>Robot utilization expectation</td>
<td>6 dimensions: Robot reliability (4); Robot necessity (4); Function (5); Robot environment (4); Government policy (5)</td>
</tr>
<tr>
<td>Zhong et al. [6]</td>
<td>Service quality of robotic hotel service</td>
<td>Scenario-based experiment (N = 214), online respondents</td>
<td>Service robots in hotel</td>
<td>t-test</td>
<td>Purchase intention</td>
<td>2 dimensions: Hotel service type (traditional vs. robot hotel service); Attitudes (6)</td>
</tr>
<tr>
<td>Wang et al. [16]</td>
<td>Artificial intelligence (AI) application service quality</td>
<td>Survey (N = 237), random respondents</td>
<td>Smart speaker</td>
<td>CFA, SEM</td>
<td>Behavior intention</td>
<td>5 dimensions: Perceived usefulness (3); Perceived ease-of-use (3); Perceived behavioral control (3); Subjective norm (3); Attitude (3)</td>
</tr>
<tr>
<td>Zhang and Qi [34]</td>
<td>Service quality of AI robotic hotel</td>
<td>Survey (N = 102), adult residents living in Beijing city for more than one year</td>
<td>AI-based service robots</td>
<td>t-test, one-way ANOVA; Regression analysis</td>
<td>Robotic service expectation</td>
<td>5 dimensions: Tangibles; Reliability; Responsiveness; Assurance; Empathy</td>
</tr>
<tr>
<td>Dou et al. [35]</td>
<td>Perceived Robot Personalities</td>
<td>Experiment (N = 15), university students</td>
<td>Humanoid robot (Pepper)</td>
<td>Factor Analysis; Multivariate Statistical Analysis</td>
<td>Perceived robot personalities</td>
<td>2 dimensions (experimental manipulation): Robot voice types; Robot gesture types</td>
</tr>
<tr>
<td>Kim et al. [36]</td>
<td>Service quality of the robot museum</td>
<td>Survey (N = 57), robot museum visitors</td>
<td>Museum robots (Genibo and Aibo)</td>
<td>Paired t-test</td>
<td>Service quality</td>
<td>4 dimensions: Reliability (2); Empathy (2); Tangibility (2); Responsiveness (2); Other factors: Sociability (3); Social attraction (intimacy, 10); Interaction (6); Social influence (5); Emotions (3); Customer loyalty (2); Customer satisfaction (3)</td>
</tr>
<tr>
<td>Kim and Lee [37]</td>
<td>Service quality on ubiquitous robot companion (URC) personal robot service</td>
<td>Survey (N = 490), Korean users who used the personal robot in their home for 4 Months</td>
<td>URC personal robot</td>
<td>EFA, CFA</td>
<td>Intention to use</td>
<td>5 dimensions: Tangibility (tangibles, 7); Motion quality (responsiveness + assurance, 4); System quality (4); Perceived usefulness (7)</td>
</tr>
<tr>
<td>Bhat et al. [38]</td>
<td>Branding effects of social robots</td>
<td>Experiment (N = 530), a random sample</td>
<td>Social robots</td>
<td>SEM</td>
<td>Brand Trust; Brand experience</td>
<td>5 dimensions: Anthropomorphism, Attractiveness; Likeability; Perceived Intelligence; Perceived Safety</td>
</tr>
<tr>
<td>Merkle [39]</td>
<td>Customer Responses to Service Robots</td>
<td>Experiment (N = 120), random participants</td>
<td>Human service robot (Pepper)</td>
<td>ANOVA; Scheffe’s Post Hoc Test</td>
<td>Customer satisfaction</td>
<td>2 dimensions (experimental manipulations): Service provider (Service robots vs. Frontline employees); Service situation (appropriate service vs. service failure)</td>
</tr>
<tr>
<td>Stock and Mickle [40]</td>
<td>Customer responses to robotic innovative behavior</td>
<td>Experiment (N = 132); university students</td>
<td>Human service robot (Pepper)</td>
<td>MANOVA, Bonferroni post hoc test; Polynomial regression analysis</td>
<td>Customer satisfaction; Customer delight</td>
<td>3 dimensions: Perceived robotic interactive service behavior (ISB); Expectations; Confirmation between expected and perceived robotic ISB</td>
</tr>
<tr>
<td>Musavi and Koufaris [28]</td>
<td>Service quality provided Personal Intelligent Agents</td>
<td>Survey (N = 232), undergraduate college students at a Northeastern university US.</td>
<td>Personalized intelligent software systems</td>
<td>CFA, SEM</td>
<td>Continuance of use intention</td>
<td>6 dimensions: Perceived intelligence (5); Perceived anthropomorphism (6); Perceived usefulness; Disconfirmation of expectation; Satisfaction with use; Subjective norms</td>
</tr>
<tr>
<td>Sohn et al. [41]</td>
<td>Massaging service quality</td>
<td>Experiments (N = 74; N2 = 64); participants recruited from MTurk</td>
<td>Conversational user interfaces</td>
<td>One-way ANOVA, PLS-SEM</td>
<td>Privacy concerns</td>
<td>3 dimensions (experimental manipulations): The presence of CUI; Perceived social presence; Perception of being watched</td>
</tr>
<tr>
<td>Li et al. [8]</td>
<td>Intelligent Advisory Service quality</td>
<td>Survey (N = 83), guests recruited via emails and official online messages on personal contact lists</td>
<td>Virtual Advisory Service</td>
<td>-</td>
<td>Service reuse intentions</td>
<td>6 dimensions: Communication style similarity; Perceived clarity; Perceived engagement; Perceived enjoyment; Perceived credibility; Social presence</td>
</tr>
<tr>
<td>Schuetzler et al. [18]</td>
<td>Responses to online conversational agents</td>
<td>Experiment (N = 103), students a MIS course at a public university in U.S.</td>
<td>Conversational agents</td>
<td>SEM</td>
<td>Perceived humanness; Partner engagement</td>
<td>2 dimensions (experimental manipulation): Conversational relevance; Social presence</td>
</tr>
<tr>
<td>Bruckes et al. [42]</td>
<td>Robo-advisors service quality in bank</td>
<td>Survey (N = 246), participants familiarized with the concept of robo-advisory</td>
<td>Bank Robo-advisors</td>
<td>PLS-SEM</td>
<td>Intention to use</td>
<td>4 dimensions: Structural assurances; Trust in Banks; Perceived Risk; Initial Trust</td>
</tr>
<tr>
<td>Authors</td>
<td>Study Type</td>
<td>Methodology</td>
