The Reduced-Length Connection–Coordination Rapport (CCR) Scale

TING-HAN LIN and GUAN CHEN, The University of Chicago, Chicago, Illinois, USA
BILGE MUTLU, University of Wisconsin-Madison, Madison, Wisconsin, USA
J. GREGORY TRAFTON, US Naval Research Laboratory, Washington, District of Columbia, USA
SARAH SEBO, The University of Chicago, Chicago, Illinois, USA

Robots such as those serving as educational tutors, healthcare supporters, and collaborative partners must develop "rapport," a construct that encompasses mutual understanding and interpersonal connection with people, to ensure their long-term success. In our earlier work, we constructed, evaluated, and validated an 18-item Connection–Coordination Rapport (CCR) scale to measure human–robot rapport (Studies 1–3). Even though the full-length 18-item CCR scale measures rapport thoroughly, it may not always be practical for researchers to adopt given its relatively long length. Therefore, in this work, we developed a reduced-length version of the CCR scale that still effectively measures rapport using just 8 items. Following recommended practices for short-form development and validation, we leveraged the input of Human–Robot Interaction (HRI) experts (Study 4, $N = 30$) to shorten the CCR scale from 18 items to 8 items (4 items per factor). Then, we evaluated this reduced-length CCR scale on a new sample (Study 5, $N = 186$) where online participants watched a HRI video and evaluated it using both the full-length and reduced-length CCR scales. We validated the reduced-length CCR scale by showing that it has high internal reliability, high overlap with the full-length CCR scale, a consistent factor structure, high construct validity, and significant time savings.

CCS Concepts: • **Human-centered computing → Empirical studies in HCI**; **User studies**; **HCI theory, concepts and models**; **HCI design and evaluation methods**;

Additional Key Words and Phrases: rapport, human-robot rapport, human-robot interaction, scale development, scale reduction

ACM Reference format:
Ting-Han Lin, Guan Chen, Bilge Mutlu, J. Gregory Trafton, and Sarah Sebo. 2026. The Reduced-Length Connection–Coordination Rapport (CCR) Scale. *ACM Trans. Hum.-Robot Interact.* 15, 3, Article 57 (April 2026), 29 pages.
https://doi.org/10.1145/3796533

This work was supported by the National Science Foundation (awards #2339581 and #2312352).
Authors' Contact Information: Ting-Han Lin (corresponding author), The University of Chicago, Chicago, Illinois, USA; e-mail: tinghan@uchicago.edu; Guan Chen, The University of Chicago, Chicago, Illinois, USA; e-mail: guanc@uchicago.edu; Bilge Mutlu, University of Wisconsin-Madison, Madison, Wisconsin, USA; e-mail: bilge@cs.wisc.edu; J. Gregory Trafton, US Naval Research Laboratory, Washington, District of Columbia, USA; e-mail: greg.trafton@nrl.navy.mil; Sarah Sebo, The University of Chicago, Chicago, Illinois, USA; e-mail: sarahsebo@uchicago.edu.

ACM 2573-9522/2026/4-ART57
https://doi.org/10.1145/3796533

1 Introduction

From offering caregiving for children and the elderly [44, 69, 87, 88, 100, 102], giving educational assistance [55, 56, 64, 96], to providing customer service [23, 51, 58, 81, 90], robots are steadily integrating into our everyday lives. For these robots to be successful over the long term, they need to establish harmonious relationships with people. These harmonious human–robot relationships can be characterized by the construct of *rapport*, which we define as "*a feeling of mutual understanding and interpersonal connection among individuals developed through interactions*" [60].

Developing human–robot rapport has several benefits, including promoting more effective communication and resolution of conflicts [14, 79, 80] and increasing people's willingness to engage with the robot [53, 64, 77, 96] even after the novelty effect of the robot fades [36]. Although rapport has the potential to significantly improve **Human–Robot Interactions (HRIs)**, we found that the HRI community lacked a psychometrically validated rapport scale that follows traditional scale construct guidelines [19, 26, 68, 83] to enable an accurate measure of human–robot rapport.

Therefore, in our recently published work (Lin et al. [60]), we designed a psychometrically validated rapport scale, the 18-item **Connection–Coordination Rapport (CCR)** Scale, that can be used for different types of robots and contexts (see Table 1 and its detailed usage in Box 1). The 18-item CCR scale offers a comprehensive assessment of rapport using two factors reinforced by psychological literature [97]: connection and coordination. It contains short, forward-coded items and can be used from both first-person and third-person perspectives. Our multiple online and in-person validation studies in Lin et al. [60] demonstrate that the CCR scale has a consistent two-factor structure, shows high internal validity and reliability, and outperforms the current best alternative scale [37] used to measure human–robot rapport.

Although our 18-item CCR scale provides a psychometrically validated method to measure rapport, we considered whether having access to a shorter scale that maintained its validity would be beneficial for the HRI community. Performing scale reduction on the 18-item CCR scale offers several key advantages. It decreases participant fatigue by yielding better data quality and lower dropout rate [25, 29, 32, 45, 62, 98]. It gives researchers more feasibility to collect participant responses if they need to repeatedly ask participants to rate rapport [30]. It also allows researchers who are exploring rapport-adjacent topics to adopt the scale more easily since HRI user studies (e.g., [61, 76, 105, 106]) usually administer multiple scales at once.

In this work, we developed and validated a reduced-length 8-item CCR scale. We first summarize the development of the full-length 18-item CCR scale from Lin et al. [60], where we conducted Studies 1–3 to construct, evaluate, and validate the scale. We then present our methodology for scale shortening, which can broadly be applied to any future scale-shortening efforts in HRI. Next, we describe how we developed the reduced-length CCR scale by performing content validation (Study 4). We then detail how we validated the reduced-length CCR scale by administering both the full-length and reduced-length CCR scales to an independent sample (Study 5). Our resulting reduced-length 8-item CCR scale has high internal reliability, convergent validity, divergent validity, and high correlation and similar inter-item correlation with the full-length CCR scale. It also takes participants significantly less time to complete than the full-length CCR scale. Our validation of the reduced-length scale also provided further evidence of the full-length CCR scale's construct validity and the strong influence of a robot's responsiveness on human–robot rapport.

