A Systematic Validation of the Robotic Social Attributes Scale (RoSAS)

Pawinee Pithayarungsarit
George Mason University
Fairfax, VA, USA
ppithaya@gmu.edu

Laura Saad
Naval Research Laboratory
Washington DC, USA
laura.s.saad.ctr@nrl.navy.mil

J. Gregory Trafton
Naval Research Laboratory
Washington DC, USA
greg.trafton@nrl.navy.mil

Eileen Roesler
George Mason University
Fairfax, VA, USA
eroesle@gmu.edu

***Abstract*—The Robotic Social Attributes Scale (RoSAS) is widely used in human-robot interaction research to measure the social perception of robots, including warmth, competence, and discomfort. As previous researchers have found ambiguous support for the RoSAS's three-factor structure, the current study aims to evaluate the proposed structure by conducting a confirmatory factor analysis (CFA) using openly available datasets. The CFA (n = 1107) showed that the three-factor model had a poor model fit. This suggests that the RoSAS's three dimensions might better be used as separate scales instead of measuring a broad concept of social perception. When separating by stimulus type, only stimuli using words and vignettes had an acceptable model fit, indicating that the RoSAS might be more suitable for word/vignette stimuli. We recommend using the RoSAS's individual subscales as separate constructs rather than measuring social attributes in general. This approach also aligns with what most research has already adopted.**

***Index Terms*—social perception, robot, RoSAS, factor analysis**

I. Introduction

With the increasing frequency and importance of interaction between people and robots in various domains such as healthcare, education, and hospitality [1]–[4], it is crucial to understand how people perceive and evaluate robots' social attributes [5]. Social attributes are traits and characteristics people ascribe to robots (e.g., friendly, smart, or assertive) based on design features or perceived social category membership when forming impressions of robots [5], [6]. The perception of these attributes can influence people's acceptance, expectations, and interaction with robots [5], [7]. Therefore, understanding how robots with different features will likely be perceived can be useful for effective human-robot interaction (HRI) [8]. Thus, it is necessary to have psychometrically valid tools that can accurately measure aspects of the social perception of robots.

The Robotic Social Attributes Scale (RoSAS) is one of the most widely used scales in HRI research designed to measure people's judgment of the social attributes of robots [5]. Developed based on a set of social attributes from existing scales in robot and social perception [6], [9], [10], the RoSAS consists of three dimensions underlying the social evaluation of robots: warmth, competence, and discomfort. Warmth is related to how likeable and friendly the robot is perceived, while competence is related to how capable and dependable the robot is perceived [5], [8]. These two dimensions are similar to the two universal dimensions of person perception in social psychology [6]. Discomfort was considered a unique attribute of robots' social perception involving feelings of scariness and danger [5].

The RoSAS consists of 18 items with each dimension containing six items as shown in Table I. Participants are instructed, "Using the scale provided, how closely are the words below associated with the category robots?" Each item is then measured on a 9-point Likert scale from 1 (definitely not associated) to 9 (definitely associated).

TABLE I
RoSAS items in each dimension

Warmth	Competence	Discomfort
Happy	Capable	Scary
Feeling	Responsive	Strange
Social	Interactive	Awkward
Organic	Reliable	Dangerous
Compassionate	Competent	Awful
Emotional	Knowledgeable	Aggressive

The original study demonstrated that the RoSAS has good reliability and validity. All three dimensions were shown to have high internal consistency (Study 4: warmth; α = .92, competence; α = .95, and discomfort; α = .90 [5]). The validity assessment also showed that people rated each dimension differently depending on whether the robots had feminine or masculine faces and whether the robots were more humanlike or machinelike, consistent with the expectation of gender-typicality and machinelikeness [5], [11].

The psychometric evaluation of the RoSAS was also provided by previous studies that have used the RoSAS to study a range of research questions in HRI with various types of robots and methods [8]. Several studies showed that each individual dimension of the RoSAS demonstrated high reliabilities with Cronbach's α ranging from .81 to .95 [11]–[14]. Moreover, some studies also supported the validity of the RoSAS. For example, robots were perceived as warmer when they showed high vulnerability [15], and as more competent when they provided help through physical collaboration [16]. Together, these findings suggest that each dimension of the RoSAS reliably measures its intended construct.