<td>Findings</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-------------------</td>
<td>-----------------------------------</td>
<td>-----------------------------------------------------------------------------</td>
<td>-------------------------------------------------------------------------</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vosybladha and Pathik [14]</td>
<td>Hotel robotic service quality</td>
<td>Online survey (N = 841), random sample. Laboratory experiment (N = 32), respondents invited through personal communication in a professional network setting</td>
<td>Hotel service robots: PLS-SEM; Adoption intention</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lu et al. [3]</td>
<td>Service robot integration willingness (SRiW) scale</td>
<td>Survey (N = 13483), consumer samples in the United States</td>
<td>Service robots in four service industries (e.g., hotels, restaurants, airlines, and retail stores): Hermeneutical approach; EFA; CFA; SEM; Invariance analysis; Willingness to use service robots</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Señor and Webber [7]</td>
<td>Tourism service quality</td>
<td>Survey (N = 1003), respondents recruited via email and social media</td>
<td>Service robots in tourism: EFA; Use intention</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stock and Merkle [43]</td>
<td>Robotic service quality</td>
<td>Experiment (N = 82), undergraduate and graduate students from a medium-sized university</td>
<td>Service frontline robots in hotel settings:ting test; Robot acceptance</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lu et al. [10]</td>
<td>Hotel robotic service quality</td>
<td>Experiment (N = 587), Consumer participants were recruited from Amazon Mechanical Turk</td>
<td>Service robots in a casual dining restaurant: Three-way ANOVA; Service encounter evaluation; Revisit intentions; WLM intentions</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chen and Tung [44]</td>
<td>Hotel robotic service quality</td>
<td>Experiment 1 (N = 60), university students; Experiment 2 (N = 180), participants recruited at the entrance of Tsim Sha Tsui Star Ferry Pier in Hong Kong</td>
<td>Hotel service robot: MANOVA; Ratings of brand experience;</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lee et al. [45]</td>
<td>Robotic service quality (the situation of service breakdown)</td>
<td>Scenario-based survey (N = 337), participants recruited from Amazon mTurk</td>
<td>Service robots: One-way analyses of variance;</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lin et al. [47]</td>
<td>Hotel service quality</td>
<td>Survey (N = 605), participants recruited from Amazon mTurk</td>
<td>Hotel service robots: Sentiment analysis; Customer acceptance of service robots in hotel</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Garvey et al. [26]</td>
<td>Hotel service quality</td>
<td>Survey (N = 439), participants recruited from Amazon mTurk</td>
<td>Hotel service robots: Artificially intelligent robotic device in hotel settings: CB-SEM; Willingness to use AI Devices; Objection of Using AI Devices</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Choi et al. [48]</td>
<td>Robotic service quality</td>
<td>Experiment (N = 173), US adult consumers recruited via Amazon mTurk</td>
<td>Service robot: ANOVA; Service encounter evaluation;</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lin and Mattila [23]</td>
<td>Hotel robotic service quality</td>
<td>Interview (N = 30), participants recruited in tourist spots and online; Survey (N = 215), individuals over the age of 18, recruited from Quatrax</td>
<td>Hotel service robot: CFA; SEM; Acceptance of service robots</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lee et al. [49]</td>
<td>Hotel service quality</td>
<td>Survey (N = 594), random consumers</td>
<td>Hotel assistant robots: EFA; Cluster analysis; Dis- criminant analysis; Intention to use robot assistant hotel</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tsao et al. [50]</td>
<td>Hospitality service quality</td>
<td>Exploitation service experiment (N = 30, prototype1: N = 18, prototype 2): participants from an academic conference focused on technology and tourism in the UK</td>
<td>Humanoid service robots: Qualitative multi-method approach, including exploratory service experimenta- tion, accompanied with ob- servation, questionnaire, interaction and photo- elicitation; Humanoid robot adoption in hospitality service encounters</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Zhong et al. [17]</td>
<td>Hotel service quality</td>
<td>Survey (N = 217), hotel guests who stayed in the rooms with service robots as the workforce.</td>
<td>Hotel service robots: EFA; CFA; Grouped regression analysis; SEM; Behavioral Intention</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bhat et al. [51]</td>
<td>Robotic service quality</td>
<td>Literature retrieval (N = 71)</td>
<td>Physical robots, chatbots, and other AI: Meta-analysis; Intention to use</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chi et al. [52]</td>
<td>Social Service Robot Interaction Trust (SSRIT) Scale</td>
<td>Survey (N = 316), a customer panel was recruited through Amazon MTurk.</td>
<td>Social service robot: EFA; Social service robot interaction trust</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Notes: SEM means Structural Equation Modeling; EFA means Exploratory Factor Analysis; CFA means Confirmatory Factor Analysis; CB-SEM means Covariance-based Structural Equation Modeling; (M)ANOVA means (Multivariate) Analysis of Variance; The numbers with brackets mean the number of items of the respective construct (that can be found in the reviewed papers).
2. Literature Review