2 Development of the Full-Length 18-Item CCR Scale

The development of the full-length 18-item CCR Scale is fully detailed in our prior work, Lin et al. [60]. In this section, we briefly summarize the key methods, studies, and results from the development of the full-length CCR scale and specifically emphasize details relevant to our efforts

Table 1. Scale Items for the Full-Length 18-Item CCR Scale and the Reduced-Length 8-Item CCR Scale

Factor One: Connection	Full-Length 18-Item CCR Scale	Reduced-Length 8-Item CCR Scale
Connection	✓	✓
Getting along	✓	✓
Empathy	✓	✓
Respect	✓	✓
Closeness	✓	
Friendliness	✓	
Liking each other	✓	
Sympathy	✓	
Warmth	✓	
Positivity	✓	
Enthusiasm	✓	
Excitement	✓	
Factor Two: Coordination	Full-Length 18-Item CCR Scale	Reduced-Length 8-Item CCR Scale
Coordination	✓	✓
Engagement	✓	✓
Smooth flow	✓	✓
Attentiveness	✓	✓
Focus	✓	
Equal participation	✓	

Question wording for both scales: "Rate how much you think the following was present in the interaction." *A five-point scale was used for both scales*: Strongly Disagree, Disagree, Neither Agree nor Disagree, Agree, Strongly Agree. *Rapport calculation for both scales*: (1) Average item ratings from the Connection factor, (2) average item ratings from the Coordination factor, and (3) average values from (1) and (2) to get one score for rapport.

to reduce its length. The protocols for Studies 1–3 presented in this section were approved by the University of Chicago's Institutional Review Board (#IRB24-0884).

2.1 Initial Scale Item Generation

To create a set of initial scale items to capture the construct of rapport, we extracted phrases from dictionary definitions of rapport [1–10], papers retrieved from Google Scholar with the search term "rapport" [11, 15, 16, 28, 31, 34, 38–41, 47, 54, 57, 70, 71, 75, 91, 92, 97, 101] and "rapport in human–robot interaction" [12, 14, 23, 33, 35, 48, 51, 52, 63, 64, 72, 74, 77, 79, 80, 84, 86, 90, 93, 103], and the general public's perceptions of rapport (collected in an online pre-study, Study 0). We then employed the Delphi method [21] to combine relevant themes and identified a list of 67 candidate scale items (see Supplemental Materials Table SI). Through numerous rounds of internal discussion, we eventually narrowed the list to 27 scale items that describe positive characteristics of interaction (see the top 27 items listed in Supplemental Materials Table SI). The reasons for rejecting the other 40 items are also detailed in Supplemental Materials Table SI (e.g., some of the items are excluded

because they are idiomatic or too colloquial, which could be difficult for respondents to understand or translate into other languages).

2.2 Scale Construction (Study 1)

We conducted Study 1 to determine which (if any) of the 27 scale items chosen in Section 2.1 would be best to measure human–robot rapport. In an online between-subjects study ($N = 288$), participants were randomly assigned to watch one of four videos each depicting a HRI[1] and rate the video using the 27 scale items. For each item, participants were presented with the prompt "Rate how much you think the following was present in the interaction" on a five-point Likert scale ranging from 1 (Strongly Disagree) to 5 (Strongly Agree). From participants' responses, we found that **Exploratory Factor Analysis (EFA)**, Very Simple Structure, Empirical Bayesian Information Criterion, and Parallel Analysis all indicated a clear two-factor scale structure. We thus proceeded to analyze participants' responses using EFA with two factors under promax rotation (refer to each item's factor loadings in Table 2). After inspecting the EFA results, we kept scale items that had one loading above or equal to 0.6 and had no cross-loadings above 0.3 (suggested by [27, 66]). From the remaining 19 scale items (see the top 19 items listed in Table 2), we found high internal reliability using Cronbach's alpha and McDonald's omega ($\alpha = 0.96$, $\omega_{total} = 0.97$) and named the two factors based on the item's contextual definitions: connection (13 items) and coordination (6 items). The names of the two factors give the scale its name—CCR Scale.

2.3 Scale Evaluation (Study 2)

We then executed Study 2 to evaluate our 19-item CCR scale from Study 1 and compare its effectiveness with an alternative state-of-the-art scale: Rapport Scale 4 from Gratch et al. [37] (which we will refer to as the Gratch Rapport Scale; see Supplemental Materials Table SII). In an online within-subjects study ($N = 201$), we asked participants to watch a new set of four videos of HRI[2] and rate all four videos using our 19-item CCR scale and the Gratch Rapport Scale. We also asked participants to rank the four videos from highest to lowest rapport. Based on participants' responses, we first removed the scale item "Deep conversation" as it was notably more context-specific than the other items. We then performed **Confirmatory Factor Analysis (CFA)** on the participant's ratings for the remaining 18 scale items and found that our CCR scale has a good fit based on **Comparative Fit Index (CFI)**, **Tucker–Lewis Index (TLI)**, and **Standardized Root Mean Square Residual (SRMR)**: $CFI = 0.997$ (≥ 0.95), $TLI = 0.996$ (≥ 0.95), and $SRMR = 0.046$ (≤ 0.08).