Despite the evidence for good psychometric properties of each dimension, the overall factor structure of the RoSAS

979-8-3503-7893-1/25/$31.00 ©2025 IEEE

was not sufficiently validated. When developing a new scale, a test of factor structure such as a confirmatory factor analysis (CFA) is important as it allows researchers to evaluate if the hypothesized factor structure extracted from a previous scale development is the same across different time points or samples [17]. Thus, confirming the underlying structure of the construct being assessed is recommended to ensure the quality and utility of a scale. However, the original RoSAS study did not conduct a factor structure test of the RoSAS to confirm the proposed three-factor structure in a different sample [5]. Moreover, many previous researchers used individual subscales rather than the full scale (e.g., [18]–[21]) which did not allow for testing the factor structure. This implies that researchers might be more interested in measuring each individual attribute to answer their research questions rather than measuring social perception as a whole. Nevertheless, given the RoSAS's original intention as a general measure to capture the fundamental aspects of robot social perception [5], [8], it is crucial to evaluate whether the RoSAS measures the general construct of social perception.

A few studies have attempted to test the RoSAS's latent factor structure and found no convincing evidence supporting the three-factor model. For example, Pan et al. (2018) [12] found evidence supporting high internal consistency and unidimensionality of each RoSAS dimension. However, a full factor structure evaluation using a CFA was not performed due to a small sample size of 22 participants. Similarly, Oliveira et al., (2021) [22] translated the original RoSAS to Portuguese (European) and found that the three-factor structure of the translated RoSAS had a poor model fit, with several items not loading onto their expected dimensions. After removing seven items, the adapted 11-item RoSAS indicated an adequate model fit. However, the initial poor model fit could potentially be due to the translation. Lastly, Kettle et al. (2022) [23] performed a CFA to validate the three-factor structure using the pre-existing dataset from Schadenberg et al. [24]. Although the result showed that the scale demonstrated good reliability for each dimension, the three-factor model did not provide a good fit with the data. Together, the previous tests of factor structure do not provide support that the three dimensions of the RoSAS are measuring a coherent construct of social perception.

Moreover, the RoSAS was originally developed using only the word "robot" without any images, definitions, or specifications of the type of robot to avoid any preconceptions or expectations that people might have of the robots. Thus, the scale was developed to be used as a standardized measurement of general robot perception that was not tied to any particular type of robot or stimuli [5]. Previous research has used the RoSAS with various types of robots (e.g., images [25], videos [26], or physical interaction [27]). However, no previous research has tested if the RoSAS's three-factor structure holds true for attitudes towards specific robots presented via images, videos, or in-person, instead of imagined robots in general. Thus, it is important to evaluate the three-factor structure across different stimulus types to ensure the generalizability of the scale.

Taken together, the current literature suggests that there is strong support for the individual subscales of the RoSAS, but less support for the three-factor structure as a whole, and no validation with different stimulus types. Therefore, the current study aims to examine the proposed three-factor structure of the RoSAS by conducting CFAs on the original English version of the RoSAS using existing publicly available datasets and also examining the factor structure by types of stimuli.

II. Methods

A. Systematic Dataset Search

A systematic database search was conducted to search for the publicly available datasets for the RoSAS. Utilizing the Publish or Perish software [28], a list of 563 articles citing the original scale paper "The Robotic Social Attributes Scale (RoSAS): Development and Validation" [5] were obtained from the Google Scholar search (up until May 2024). These articles were reviewed and then coded based on the following inclusion criteria: 1) If the RoSAS was used for data collection, 2) If a full original scale was used (18 items of the three subscales: warmth, competence, discomfort), 3) If the English version of the RoSAS was used, and 4) If the dataset with raw item scores was publicly available. Articles that met these criteria were considered eligible for the analysis, and the datasets from these articles were obtained. Moreover, some researchers shared their datasets upon request. The Google Sheet used for the coding process is available in the following OSF Repository: https://osf.io/kzx47.