The literature retrieval was carried out in January 2021 through three databases of AIS Library, Scopus, and Web of Science. In addition, Google Scholar was used as a supplementary source of literature. These databases cover most of the current literature and are, in turn, the most consulted by academic staff from various fields of knowledge [53]. After gathering all the retrieval records, removing duplicates, and screening out unqualified papers, the final sample consists of 55 articles (see Figure 1). These studies either focus on developing an instrument for measuring r-service or aim at consumer responses to r-service. They are subjected to a comprehensive, in-depth content analysis of the key methodological aspects of developing various r-service quality scales and their proposed dimensions. Table 1 lists the key studies reviewed in this study.

![Figure 1. Flow Diagram of Literature Selection](image)

2.1. Adequacy of dimensions

There is a lack of a widely accepted measure of r-service quality in the current literature. Existing r-service quality measures typically concern the design of service robots and quality of service delivery, including factors triggering consumer willingness [3], [19], [47], consumer satisfaction [11], [40], [54], and/or intention to (re)use [14], [49], [51]. In this regard, Lu et al. [3] develop a six-dimensional SRJW scale: performance efficacy, intrinsic motivation, social influence, anthropomorphism, emotions, and facilitating conditions. Stock and Merkle [40] identify three constructs dominating consumer evaluation of satisfaction, i.e., perceived robotic innovative service behavior (ISB), expectations, and confirmation between perceived and expected robotic ISB. Tussyadiah and Parkh [14] verify the determinants of consumer intention to adopt hotel service robots: anthropomorphism, perceived intelligence, and security.