Our next objective was to compare our CCR scale with the Gratch Rapport Scale in terms of its effectiveness in measuring rapport. We found that both scales were strongly correlated ($R = 0.84$, $p < 0.001$), and they were both highly reliable: CCR scale ($\alpha = 0.97$, $\omega_{total} = 0.97$) and Gratch Rapport Scale ($\alpha = 0.89$, $\omega_{total} = 0.94$). This indicates that both scales have items that, together, measure a cohesive construct and that the constructs measured by both scales are highly similar. Finally, to compare the efficacy of both scales in measuring rapport, we evaluated the ordinal regression model fit using participants' video ranking of rapport and observed that our CCR scale has a significantly better model fit ($AIC = 1874.23$) than the Gratch Rapport Scale ($AIC = 1914.33$). This observation is reinforced by the Nagelkerke Pseudo R^2_N metric, with our CCR scale achieving a better model fit ($R^2_N = 0.39$) than the Gratch Rapport Scale ($R^2_N = 0.35$). These metrics, AIC and Nagelkerke Pseudo R^2_N, demonstrate that our CCR scale is more effective at measuring rapport than the Gratch Rapport Scale.

[1]OSF link to view Study 1 videos: https://osf.io/5ezga/overview?view_only=f857949cfe1e49dd8f75ce1aac206b9f.
[2]OSF link to view Study 2 videos: https://osf.io/5ezga/overview?view_only=f857949cfe1e49dd8f75ce1aac206b9f.

Table 2. Study 1 EFA Results

Item	Connection	Coordination
Warmth	**1.03**	−0.19
Empathy	**1.00**	−0.17
Friendliness	**0.97**	−0.05
Sympathy	**0.92**	−0.12
Closeness	**0.85**	−0.02
Positivity	**0.84**	0.10
Liking each other	**0.82**	0.13
Enthusiasm	**0.76**	0.02
Deep conversation	**0.74**	−0.22
Respect	**0.74**	0.21
Getting along	**0.72**	0.25
Excitement	**0.71**	0.06
Connection	**0.63**	0.23
Coordination	−0.17	**0.97**
Focus	−0.19	**0.95**
Attentiveness	−0.03	**0.82**
Smooth flow	0.05	**0.79**
Equal participation	−0.15	**0.65**
Engagement	0.13	**0.60**
Enjoyment	0.64	0.32
Agreement	0.61	0.35
Understanding	0.58	0.35
Trust	0.56	0.37
Comfortable with each other	0.52	0.42
Satisfaction	0.44	0.50
Harmony	0.42	0.52
Cooperation	0.42	0.54

Bolded values indicate factor loadings above or equal to 0.6 for scale items with no cross-loadings above 0.3.

2.4 Scale Validation (Study 3)

We conducted Study 3 to validate our 18-item CCR scale from a first-person perspective in an in-person HRI study. We decided to replicate a study from Birnbaum et al. [18] who investigated the influence of a robot's responsiveness on a HRI. We hypothesized that people would perceive greater rapport between themselves and the robot if the robot was programmed to be responsive as opposed to unresponsive. As illustrated in Figure 1(a), participants ($N = 44$) were told to disclose a current problem to either a responsive robot (i.e., one that nods its head and says one template-based sentence that was personalized and positive to the participant) or an unresponsive robot (i.e., one that does not perform any gestures, instead simply giving a one-sentence utterance that asked participants to move on). After the interaction, participants completed a questionnaire that included our 18-item CCR scale. Similar to Studies 1 and 2, we found our CCR scale has high internal reliability ($\alpha = 0.95$, $\omega_{total} = 0.96$), this time from a first-person perspective. As shown in Figure 1(b), we also found that participants perceived a greater degree of rapport between themselves

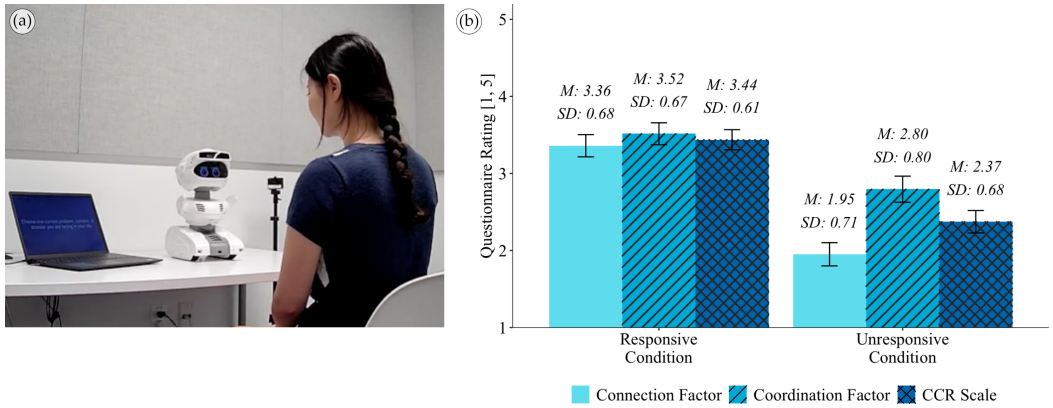

Fig. 1. (a) In Study 3, participants were asked to disclose a personal concern to either a responsive or unresponsive robot. (b) Study 3 participants' responses for the connection factor, the coordination factor, and the CCR Scale between the responsive condition and the unresponsive condition. Error bars show one standard error from the mean.

and the robot when the robot was responsive (M = 3.44, SD = 0.61) as opposed to unresponsive (M = 2.37, SD = 0.68, $p < 0.001$). This finding confirms our CCR scale's construct validity, supports the connection between robot responsiveness and rapport, and shows its applicability to in-person usage.

2.5 Strengths and Limitations of the Full-Length CCR Scale

The full-length 18-item CCR Scale has several strengths. For example, it enables a comprehensive measurement of the construct of rapport, can detect subtle changes in the ratings of rapport, and shows strong internal reliability. However, administering the full-length CCR scale has some limitations. Asking participants to rate 18 scale items may take a relatively longer time, which could potentially increase participants' fatigue [25, 29, 32, 45, 62, 98]. Due to its length, the full-length CCR scale may also be less practical for other researchers who are researching rapport-adjacent topics to include or for other researchers who are conducting repeated measurements of rapport to adopt [30]. For instance, in Study 2, we noticed 49 out of 250 participants (19.60%) failed one or more attention checks when we asked them to rate the full-length 18-item CCR scale four times in one session. This heightened attention check failure rate could be a result of participant fatigue. To address these limitations, this work seeks to reduce the full-length CCR scale to a shorter version that still enables effective measurement of rapport.