B. Combined Dataset of the RoSAS

Based on the systematic search, four datasets were obtained from four published studies [24], [29]–[31] provided via the online repositories or direct contact with the authors. Two additional datasets were provided directly by researchers[1]. For the purpose of the current study, the de-identified data only include the raw item scores of the RoSAS and the types of stimuli used when responding to the scale. Together, a total of six datasets was obtained, consisting of 663 participants with a total of 1107 observations.

Across all six datasets, the types of stimuli used include robot images, robot videos, VR drones, physical robots, the word "robot", and vignettes related to robots. We categorized the stimuli into three groups: Described (words and vignettes), Depicted (robot images and videos), and Embodied (physical robots and VR drone), with the latter two groups based on the HRI taxonomy for categorizing exposure to robots [2]. The number of observations for each stimulus type is Described (n = 222), Depicted (n = 664), and Embodied (n = 221).

C. Statistical Analysis

To evaluate the scale's reliability, the internal consistency of each subscale was assessed using Cronbach's alpha (α). The McDonald's Omega ω reliability was also conducted with the

[1]https://osf.io/srv8c, (Trafton & McCurry, unpublished)

Omega total (ω_t) for each subscale and the Omega hierarchical (ω_h) for the overall scale. The ω_t represents the reliability estimate of overall variance that is due to a general factor and specific factors (i.e., dimensions), while ω_h represents the reliability of a general factor after controlling for the variance of specific factors [32]. In the presence of a strong general factor, the ω_h value will be high and the difference between ω_h and ω_t will be low.

Multiple CFAs were conducted to analyze the RoSAS's three-factor structure (warmth, competence, and discomfort). First, all six datasets were combined into a single large dataset and a CFA was conducted on this combined dataset. Additionally, the datasets were grouped based on the types of stimuli (Described, Depicted, and Embodied) and CFAs were run separately for each type of stimuli. This allowed us to investigate if the RoSAS's three-factor structure would fit for the different types of stimuli.

To evaluate the model fit indices (see Table II), the following guidelines were used. The non-significant chi-square test of goodness of fit indicates a good model fit. However, since the chi-square is sensitive to larger sample sizes, additional fit indices were also considered [33], [34]. Comparative Fit Index (CFI) and Tucker-Lewis Index (TLI) values of more than .95 are considered a good fit, while the values between .90 and .95 are considered an acceptable fit, and the values less than .90 are considered a poor fit [35]. Root Mean Square Error of Approximation (RMSEA) and Standardized Root Mean Square Residual (SRMR) with values less than .05 are considered a good fit, while the values between .05 and .08 are considered a reasonable fit, and values more than .08 are considered a poor fit [36]. Lastly, the correlations between each dimension were also evaluated. We expected low correlations between each dimension as was reported in the original RoSAS study [5].

III. Results

A. Reliability

The reliability of each subscale was calculated based on the combined dataset. For Cronbach's alpha, each subscale demonstrated a good internal consistency with α = .88 for warmth, α = .86 for competence, and α = .83 for discomfort. The reliability for each subscale based on Omega total was also high: ω_t = .89 for warmth, ω_t = .86 for competence, and ω_t = .84 for discomfort. However, Omega hierarchical for the overall scale was poor with ω_h = .42. The high ω_t values and low ω_h value suggest that variances in the data were explained more by the specific factors (subscales) and less by a general factor.

B. Confirmatory Factor Analysis for Combined Dataset

The multivariate normality assumption was violated (Mardia's Skewness = 4915.15, $p < .001$ and Mardia's Kurtosis = 31.42, $p < .001$). As the items were assessed using a 9-point Likert scale, the data can be considered as continuous [37]. Therefore, the Satorra-Bentler chi-square correction [38] (MLM estimator), which is robust to non-normality, was used to perform the CFA for our datasets.

To test the RoSAS's three-factor model, the CFA using the MLM estimator was performed on a combined dataset with a total of 1107 observations[2]. The results showed that the three-factor model did not indicate an adequate fit (as shown in Table II). However, each item demonstrated a high factor loading of more than .40 onto its respective dimension, with the factor loadings for all items ranging from .50 to .85. Upon inspecting the Schmid Leiman Factor loadings, the items "Dangerous" and "Aggressive" did not load onto the general factor of perceived social attributes, and the item "Interactive" had cross-loadings between the warmth and competence dimensions.