In addition to the humanlike characteristics of service robots, some other scholars bend their efforts to develop more direct measures of the instruments of r-service quality. This research stream typically concentrates on two views: i) replicating or modifying the renowned scale dubbed SERVQUAL [55], [56]; ii) adopting technology acceptance theories, such as TAM, to develop robot-contextualized constructs [9], [16], [17].

“SERVQUAL is a generic instrument with good reliability and validity and broad applicability” [56, p. 445], which endorses five dominant dimensions: tangibility, responsiveness, reliability, empathy, and assurance [55]. Its principle is to assess service quality through the gap between delivered service performance and service expectations [55], [56]. A wealth of evidence shows that SERVQUAL has been verified and extensively applied in human-delivered services, e.g., hospitality and bank service [57], as well as e-service [58]. However, problems with SERVQUAL still arise concerning conceptualization and operationalization [57]. As proof, challenges occur to the applicable generalization of the five dimensions in different service industries because of the context-bounded attribute of service quality [57]. In respect of this, Zhang and Qi [59] apply SERVQUAL to r-service in hotels, and their results collapsed the five dimensions into two dimensions of tangibles and responsiveness. To evaluate service quality in the context of multi-robot café, Morita et al. [11] extend SERVQUAL dimensions from five to seven dimensions by including interactivity and entertainment factor.

A string of literature regarding r-service quality is established based on a consumer version of technology acceptance. According to the earliest TAM [60], perceived usefulness and ease of use are the two dominants affecting personal attitudes, thereby intention and actual behavior to use the technology. Nevertheless, the AI attributes allow service robots to gain several characteristics, such as bilateral interaction and anthropomorphism, differentiated from regular technologies (e.g., information systems) [21], [61]. This gives rise to difficulties in the applicability of the core factors from TAM or its extended theories. To address this issue, previous studies normally adopt other elements involving robot design, interactional components, and consumer emotional constructs. To illustrate, Zhong et al. [17] build an acceptance model of hotel service robots, and besides confirming the factors of usefulness, ease of use, and attitude, they also verify the significant roles of perceived value, self-efficacy, and perceived behavioral control. Stock and Merkle [43] combine TAM and role theory and developed a humanoid robot acceptance model with three dimensions: functional components (ease of use and usefulness), informational component (informativeness of interaction), and relational component (benevolence, user satisfaction, and understanding). Wang et al. [16] develop a consumer acceptance model for AI service with usefulness, ease of use, attitude, perceived behavioral control, and subjective norm.

To sum, both views mentioned above warrant further consideration. SERVQUAL is initially developed for evaluating personal-interactional services. As the saying goes, “the definitions and relative importance of
the five service quality dimensions change when customers interact with technology rather than with service personnel” [62, p. 171], its dimensions might not directly transpose to r-service. On the other hand, a service robot is not just a regular technology but goes beyond standard technologies to enter the field preserved for human beings. This may lead to the general technology acceptance theories being inapplicable in the r-service context. As a result, neither SERVQUAL nor technology acceptance theories constitute a comprehensive instrument for assessing r-service quality. Several studies attempt to develop specific measurement scales for r-service quality, but without considering the overall picture of the factors introduced by different studies, which motivates and shapes the substance of this study.

2.2. Dimensionality of the r-service quality

Based on the content analysis of the reviewed studies, several assertions about the dimensionality of r-service quality can be concluded. First, there is a lack of a consensus in the construct of r-service quality regarding its dimensions. However, some dimensions are often considered, such as SERVQUAL five dimensions, dimensions related to technology acceptance, and robot-design characteristics. Second, several dimensions of r-service quality in the reviewed papers are similar with or recur from conventional service quality.