3 Scale Reduction Approach

Our approach to developing and validating a reduced-length version of the CCR Scale was primarily informed by the guidelines for (1) short-form development and (2) short-form validation detailed by Smith et al. [89].

3.1 Short-Form Development

Here, we list "A Priori Steps in Rigorous Short-Form Development" from Smith et al. [89], which provide guidance in how researchers can best go about determining which items should be included in a short-form scale. Each item text is identical to the corresponding item listed in Smith et al.

[89]. We have added descriptive titles to each step and alphanumeric values (A1–A7) to which we can later refer to each step in the article.

(A1) *Validate Parent Measure:* Ensure that the parent measure has, itself, been sufficiently validated for the intended purpose.
(A2) *Clarify Use and Choose Items:* Clarify the intended use of the short form (e.g., screening measure vs. diagnosis measure) and choose items or observation classes that meet the goal.
(A3) *Estimate Reliability:* Compute an a priori estimate of the short form's reliability.
(A4) *Estimate Parent Overlap:* Compute an a priori estimate of the likely overlap between the short form and its parent.
(A5) *Estimate Validity Correlations:* Compute an a priori estimate of validity correlations of your short form with key criteria.
(A6) *Estimate Classification Accuracy:* Where appropriate, compute an a priori estimate of the classification accuracy (whether hit rates or elimination of false negatives) of your short form.
(A7) *Estimate Time Saved and Validity Lost:* Compute a priori estimates of the time saved and the validity lost.

In addition to closely following the steps listed above, we also performed content validation as specified by Yusoff [104] to determine the quality of each item within the full-length CCR scale and evaluated the quality of each reduced-length candidate scale as recommended by Cortina et al. [26]. We used these two additional components to choose the items contained within the reduced-length scale as a part of step "(A2) Clarify Use and Choose Items."

Further, our approach to the steps listed above leveraged two data sets: participant item ratings from Study 1 and participant item ratings from Study 2. We used participants' item ratings from Study 1 as a sort of "validation set" to evaluate the quality of each of the candidate reduced-length CCR scales in A2. We then followed the remainder of the steps (A3–A7) using participants' item ratings from Study 2, which we consider our "test set."

3.2 Short-Form Validation

We list each step in "Methodological Steps During Short-Form Validation" from Smith et al. [89], which provides guidance for validating a short-form version of a scale after its items have been chosen. Each list item below is identical to the corresponding item listed in Smith et al. [89]. We have added descriptive titles to each step and alphanumeric values (B1–B9), which we refer to later in the article.

(B1) *Show Time Saved and Validity Lost:* Show time or resource savings and their relation to loss of validity, empirically.
(B2) *Conduct Content Analysis:* Conduct and describe content analyses of each factor in the measure to preserve as much content coverage as possible. Describe clearly any decisions to eliminate a content domain from the parent measure.
(B3) *Administer to a New Sample:* Administer the short form on an independent sample, to enact Steps B4 through B8.
(B4) *Show Reliability:* Show that each factor meets reasonable reliability standards.
(B5) *Calculate Parent Overlap:* Calculate an estimate of overlapping variance between the two versions.[3]

[3]We removed the following clarification text from step B5 as originally listed by Smith et al. [89] to improve readability of this step: "As noted above, extracting the short form from the full form during one administration will overestimate the

(B6) *Demonstrate Factor Structure:* Demonstrate the factor structure or dimensionality of the short form. Make readers aware of any significant differences between the factor structures or dimensions of the short- and full-length versions.
(B7) *Validate Short Form:* Validate the short form. By definition, it is harder to validate a short measure, and the validity evidence for the full-length measure cannot be presumed. It is important to validate the short form in the form it will be used, rather than by extracting its items or observations from the full-length assessment.
(B8) *Assess Classification Accuracy:* Where appropriate, assess the short form's classification rates.
(B9) *Justify Omitted Subfacets:* If subfacets that are included in the full-length form are omitted and only aggregate factors are kept, then content analyses must be conducted to show preservation of the meaning of the aggregate factors or to show that the aggregate factors represent a more narrow construct domain.

4 Development of the Reduced-Length CCR Scale (Study 4)

To develop a proposed reduced-length version of the CCR scale, in this section, we detail how we followed the steps from A1 to A7 detailed in the prior section (Section 3.1—Short-Form Development). After confirming the validity of the parent measure (A1), we developed a reduced-length scale (A2) by conducting a human-subjects study (Study 4), where we asked HRI experts to rate each item in the full-length CCR scale to determine which are the most fundamental to the construct of rapport. Guided by these ratings, we developed a candidate reduced-length CCR scale and estimated its potential using data we had previously collected when validating the initial full-length CCR scale (A3–A7). Study 4 was approved by the University of Chicago's Institutional Review Board (#IRB25-0199). We have made the data used in this section and the code we used to analyze them available in this **Open Science Framework (OSF)** repository.[4]

4.1 Reduced-Length CCR Scale Development

To develop the reduced-length CCR scale, we describe in this section how we completed the steps A1 through A7 from "A Priori Steps in Rigorous Short-Form Development" by Smith et al. [89] detailed in Section 3.1.

4.1.1 (A1) Validate Parent Measure. The parent measure for the reduced-length CCR scale is the full-length CCR scale, whose construction, evaluation, and validation are thoroughly detailed in Section 2 and Lin et al. [60]. In summary, the full-length CCR scale has high internal reliability, high construct validity, and a validated two-factor structure. It outperforms the Gratch Rapport Scale [37] in measuring rapport and was evaluated in both online and in-person human subjects studies from both first-person and third-person perspectives. Thus, the parent measure has been validated.

4.1.2 (A2) Clarify Use and Choose Items. We first clarify our intended use of the reduced-length CCR scale with the following three objectives. The reduced-length CCR scale should:

(1) Preserve the two-factor structure (factors: connection and coordination),
(2) Contain a similar number of items within each factor, and
(3) Save time for participants who complete the questionnaire.