TABLE II
Fit Indices of the RoSAS's Three-Factor Model

Model	χ^2(df)	CFI	TLI	RMSEA	SRMR
Combined	1245.05(132)	0.874	0.854	0.087	**0.069**
Described	235.37(132)	**0.929**	**0.918**	**0.059**	**0.058**
Depicted	981.91(132)	0.828	0.801	0.098	0.082
Embodied	353.13(132)	0.87	0.85	0.087	**0.072**

Bolded indicates a reasonable (not good) fit with the following criteria: CFI and TLI between 0.90 and 0.95; RMSEA and SRMR between 0.05 and 0.08

The correlations between the subscales were also low, such that warmth and competence were positively correlated (r = .33) while discomfort was negatively correlated with warmth (r = -.19) and competence (r = -.21). This is consistent with the original RoSAS development which suggested that the three dimensions measure independent constructs of robot perception [5].

C. Confirmatory Factor Analyses Based on Stimulus Types

When analyzing the three-factor model based on the types of stimuli (see Table II), the CFAs showed that only Described (words and vignettes) had an acceptable model fit. For the Described stimuli, all items had high factor loadings of more than .40 on their respective factors, except for the item "Interactive" which had a factor loading of .36 on the competence factor.

Depicted (robot images and videos) and Embodied (physical robots and VR drone) had a poor model fit. However, all items still had high factor loadings of more than .50 onto their respective factors for both the Depicted and Embodied stimuli.

IV. Discussion

Overall, the CFA results based on the combined dataset across different stimuli showed that the three-factor model had a rather poor model fit indicating that the RoSAS's three dimensions (warmth, competence, and discomfort) did not adequately explain the variation in social perception as a common latent variable. Thus, the three dimensions might not measure the same general concept of social perception.

[2]The CFA was conducted without reverse-coding the discomfort dimension as it makes no difference for this analysis.

However, most items had a high factor loading onto their respective dimension, implying a good item-factor relationship. This implies that while the relationships between the three dimensions might not match the hypothesized structure of the social perception construct, the items are well-aligned with their respective dimensions and measure what each dimension intends to capture.

This poor model fit for the three-factor model might be due to two main reasons. First, social perception is a broad, multidimensional concept that can be broken down into different dimensions and sub-dimensions [39]. Previous research has proposed different dimensions of social perception of robots which imply the heterogeneity in dimensions underlying the construct. For example, the Social Perception of Robots Scale (SPRS) suggested three dimensions of social perception of robots including anthropomorphism (morphological perception and feeling toward robots), morality/sociability (perceived social personality of robots), and activity/cooperation (required attributes for smooth interaction and cooperation) [39]. Despite the good reliabilities for each dimension, the CFA for the SPRS also showed a poor model fit [39], similar to our CFA results for the RoSAS. Thus, this raises a question regarding the cohesiveness of the social perception as a single general construct. This might explain the poor fit of the three-factor model with the overall construct of robot's social perception.

Moreover, the original RoSAS did not measure an overall mean score of social perception [5], and no one has done so to our best knowledge. While understanding how people perceive robots' social attributes is important for designing and implementing effective HRI, the current finding suggests that the RoSAS's three dimensions might better be used as separate scales to measure the constructs of warmth, competence, and discomfort, instead of measuring the broad concept of social perception. According to our data, 28% of the studies using the RoSAS (86 out of 300 studies) also used only one or two subscales to measure their construct of interest. This shows that each subscale has already been used to measure independent constructs of robot attributes.

Second, the observed poor fit might be because the RoSAS's three-factor structure is more suitable for described robot stimuli. When analyzing the three-factor structure for each stimulus type, the results showed a good model fit for Described (word and vignettes) stimuli only. This is similar to the method used to develop the RoSAS which only used the word "robot" without any images or specification of the type of robot [5]. This implies that the RoSAS might be more suitable for described stimuli without any visualization of the robots because visualizations might influence the social attributes people ascribe to the robots. Thus, the RoSAS's three dimensions of robot's social perception might not be applicable to all stimulus types as proposed by the original development of the scale.