2.2.1. R-service quality constructs. Except for a few studies that use experimental manipulation to verify a specific single dimension [19], [63]–[65], most studies have multiple dimensional constructs for r-service quality, ranging from 2 [16], [48] to 11 dimensions [52]. Due to the lack of consensus regarding constructs of r-service quality on its dimensions, many dimensions merely appear in specific studies or research contexts. The determinants of r-service quality depend on involving service industries and particular service types. For example, anthropomorphism plays an essential role in service robots with physical attendance [14], [38], which is not the case for virtual robots that care more about communication patterns and language cues [63], [65]. However, some constructs, such as reliability [11], [13], [33], [59] and anthropomorphism [3], [14], [28], [38], have been frequently identified in previous studies. It is conceivable that there are several common dimensions considered by consumers when evaluating r-services. Ten dimensions are identified:

Reliability. As one of the five prominent dimensions of SERVQUAL, reliability is conceptualized as the capability to perform a promised service dependably, accurately, and timely [55]. Among the reviewed studies, reliability plays a significant role in general service quality [13], [32], [36], service expectation [33], and behavioral intention [11].

Assurance. In the r-service context, assurance refers to knowledge and courtesy of service robots and their abilities to inspire consumers’ trust and confidence in receiving service cf. [55]. By reflecting service experience, assurance indicates that qualified r-services not only cater to particular consumer requirements but also represent safe and dependable services that are trustworthy in long-term use [5], [56]. Assurance constitutes an essential component towards customer satisfaction [11], willingness [27] or intention to use [5], [37], [42], and overall r-service quality [13].

Tangibility. Tangibility refers to physical facilities, equipment, and appearance of robots in r-services cf. [55], as one of the most common factors of r-service quality [66]. In the reviewed studies, tangibility is a significant dimension that determines overall service quality [13], [36], service expectations [59], and willingness [27] and intention to use robots [5], [11], [37].

Responsiveness. It refers to the willingness to help customers and offer prompt service [55]. With increasing service robots deployed to replace human personnel to delivery services, this dimension also matters in r-services, affecting consumer satisfaction and loyalty [24], service expectations [59], overall service quality [13], [32], [36], and intention to use [11], [37].

Empathy. Empathy can be viewed as caring and individualized attention the robotics offers for customers [55]. This dimension is relevant since service robots can mimic humans and pay attention to consumers when interacting with them [11]. In this regard, researchers report that empathy affects consumer satisfaction and loyalty [24], intention to use robots [5], [11], [37], and overall service quality [32].

Functional component. This dimension derives from a technology acceptance perspective. It is covered for that, albeit robots act as a replacement for human staff, it is essentially a novel technology that can be intimidating and complex for many individuals. Ease of use/usability is a reflection of consumer friendliness, whereas usefulness manifests ones’ perception regarding the outcome of the service experience. Both play an integral part in consumers’ behavioral intention [5], [9], [16], [17], [43], [51].

Anthropomorphism. It refers to that a robot is humanlike regarding either physical appearance or psychological features, such as emotions and gestures [3]. Anthropomorphism plays an essential role in affecting human-robot interaction [1] and acts as a determinant in consumer trust [38], [52], willingness to use [3], [26], [47], and intention to (re)use [14], [28], [51]. Many studies exploring the impact of anthropomorphism draw upon Uncanny Valley Theory (UVT). Some similar constructs, such as perceived humanness [18], [63], [64], physical appearance [10], [67], and uncanniness [63] can also be seen in the reviewed studies. Note, however, that the level of anthropomorphism is not necessarily linearly associated with r-service quality, according to UVT.
**Perceived intelligence.** It means consumers’ perception that robots can learn, reason, and solve problems [51], [68]. Perceived intelligence concerning interacting with a service robot has been endorsed as a critical factor for accepting the robot [68], [69]. This dimension determines consumer trust [38], service robot adoption intention [14], and intention to (re)use [51].

**Social presence.** This dimension manifests how people react socially to robots through psychological simulations of non-human intelligence as a real creature [15], [49]. Social presence is vital in determining behavioral intention to use service robots in hotel settings [37] and advisory services [8].