To select which items to include in the reduced-length version of the CCR scale, we started by evaluating the quality of each of the 18 items within the scale. By understanding each item's quality,

overlap, which could lead to unanticipated validity problems down the road. Correlating the short form with the unused, remaining items from the full form underestimates the overlap and is not recommended."
[4]OSF link to view Study 4 data and code: https://osf.io/nr49m/overview?view_only=b1b8730560944d068bae7e3b8745ca42.

we can identify which items are stronger or weaker than others and select candidate items to keep. To assess item quality, we performed the content validation procedure from Yusoff [104]. Content validation is used to assess whether a scale item is relevant and representative of the construct intended to be measured. This procedure generally involves asking a group of experts to review and score the relevance and representativeness of the items to the target construct and calculating the experts' scores (i.e., content validity index) to determine which items to keep and which ones to remove [104].

Therefore, we conducted a human subjects study ($N = 30$) where we asked experts in the field of HRI to rate the 18 scale items from the full-length CCR scale to determine which are the most fundamental to the construct of rapport. We recruited our HRI expert participants by e-mailing all program committee members of the 2025 ACM/IEEE International Conference on HRI with a request to take our online questionnaire. After filling out an online consent form, we provided participants with the definition of rapport: "A feeling of mutual understanding and interpersonal connection among individuals developed through interactions." For each scale item, the experts were asked to rate four options ranging from 1 (Terrible) to 4 (Fundamental):

1—*Terrible* (the item is definitely not part of the domain)
2—*Less Appropriate* (the item is probably not part of the domain)
3—*Acceptable* (the item is a clear part of the domain)
4—*Fundamental* (the item is highly relevant and a core part of the domain)

These participants were not provided with any form of compensation. From participants' responses, we first followed the content validation procedure from Yusoff [104] by recoding *3—Acceptable* and *4—Fundamental* as relevance ratings of 1 and recoding *1—Terrible* and *2—Less Appropriate* as relevance ratings of 0. We then calculated the **Item-Level Content Validity Index (I-CVI)** using the following formula:

$$\text{I-CVI} = \left(\frac{\text{\# of Experts with a Relevance Rating for the Item}}{\text{Total \# of Experts}} \right).$$

# of Experts with a Relevance Rating for the Item refers to the number of experts with a relevance rating of 1 for the item, and *Total # of Experts* is the total number of experts surveyed ($N = 30$) in the formula. The calculated I-CVI for each item is shown in Table 3. I-CVI ranges from 0 to 1, with a higher value showing that a scale item is more important to measure rapport, as indicated by the experts. Yusoff [104] suggests keeping scale items with an I-CVI value higher than 0.78 when more than nine experts have provided their ratings. Using this threshold, 11 of our 18 items can be considered to have high content validity (7 in the connection factor and 4 in the coordination factor).

Given these 11 items and our goal of balancing the number of items within each factor, we decided to explore candidate reduced-length CCR scales with 4 items in the connection factor and 4 items in the coordination factor. We thus reviewed a total of 35 candidate reduced-length scales, where each of them contains the 4 items in the coordination factor and one of the 35 combinations of 4 items selected from the 7 items in the connection factor. To estimate the validity of each of the possible candidate reduced-length scales, we followed the recommendations of Cortina et al. [26], who outlines many possible psychometric indicators that researchers should attend to when designing a shortened scale (e.g., internal consistency reliability, part-total correlations, general factor loadings, convergent and discriminant validity, content coverage of the items). Therefore, we performed initial validity checks of the reduced-length scale using Study 1 data (see Section 2.2) on all possible reduced-length scales [26]. We found that all candidate reduced-length scales show relatively high internal reliability (e.g., Cronbach's alpha, McDonald's omega, Guttman's

Table 3. I-CVI for Each Item Based on Expert Ratings on the Data We Collected in Study 4

Scale Item	Factor	Experts in Agreement	I-CVI
Connection	Connection	30/30	1.00
Getting along	Connection	30/30	1.00
Coordination	Coordination	29/30	0.97
Empathy	Connection	28/30	0.93
Respect	Connection	28/30	0.93
Closeness	Connection	27/30	0.90
Engagement	Coordination	26/30	0.87
Smooth flow	Coordination	26/30	0.87
Attentiveness	Coordination	25/30	0.83
Friendliness	Connection	24/30	0.80
Liking each other	Connection	24/30	0.80
Sympathy	Connection	23/30	0.77
Warmth	Connection	23/30	0.77
Positivity	Connection	21/30	0.70
Enthusiasm	Connection	18/30	0.60
Focus	Coordination	14/30	0.47
Equal participation	Coordination	14/30	0.47
Excitement	Connection	12/30	0.40

The horizontal line between "Liking each other" and "Sympathy" indicates the I-CVI cutoff (> 0.78) suggested by Yusoff [104]. The items we selected for the reduced-length CCR scale are shown in bold.

lambda-2), high correlation with the full-length scale, and similar inter-item correlation with the full-length scale. Since all candidate scales satisfied the indicators recommended by Cortina et al. [26], we decided to select the four items from each factor with the highest content validity index (I-CVI), resulting in a reduced-length scale with the items "Connection," "Getting Along," "Empathy," and "Respect" for the connection factor and "Coordination," "Engagement," "Smooth Flow," and "Attentiveness" for the coordination factor (these items are bold in Table 3).