V. Implications and Future Research

By utilizing existing datasets, the current study was able to analyze the factor structure of the RoSAS using a large sample which ensures adequate statistical power for CFA [40]. While the observed poor model fit of the original three-factor structure implies that each dimension might not measure the same general concept of social perception, each separate dimension can still be used as a psychometrically sound measure. Therefore, the current results suggest that the RoSAS might better be used to measure people's evaluation of warmth, competence, and discomfort separately. Future researchers should ensure that the theorized factor structures of the scales are sufficiently validated and confirmed before using them to ensure the credibility and reliability of the research findings.

Moreover, the good model fit for Described stimuli suggested that the RoSAS's three-factor structures might be more suitable for evaluating the social perception of robot stimuli presented without any visualization, such as in reference to the word "robot" or when reading vignettes. Previous research found that people perceived and evaluated robots differently when the robots were physically presented, compared to video-displayed or even vignettes [41]–[44]. Thus, future research should explore whether people might attribute different social dimensions to the robots depending on the type of robot stimuli used.

Lastly, our systematic search showed that the RoSAS had been used in various different languages other than English (e.g., German, French, and Portuguese). However, we only found one validation/adaptation paper of the Portuguese version that evaluated the factor structure of the translated RoSAS [22]. Nevertheless, it cannot be guaranteed that the validations of the other translations do not exist as they may have been published in their original language. Thus, we strongly recommend future researchers validate and collect validation of translated versions of the RoSAS to evaluate the factor structure of these translated versions.

VI. Conclusion

Utilizing a large dataset, a CFA across 1107 observations failed to find evidence that supports the proposed three-factor structure of the RoSAS. When separating by stimulus types, only Described stimuli (words and vignettes) had an acceptable model fit. However, the three subscales can still be used to measure the perception of each dimension separately. Our approach of combining existing datasets allowed us to validate a scale's psychometric properties utilizing a large sample. However, only six datasets could be obtained from the current study's systematic search. Therefore, future researchers are encouraged to make their de-identified datasets publicly available to support the psychometric evaluation of scales within the HRI field.

Acknowledgment

We would like to thank the authors of the following papers [24], [29]–[31] who provided the RoSAS datasets used in this current study.

REFERENCES

[1] T. B. Sheridan, "Human–robot interaction: status and challenges," *Human factors*, vol. 58, no. 4, pp. 525–532, 2016.

[2] L. Onnasch and E. Roesler, "A taxonomy to structure and analyze human–robot interaction," *International Journal of Social Robotics*, vol. 13, no. 4, pp. 833–849, 2021.

[3] S. Thrun, "Toward a framework for human-robot interaction," *Human–Computer Interaction*, vol. 19, no. 1-2, pp. 9–24, 2004.

[4] R. de Kervenoael, R. Hasan, A. Schwob, and E. Goh, "Leveraging human-robot interaction in hospitality services: Incorporating the role of perceived value, empathy, and information sharing into visitors' intentions to use social robots," *Tourism Management*, vol. 78, p. 104042, 2020.

[5] C. M. Carpinella, A. B. Wyman, M. A. Perez, and S. J. Stroessner, "The robotic social attributes scale (rosas) development and validation," in *Proceedings of the 2017 ACM/IEEE International Conference on human-robot interaction*, pp. 254–262, 2017.

[6] S. T. Fiske, A. J. Cuddy, and P. Glick, "Universal dimensions of social cognition: Warmth and competence," *Trends in cognitive sciences*, vol. 11, no. 2, pp. 77–83, 2007.

[7] L. Bishop, A. van Maris, S. Dogramadzi, and N. Zook, "Social robots: The influence of human and robot characteristics on acceptance," *Paladyn, Journal of Behavioral Robotics*, vol. 10, no. 1, pp. 346–358, 2019.

[8] S. J. Stroessner, "On the social perception of robots: measurement, moderation, and implications," in *Living with robots*, pp. 21–47, Elsevier, 2020.

[9] C. Bartneck, D. Kulić, E. Croft, and S. Zoghbi, "Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots," *International journal of social robotics*, vol. 1, pp. 71–81, 2009.

[10] S. L. Bem, "The measurement of psychological androgyny.," *Journal of consulting and clinical psychology*, vol. 42, no. 2, p. 155, 1974.