Despite that these dimensions are relatively frequently identified, they are neither necessarily generic nor exhaustive. Instruments for r-service quality measurement have to vary and are contingent on specific service industries and service types. In general, dimensions of r-service quality in the reviewed studies, except for the common dimensions mentioned above, can be subdivided into three categories: i) robot-related component, such as sociability [36], social attraction [36], autonomy [70], safety [70], animacy [69], likability [69], imitation [70], and benevolence [43]; ii) functional component, such as understanding [43], performance efficacy [3], interactivity [11][31], and scalability [70]; iii) consumer-related component, such as perceived safety [69], entertainment [11], and enjoyment [25]. Note that the common dimensions could be utilized as a starting point for instrument development of r-service quality.

### 2.2.2. Comparison with conventional service quality.

While some new dimensions of r-service quality have been extracted, several dimensions are similar to or recur from conventional human service and e-service. Concretely, reliability and assurance, both prominent in the offline context of human service, are reported as top priorities of r-service quality [13]. The other three SERVQUAL dimensions — tangibility, responsiveness, and empathy — are also reported in several studies of r-service quality, e.g., [11], [13], [36]. However, mixed results exist in the reviewed literature. For instance, Zhang and Qi [59] show that tangibility and responsiveness significantly increase consumer expectations of robotic hotels, whereas the effects of reliability, assurance, and empathy are insignificant. Morita et al. [11] report the high importance of reliability and tangibility when evaluating r-services, while the responsiveness dimension is subscribed as low importance. The dispute may result from differentiated interpretations of these dimensions when service robots are deployed to replace human personnel. More specifically, assurance and empathy are different in the r-service context from its connotations in human service, since robots can always be polite and work consistently within rules to fulfill consumer needs while human staff may show extra caring attitude and go beyond rules to solve problems.

Furthermore, ease of use/usability and usefulness, which are widely used in e-service quality, have been adapted to r-service quality [5], [9], [16]. Such dimensions play important roles in evaluating r-service since robots can be novel technologies for many individuals, and induced unfamiliarity can intimidate them and make them feel complex to be involved in r-services. One issue requiring more attention is that some overlaps exist concerning connotations of SERVQUAL dimensions and ease of use/usability. Specifically, there is an intersectional area between tangibility and ease of use when considering the robot design and aesthetics.

Notably, several dimensions that are tailored for robotics, particularly humanoid robots, take essential parts in r-service quality. These dimensions include anthropomorphism [1], [3], [14], [26], [28], [38], [47], [51], [52], perceived intelligence [14], [28], [38], [51], social presence [8], [49], autonomy [70], animacy [69], imitation [70], etc. Past studies usually allude that human appearance tends to trigger positive perceptions and attitudes towards robots [14], [23].

### 2.3. Methodological issues

Studies concerning r-service quality utilize various methodologies, e.g., qualitative, quantitative, and hybrid methods. The first stage of establishing a measurement scale is to conduct qualitative research to identify multiple dimensions, which can be fulfilled with different qualitative approaches, e.g., the critical incident technique (CIT). CIT helps recall impressive events and identify important factors for the subject through qualitative interviews, which has proved valuable in developing service quality scales [71], [72]. Whereas some of the reviewed works use interviews to identify constructs of r-service quality, the application of CIT, as well as other qualitative methods, e.g., focus-group study and Delphi method, are recommended in future studies at the early stage of identifying r-service quality dimensions.

#### 2.3.1. Sampling.

The reviewed studies collected samples on r-service quality from various populations. Convenience sampling [7], [31], [44], [63] has been frequently used, whereas random sampling appear in some studies [14], [16], [38], [39]. A few studies utilize sampling of guests in real service settings, such as hospitality [13], [31], [32] and restaurants [11], [29]. Many studies recruit students in their surveys [18], [28], [40], [43], [73].