Here, we detail the strength of this proposed reduced-length CCR scale on the indicators recommended by Cortina et al. [26] on the data from Study 1. First, we found the reduced-length CCR scale to have high internal reliability ($\alpha = 0.92$, $\omega_{total} = 0.95$, $\lambda_2 = 0.92$). Both factors' reliability was high as well (connection factor: $\alpha = 0.93$, $\omega_{total} = 0.94$; coordination factor: $\alpha = 0.87$, $\omega_{total} = 0.90$). These results show that the reduced-length 8-item CCR scale preserves internal reliability and serves as a candidate shorter alternative to the full-length scale. We also found that the reduced-length 8-item CCR scale has a strong correlation with the full-length 18-item CCR scale ($R = 0.99, p < 0.001$). Given that the coefficient of determination (R^2) captures overlapping variance, we discovered that the reduced-length scale captures 98% ($R^2 * 100\% = 0.99^2 * 100\% \approx 98\%$) of variability in the full-length scale. This strong overlap indicates that participants would likely rate rapport similarly on both scales, which means that the reduced-length scale preserves the measurability of rapport from the full-length scale.

Separately for the reduced-length scale and full-length scale, we examined their inter-item correlation, which measures the breadth of the construct that a scale is measuring. Scale items that measure a broad construct will correlate less highly with each other than will those that reflect a

narrow construct. If the reduced-length scale is measuring the same construct as the full-length scale, it should have a similar inter-item correlation compared to the full-length version of the scale. We found that the inter-item correlation of the reduced-length 8-item CCR scale ($\bar{r} = 0.59$) is similar to that of the full-length 18-item CCR scale ($\bar{r} = 0.59$). This observation ensures that the reduced-length scale does not narrow (or broaden) the measurement ability of the full-length scale.

4.1.3 (A3) Estimate Reliability. Using Study 2 data (see Section 2.3), we estimated a strong internal reliability of the reduced-length scale ($\alpha = 0.91$, $\omega_{total} = 0.94$, $\lambda_2 = 0.90$). We also estimated that both factors maintain their reliability: connection factor ($\alpha = 0.87$, $\omega_{total} = 0.92$) and coordination factor ($\alpha = 0.85$, $\omega_{total} = 0.87$). This indicates that the reduced-length 8-item CCR scale is likely to maintain high internal consistency.

4.1.4 (A4) Estimate Parent Overlap. From the data from Study 2, we also estimated a highly positive correlation between the reduced-length and full-length CCR scales ($R = 0.98$, $p < 0.001$). The reduced-length scale captures an estimated 96% ($R^2 = 0.96$) of variability in the full-length scale. Furthermore, the reduced-length 8-item CCR scale has a similar inter-item correlation ($\bar{r} = 0.56$) when compared to the full-length CCR scale ($\bar{r} = 0.60$) with a difference of 0.04. This indicates that the reduced-length scale will preserve the measurability of rapport from the full-length scale.

4.1.5 (A5) Estimate Validity Correlations. From Study 2 data, we also examined the convergent validity of the reduced-length scale by confirming whether its measure correlates with another scale that measures rapport. We estimated that the reduced-length scale will have a high correlation with the Gratch Rapport Scale (convergent scale) ($R = 0.82$, $p < 0.001$), which points to a high convergent validity.

4.1.6 (A6) Estimate Classification Accuracy. Since our reduced-length 8-item CCR scale is a Likert scale that is not used for classification of distinct categories, this step does not apply to our specific case.

4.1.7 (A7) Estimate Time Saved and Validity Lost. We calculated the estimated time saved using the following formula:

$$\text{Estimate Time Saved} = \left(1 - \frac{\#\text{ of Items in the Reduced-Length Scale}}{\#\text{ of Items in the Full-Length Scale}}\right) * 100\%.$$

Given that the reduced-length scale has 8 items as opposed to 18 items in the full-length scale, it is estimated to have time savings of 55.56%.

From examining Study 2 data using A4, the reduced-length scale captures 96% of the variability in the full-length scale. This finding indicates that the estimated validity lost is 4% (100−96%).

4.2 Reduced-Length CCR Scale Development Discussion

We followed the steps from A1 to A7 in Section 3.1 to develop the reduced-length CCR scale. From a priori analysis of this reduced-length scale using our existing data (Studies 1 and 2), this scale has demonstrated strong estimated internal reliability, a high correlation with the full-length scale, and decent convergent validity. To ensure that these psychometric properties are generalizable beyond our current data, additional validation is needed for the reduced-length CCR scale using an independent sample, carried out in the next section.

5 Validation of the Reduced-Length CCR Scale (Study 5)

Equipped with a short-form version of the CCR scale, we next sought to validate it following the steps B1 to B9 detailed in Section 3.2—Short-Form Validation. This involved administering the short

form to a new sample (B3); therefore, we conducted a human-subjects study (Study 5), where a new sample of participants ($N = 186$) watched videos of HRIs with varied degrees of human–robot rapport and completed a series of questionnaires. Using these data, we were able to assess the validity and time saved of the reduced-length CCR scale detailed in the remaining applicable steps (B1–B2 and B4–B9). This study was approved by the University of Chicago's Institutional Review Board (#IRB25-0199), and the materials (video stimuli), data, and data analysis code are available in our OSF repository.[5]

5.1 Experiment Design

In our prior work, Lin et al. [60], we conducted an in-person human-subjects study (Study 3) replicating prior work (Study 1 in Birnbaum et al. [18]), where participants were asked to disclose a current problem, concern, or stressor to the robot. These works demonstrated that a responsive robot was perceived more positively [18, 60], participants had a greater desire for companionship with the robot [18, 60], and participants perceived higher human–robot rapport [60].