[11] S. J. Stroessner and J. Benitez, "The social perception of humanoid and non-humanoid robots: Effects of gendered and machinelike features," *International Journal of Social Robotics*, vol. 11, pp. 305–315, 2019.

[12] M. K. Pan, E. A. Croft, and G. Niemeyer, "Evaluating social perception of human-to-robot handovers using the robot social attributes scale (rosas)," in *Proceedings of the 2018 ACM/IEEE international conference on human-robot interaction*, pp. 443–451, 2018.

[13] J. Benitez, A. B. Wyman, C. M. Carpinella, and S. J. Stroessner, "The authority of appearance: How robot features influence trait inferences and evaluative responses," in *2017 26th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, pp. 397–404, IEEE, 2017.

[14] L. Hoffmann, N. Bock, and A. M. Rosenthal vd Pütten, "The peculiarities of robot embodiment (emcorp-scale) development, validation and initial test of the embodiment and corporeality of artificial agents scale," in *Proceedings of the 2018 ACM/IEEE international conference on human-robot interaction*, pp. 370–378, 2018.

[15] S. S. Sebo, M. Traeger, M. Jung, and B. Scassellati, "The ripple effects of vulnerability: The effects of a robot's vulnerable behavior on trust in human-robot teams," in *Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction*, pp. 178–186, 2018.

[16] M. Bonani, R. Oliveira, F. Correia, A. Rodrigues, T. Guerreiro, and A. Paiva, "What my eyes can't see, a robot can show me: Exploring the collaboration between blind people and robots," in *Proceedings of the 20th International ACM SIGACCESS Conference on Computers and Accessibility*, pp. 15–27, 2018.

[17] G. O. Boateng, T. B. Neilands, E. A. Frongillo, H. R. Melgar-Quiñonez, and S. L. Young, "Best practices for developing and validating scales for health, social, and behavioral research: a primer," *Frontiers in public health*, vol. 6, p. 149, 2018.

[18] E. Senft, S. Satake, and T. Kanda, "Would you mind me if i pass by you? socially-appropriate behaviour for an omni-based social robot in narrow environment," in *Proceedings of the 2020 ACM/IEEE International Conference on Human-Robot Interaction*, pp. 539–547, 2020.

[19] F. I. Doğan, I. Torre, and I. Leite, "Asking follow-up clarifications to resolve ambiguities in human-robot conversation," in *2022 17th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pp. 461–469, IEEE, 2022.

[20] H. Nemlekar, D. Dutia, and Z. Li, "Object transfer point estimation for fluent human-robot handovers," in *2019 International Conference on Robotics and Automation (ICRA)*, pp. 2627–2633, IEEE, 2019.

[21] X. Z. Tan, S. Reig, E. J. Carter, and A. Steinfeld, "From one to another: how robot-robot interaction affects users' perceptions following a transition between robots," in *2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pp. 114–122, IEEE, 2019.

[22] R. Oliveira, P. Arriaga, S. J. Stroessner, and A. Paiva, "Preliminary validation of the european portuguese version of the robotic social attributes scale (rosas)," *Human Behavior and Emerging Technologies*, vol. 3, no. 5, pp. 750–758, 2021.

[23] L. Kettle, L. Melles, and K. Simpson, "Moderation analysis of gender in social robot interactions," in *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, vol. 66, pp. 802–806, SAGE Publications Sage CA: Los Angeles, CA, 2022.

[24] B. R. Schadenberg, D. Reidsma, D. K. Heylen, and V. Evers, ""i see what you did there" understanding people's social perception of a robot and its predictability," *ACM Transactions on Human-Robot Interaction (THRI)*, vol. 10, no. 3, pp. 1–28, 2021.

[25] K. Klüber and L. Onnasch, "Appearance is not everything-preferred feature combinations for care robots," *Computers in Human Behavior*, vol. 128, p. 107128, 2022.

[26] L. Morillo-Mendez, M. G. Schrooten, A. Loutfi, and O. M. Mozos, "Age-related differences in the perception of robotic referential gaze in human-robot interaction," *International journal of social robotics*, vol. 16, no. 6, pp. 1069–1081, 2024.