Several research limitations exist. First, several studies obtain mainly their respondents through personal networks. Albeit recruiting respondents merely from personal networks can be more time-/effort-saving than other sampling methods, which need to fulfill specific requirements, sampling bias would be inevitable due to constraints derived from, e.g., geographical and social milieus, in particular when a representative sample is requested [73]. Second, a major limitation in the reviewed studies is that most samples are not actual consumers of
r-services. Many respondents in the reviewed studies are recruited online. They are generally asked to self-report their perceptions of r-services based on reading research descriptions instead of experiencing r-service delivery. Respondents’ perceptions of service quality with scenario descriptions may differ from experiencing r-service in real settings. Furthermore, pre-delivery perceptions of r-services, e.g., comfort with robots, might be more significant in physical service delivery, thus causing differences in individual perceptions for r-service quality [31]. By having respondents reflect their perceptions of r-services that they were not familiar with or even had not experienced, deliverables might have suffered limitations in the accuracy of findings. In this vein, the recruitment of respondents should carefully consider sample qualification to safeguard reliability in future studies. Third, many reviewed studies are based on relatively small-scale samples, which may challenge the robustness and generalization of results. In this light, samples with larger scales and more diversities should be considered in future studies.

2.3.2. Considered service industries. A vast body of the reviewed studies collects consumer data within a specific (or a type of) service sector [9], [10], [13], [36], whereas only minimal studies are across several service industries [3]. Among them, studies based on the hotel industry dominate this research stream, e.g., [13], [31]. Other specific sectors considered include, e.g., restaurant [11], [29], education [9], museum [36], household [37], tourism [8], and bank [7], [42]. Notably, Lu et al. [3] verify their instruments across four service industries: hotels, restaurants, airlines, and retail.

2.3.3. Survey administration. Both online and on-site approaches are used for data collection. Concerning qualitative research, online [23] and offline interviews have mainly been used, the latter of which includes open-ended interviews [27], semi-structured interviews [5], [24], and focus-group interviews [31]. A few studies also use literature analysis [51], [66] to identify factors impacting r-service quality. Regarding quantitative studies, online surveys are the most widely used by researchers. The online distribution platforms include Amazon Mechanical Turk [26], [47], [48], [52] and personal networks [8], [14], [63], [64], whereas in-person surveys are among guests in hotels [5], [13], café or restaurant [11], [29], etc. Given the importance of survey administration, the administration mode needs to be clarified in more detail. Future studies are expected to scrutinize in-person surveys, particularly respondents with real experiences of r-service.

2.3.4. Measurement items generation. Both inductive (e.g., literature reviews) and deductive methods (e.g., exploratory research) are utilized to study r-service quality. Many studies strive to establish a research model to verify factors that affect behavioral intention [16], [42], [49], [51]. A few studies devote themselves to systematically developing related scales [3], [52]. Specifically, through a systematic literature review, interviews, and focus-group study, Chi et al. [52] launched a scale that measures consumer trust toward interaction with service robots. With rigorous quantitative studies, the SSRIT scale with 50 items is validated [52]. Based on a literature review and qualitative interviews, Lu et al. [3] established the SRIT scale consisting of 36 items. Moreover, in several studies, interviews are conducted among employees or managers for constructs and items generation [5], [24], [31], [74].

No consensus has been reached yet regarding the conceptualization and dimensions of r-service quality. Taking the dimensions of robot design as an example, some studies include communication pattern [8], [63] into robot-design constructs; others consider more the visual presence of robots, e.g., anthropomorphism [3], [14], [26], [47], [51]. The diversity of constructs in different studies underlines the lack of a consensus regarding the components of r-service quality. This may result from two main reasons. First, the conceptualization of the definition, scope, and dimensions of r-service leaves to be framed. Second, while some qualitative research relies on literature analysis and/or interview to generate constructs, a high proportion of studies directly develop research models and use data-driven approaches, e.g., EFA, to validate measurement items. In this light, future research is expected to develop a conceptual framework more specifically, comprehensively accounting for literature, expert panels, consumers, and operators. Thereby, the components of r-service quality, its dimensions, and scale-items can be identified and validated.

2.3.5. Dimensionality analysis. Given that many of the observed studies investigate the impact of antecedents on related dependent variables, such as behavioral intention and consumer satisfaction, a number of studies utilize SEM to test research models [5], [14], [38]. Besides, the dimensionality of the measures is examined primarily with EFA [3], [7], [17], [23], [37], [49] and/or CFA [3], [16], [17], [24], [28], [33], [37], [52], [74].