While Birnbaum et al. [18] and Lin et al. [60] demonstrated a robust effect of a robot's responsiveness on the human–robot relationship, it was conducted in one specific and narrow context: a participant disclosing a stressor to a robot. *In this new study (Study 5), our goal was twofold: (1) to validate the reduced-length CCR scale and also (2) to explore whether a robot's responsive behavior would also have a significantly positive impact on human–robot rapport across a variety of high-impact HRI contexts.*

We, therefore, conducted an experiment with two between-subjects conditions: a *responsive condition* and an *unresponsive condition*. While these two conditions are similar to those in Birnbaum et al. [18] and Lin et al. [60], they differed in the HRI context. For each of the two experimental conditions, we recorded several videos representing a variety of different interaction contexts and recruited online participants to watch these videos and fill out questionnaires about them. For the videos in the *responsive condition*, the robot nodded its head and responded with a more personalized message to the person, replicating the responsive robot in Birnbaum et al [18] and Study 3 (see Section 2.4). For example, after the person in the video says "*I have an exam coming up tomorrow, and I'm so stressed,*" the responsive robot nods and responds with "*That sounds really overwhelming. Do you want to tell me more about what's making you feel so stressed?*" (Figure 2). For the videos in the *unresponsive condition*, the robot did not perform any responsive gestures (similar to the unresponsive robot in Birnbaum et al [18] and Study 3 detailed in Section 2.4) and gave a short and delayed response. For instance, after the person in the video says "*I have an exam coming up tomorrow, and I'm so stressed,*" the unresponsive robot takes a significant pause and then says "*What is the source of your stress?*" (Figure 2).

In order to select the interaction contexts for these videos, we examined the HRI contexts (e.g., education, transportation, home) represented in full papers published in the ACM/IEEE International Conference on HRI from 2020 to 2024. By examining the titles and keywords of these articles, we categorized the interaction context in each of these papers as belonging to one of the following categories: education ($N = 30$), healthcare ($N = 29$), collaboration ($N = 24$), service ($N = 22$), assistive ($N = 13$), entertainment ($N = 7$), transportation ($N = 7$), home ($N = 3$), and industrial ($N = 1$). We excluded from our analysis any papers ($N = 180$) whose focus did not include a specific context, including context-independent review and theory papers and papers whose central contribution focused on context-independent topics such as machine learning models for

[5] OSF link to view Study 5 videos, data, and code: https://osf.io/nr49m/overview?view_only=b1b8730560944d068bae7e3b8745ca42.

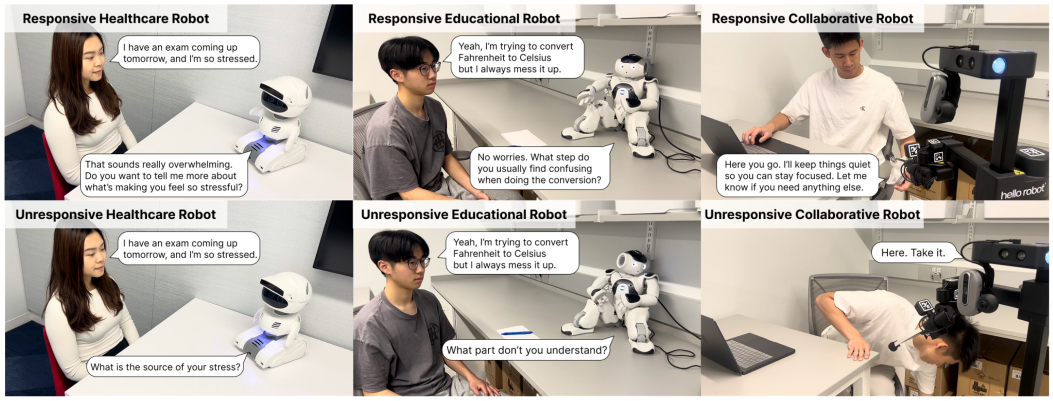

Fig. 2. For Study 5, we conducted a between-subjects study with two experimental conditions (responsive vs. unresponsive) where online participants viewed a video of a HRI in one of three possible interaction settings. The healthcare setting featured a student asking a Misty II robot for help with de-stressing. The education setting involved a student asking a NAO robot for help solving a temperature conversion problem. The collaboration setting featured a student asking for a snack from a Stretch 3 robot.

robots, robot control, and robot ethics. The three interaction contexts with the largest numbers of publications were *healthcare* (e.g., [13, 20, 24, 49, 67]), *education* (e.g., [22, 42, 46, 59, 78]), and *collaboration* (e.g., [17, 43, 94, 95, 99]). We recorded the complete list of full-paper publications in each context in Supplemental Materials Table SIII.

Therefore, we filmed three videos for each of our experimental conditions representing these three different interaction contexts (healthcare, education, and collaboration). Including these three distinct interaction contexts allows us to expand the coverage of our CCR scale, enabling generalization across multiple scenarios rather than narrowly focusing on just one. In our analysis, we grouped data from the three interaction contexts in the same condition and then compared rapport ratings from the reduced-length CCR scale between the two conditions (responsive vs. unresponsive). In this work, we did not compare the rapport ratings across the three different interaction contexts since each video showcases a different robot that demonstrates different responsive/unresponsive behaviors to the human actor.

5.2 Materials (Videos)

We filmed six short videos featuring interactions between one robot and one human to represent each possible setting (healthcare, education, collaboration) in each of our two experimental conditions (responsive condition, unresponsive condition). A description of the HRI in each video is listed below and is also shown in Figure 2.

- *Responsive Healthcare Robot*: A Misty II robot talks with a stressed student about their anxiety about an upcoming important exam with a warm and empathetic communication style with little to no delays in speech.
- *Unresponsive Healthcare Robot*: A Misty II robot talks with a stressed student about their anxiety about an upcoming important exam with a task-focused and brief communication style with larger delays between human and robot speech.
- *Responsive Educational Robot*: A NAO robot helps a student convert from Fahrenheit to Celsius with a warm and empathetic communication style while holding eye contact with the student.

— *Unresponsive Educational Robot*: A NAO robot helps a student convert from Fahrenheit to Celsius with a task-focused and brief communication style and looks away, not holding any eye contact with the student.
— *Responsive Collaborative Robot*: A Stretch 3 robot successfully delivers a requested snack to a student with a warm and empathetic communication style without interrupting the student.
— *Unresponsive Collaborative Robot*: A Stretch 3 robot unsuccessfully delivers a requested snack to a student with a task-focused and brief communication style and interrupts the student multiple times.