[27] N. Spatola and O. A. Wudarczyk, "Implicit attitudes towards robots predict explicit attitudes, semantic distance between robots and humans, anthropomorphism, and prosocial behavior: From attitudes to human–robot interaction," *International Journal of Social Robotics*, vol. 13, no. 5, pp. 1149–1159, 2021.

[28] A.-W. Harzing, *The publish or perish book*. Tarma Software Research Pty Limited Melbourne, Australia, 2010.

[29] S. Sebo, L. L. Dong, N. Chang, M. Lewkowicz, M. Schutzman, and B. Scassellati, "The influence of robot verbal support on human team members: Encouraging outgroup contributions and suppressing ingroup supportive behavior," *Frontiers in Psychology*, vol. 11, p. 590181, 2020.

[30] R. Bretin, E. Cross, and M. Khamis, "Co-existing with drones: A virtual exploration of proxemic behaviours and users' insights on social drones," *International Journal of Social Robotics*, vol. 16, no. 3, pp. 547–567, 2024.

[31] M. T. Parreira, S. Gillet, K. Winkle, and I. Leite, "How did we miss this? a case study on unintended biases in robot social behavior," in *Companion of the 2023 ACM/IEEE International Conference on Human-Robot Interaction*, pp. 11–20, 2023.

[32] S. B. Green and Y. Yang, "Evaluation of dimensionality in the assessment of internal consistency reliability: Coefficient alpha and omega coefficients," *Educational Measurement: Issues and Practice*, vol. 34, no. 4, pp. 14–20, 2015.

[33] K. A. Bollen, *Structural equations with latent variables*, vol. 210. John Wiley & Sons, 1989.

[34] R. L. Worthington and T. A. Whittaker, "Scale development research: A content analysis and recommendations for best practices," *The counseling psychologist*, vol. 34, no. 6, pp. 806–838, 2006.

[35] L.-t. Hu and P. M. Bentler, "Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives," *Structural equation modeling: a multidisciplinary journal*, vol. 6, no. 1, pp. 1–55, 1999.

[36] M. W. Browne and R. Cudeck, "Alternative ways of assessing model fit," *Sociological methods & research*, vol. 21, no. 2, pp. 230–258, 1992.

[37] G. H. Lubke and B. O. Muthén, "Applying multigroup confirmatory factor models for continuous outcomes to likert scale data complicates meaningful group comparisons," *Structural equation modeling*, vol. 11, no. 4, pp. 514–534, 2004.

[38] A. Satorra, "Scaled and adjusted restricted tests in multi-sample analysis of moment structures," in *Innovations in multivariate statistical analysis: A Festschrift for Heinz Neudecker*, pp. 233–247, Springer, 2000.

[39] S. Mandl, M. Bretschneider, F. Asbrock, B. Meyer, and A. Strobel, "The social perception of robots scale (sprs): developing and testing a scale for successful interaction between humans and robots," in *Working Conference on Virtual Enterprises*, pp. 321–334, Springer, 2022.

[40] L. S. Lambert and D. A. Newman, "Construct development and validation in three practical steps: Recommendations for reviewers, editors, and authors," *Organizational Research Methods*, vol. 26, no. 4, pp. 574–607, 2023.

[41] W. A. Bainbridge, J. W. Hart, E. S. Kim, and B. Scassellati, “The benefits of interactions with physically present robots over video-displayed agents,” *International Journal of Social Robotics*, vol. 3, pp. 41–52, 2011.
[42] J. Wainer, D. J. Feil-Seifer, D. A. Shell, and M. J. Mataric, “Embodiment and human-robot interaction: A task-based perspective,” in *RO-MAN 2007-The 16th IEEE international symposium on robot and human interactive communication*, pp. 872–877, IEEE, 2007.
[43] K. Wzietek, F. W. Siebert, and E. Roesler, “Imagination vs. reality: Investigating the acceptance and preferred anthropomorphism in service hri,” in *Proceedings of the 2024 ACM/IEEE International Conference on Human-Robot Interaction*, pp. 800–809, 2024.
[44] J. zu Putlitz and E. Roesler, “Let’s get physical: The influence of embodiment on industrial human-robot interaction,” in *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, p. 10711813241264206, SAGE Publications Sage CA: Los Angeles, CA, 2024.