Whereas the purpose of EFA is “to identify the factor structure or model for a set of variables” [75, p. 10] via dropping underqualified items, its use has been challenged with its demerits, such as the nonuniqueness of the estimates accounted for factor loadings and the lack of indicators of goodness-of-fit as the case of CFA does [72]. Furthermore, the possibility in EFA that items load on more than one factor may impact the distinctiveness and interpretation of items [76]. Taking together the merits of CFA, such as allowing a comparison of different model specifications, a combination of EFA and CFA is expected in future studies.

3. Conclusion and Implications

The present study reviews the current knowledge on the instruments of r-service quality and contributes
to i) identifying common dimensions for r-service quality; ii) outlining the methodology of instrument development for r-service quality.

This study detects heterogeneity in dimensions in the reviewed studies concerning r-service quality. However, several critical dimensions are identified in previous studies, including SERVQUAL five dimensions, ease of use/usability, usefulness, and three robot-related dimensions. This study demonstrates that the measurement of r-service quality shares several dimensions with traditional human service and e-service. Meanwhile, some dimensions of r-service quality are distinctive from conventional service settings. These distinctive dimensions focus upon social-emotional factors induced mainly by robot characteristics.

3.1. Research implications

This study offers several research implications. First, merely a few studies specifically develop and validate related measurement scales, i.e., the SRIW [3] and SSRIT scale [52]. Given a scarcity of knowledge on r-service quality, it calls for more studies to develop measurement scales for r-service quality.

Second, most of the identified common dimensions are function-oriented dimensions that reflect the service delivery process, including SERVQUAL five dimensions and ease of use/usability. However, the high dependence on these dimensions has been criticized by scholars for constituting the misspecification of service quality in both human service and e-service. Thus, future studies are expected to integrate other views from pre/post-delivery and reexamine the conceptualization of r-service quality.

Third, this study shows that more specific dimensions are contingent on particular service industries and service types. It is reasonable since different service contexts have different determinants to foster better service quality. There is no utterly generic measurement instrument of service quality, and even the widely-utilized SERVQUAL do not apply universally. Thus, a valid measurement scale of r-service quality for specific contexts should include service industry/type-specific dimensions as supplements for the generic dimensions. It would also be interesting to assess the weights of different dimensions across different robots in future studies.

Finally, more attention should be paid to methodological issues. Future studies should make more efforts in the methodological approaches to identifying dimensions and generating measurement items of r-service quality, as well as the sampling methods and size. Random and relatively bigger sample sizes across multiple service industries are warranted in future studies.

3.2 Managerial implications

These findings allow us to propose several suggestions for business practitioners designing/manufacturing/adopting r-services. First, considering the identification of SERVQUAL five dimensions in r-service quality, r-service managers should fully understand the keys to effective deployment of service robots: i) ensuring the delivery of promised services occur in a reliable, accurate, and timely manner; ii) having a suitable appearance (it is important to take UVIT into account), equipment, and interacting skills for the specific service; iii) helping consumers actively solve problems and providing prompt service; iv) performing reliable services consistently and politely; v) paying caring and individualized attention to customers. Since r-services are still in an infant stage, service failures are inevitable. Under this circumstance, assistance from human staff is necessary for r-service delivery, particularly when consumers encounter interaction difficulties. In this vein, the possible negative perceptions induced by service failures could be alleviated.

Second, considering that several reviewed studies emphasize the importance of ease of use/usability, robot manufacturers should pay more attention to the function design of service robots to make them easier to navigate and interact with. Third, as a replacement for human personnel, robot characteristics are of significance for consumer perceptions. Consumers need to feel emotionally positive during service transactions. Thus, robot manufacturers should focus on the psychological evaluation of robots as social entities and account for social-emotional elements in robot design.

4. Acknowledgement

Yanqing Lin gratefully acknowledges financial support from the Marcus Wallenberg Foundation (Grant Nos. 12-3407-40; 13-3998-14, 14-4368-17). The project is partly supported by the Finnish National Agency for Education (Project ID: AIRobinSers), The Jenny and Antti Wihuri Foundation (Project ID: Embracing the Future: Service Robots for Business Values and Social Welfare), and the National Natural Science Foundation of China (Grant No.71962008).

5. References


for Information Systems Can Conversational User Interfaces Be Harmful? The Undesirable Effects on Privacy Concern Can Conversational User Interfaces Be Harmful? The Undesirable Effects on Privacy Concern,” 2019.