We filmed each video to be approximately 60 seconds in length to minimize participant fatigue while still sufficiently displaying the responsive (or unresponsive) nature of the robot. We also selected a different robot for each setting to allow this study's results to better generalize to a wider array of different robot embodiments. Similarly, we showed different methods of being unresponsive in each of the unresponsive videos (e.g., delaying a response or not looking at the person). We have made all six of these videos available through our OSF repository.[6]

5.3 Procedure

After consenting to join the study, online participants were randomly assigned to an experimental condition and watched one of the six videos detailed in Section 5.2. They then rated the video with the following scales:

(1) The full-length 18-item CCR scale [60] (see Table 1 and Box 1 for its detailed usage).
(2) The reduced-length 8-item CCR scale (see Table 1 and Box 2 for its detailed usage).
(3) The Gratch Rapport Scale [37] as our convergent scale (see Supplemental Materials Table SII).
(4) The Perceived Danger Scale [73] as our divergent scale (see Supplemental Materials Table SIV).

We administered these scales to participants in order to assess the reduced-length CCR scale's convergent validity and divergent validity. We evaluate convergent validity by comparing participant responses on the reduced-length CCR scale with those on the full-length CCR scale and a convergent scale (the Gratch Rapport Scale [37]) to ensure that the reduced-length CCR scale is capturing the same construct. Additionally, we evaluate divergent validity by comparing participant responses on the reduced-length CCR scale with those on a divergent scale (the Perceived Danger Scale [73]), for which rapport was not expected to be similar. The Perceived Danger Scale [73] is a rigorously validated scale that measures a person's anticipation of harmful consequences of an interaction with another entity (e.g., robot). It includes items such as "How menacing was the robot?" and "How severely might you be injured?" on a six-point scale from 1 (Not at all) to 6 (Extremely). As opposed to selecting a construct more similar to rapport (e.g., competence, anthropomorphism), our choice of the Perceived Danger Scale [73] gives us the best possible chance of demonstrating the lack of correlation between rapport and the scale chosen to demonstrate divergent validity—perceived danger.

Participants were randomly assigned to complete the four scales in one of the following orders where the convergent and divergent scales were always rated in between the reduced and full scales: (1)-(3)-(4)-(2), (1)-(4)-(3)-(2), (2)-(3)-(4)-(1), (2)-(4)-(3)-(1). These scale orders were created to ensure as much separation as possible between the full-length and the reduced-length CCR scales so that the ratings of the second scale are not directly affected by the first scale [89]. By inserting two different scales between them, we minimized the likelihood of these carryover effects.

[6]OSF link to view Study 5 videos: https://osf.io/nr49m/overview?view_only=b1b8730560944d068bae7e3b8745ca42.

Table 4. The Counts of Participants Who Watched Each Video in the Two Conditions

Responsive Condition	# of Participants	Unresponsive Condition	# of Participants
Responsive Healthcare Robot	35	Unresponsive Healthcare Robot	31
Responsive Educational Robot	30	Unresponsive Educational Robot	30
Responsive Collaborative Robot	29	Unresponsive Collaborative Robot	31
Total Number of Participants	94	Total Number of Participants	92

Additionally, the scale items within each of the four scales were also randomized in the study. At the end of the study, participants were also asked to share their demographic information, such as gender, ethnicity, and age. Participants took an average of 7.20 minutes (SD = 4.29 mins) to complete the study and were compensated with $1.50 USD.

5.4 Participants

We computed the required sample size of participants a priori using a power analysis with a power of 0.95 and an effect size d of 0.5 at $p < 0.05$, which suggested 176 participants (88 for each of the two conditions). We thus recruited a total of 200 participants from the Prolific platform who were native English speakers in the United States and had an approval rating higher than 99. We removed the data from 14 participants because they missed one or more attention checks. The remaining 186 participants include 141 White, 29 Black or African American, 6 Asian, 2 other ethnicity, and 8 identified as two or more ethnicities. The ages of the participants range from 19 to 81 ($M = 39.99$, SD = 12.83). Ninety-five participants identified as men, 87 as women, and 4 as non-binary. Ninety-four participants were in the responsive condition, and 92 participants were in the unresponsive condition. The counts for the number of participants who watched which video in the two conditions are summarized in Table 4.

5.5 Reduced-Length CCR Scale Validation

To validate the reduced-length CCR scale, we describe in this section how we completed the steps from B1 to B9 from "Methodological Steps During Short-Form Validation" by Smith et al. [89] detailed in Section 3.2.

5.5.1 (B1) Show Time Saved and Validity Lost. As shown in Figure 3, we found that participants spent significantly less time completing the reduced-length 8-item CCR scale ($M = 29.69s$, SD = 24.19s) compared to the full-length CCR scale ($M = 51.96s$, SD = 38.39s, $t_{(185)} = -7.67$, $d = 0.69$, $p < 0.001$). The average time saved from the reduced-length CCR scale is 22.27 seconds, which is 42.86% of the time spent on the full-length CCR scale. It is noteworthy that participants also spent significantly less time on the reduced-length CCR scale compared to the Gratch Rapport Scale ($M = 65.90s$, SD = 60.28s, $t_{(185)} = -9.78$, $d = 0.79$, $p < 0.001$; Figure 3), whose prior version [38] is the most cited rapport scale on Google scholar before the creation of the full-length CCR scale.

To examine the validity lost by shortening the scale, we examined the correlation between the reduced-length CCR scale and the full-length CCR scale and found a highly positive correlation ($R = 0.88$, $p < 0.001$). Based on the coefficient of determination ($R^2 = 0.77$), the reduced-length scale captures 77% of the variability in the full-length scale, which suggests that the loss of validity is 23%. While there are no strict guidelines for how good or bad 23% validity loss is, other well-accepted shortened scales have lost even more and are considered acceptable. For example, Rammstedt and John [82] created a 10-item shortened scale of the 44-item **Big Five Personality Inventory (BFI)** [50]. The shortened BFI scale had a loss of validity of approximately 30% and is widely used today. Given this 23% validity loss in our reduced-length CCR scale, we have confidence that the

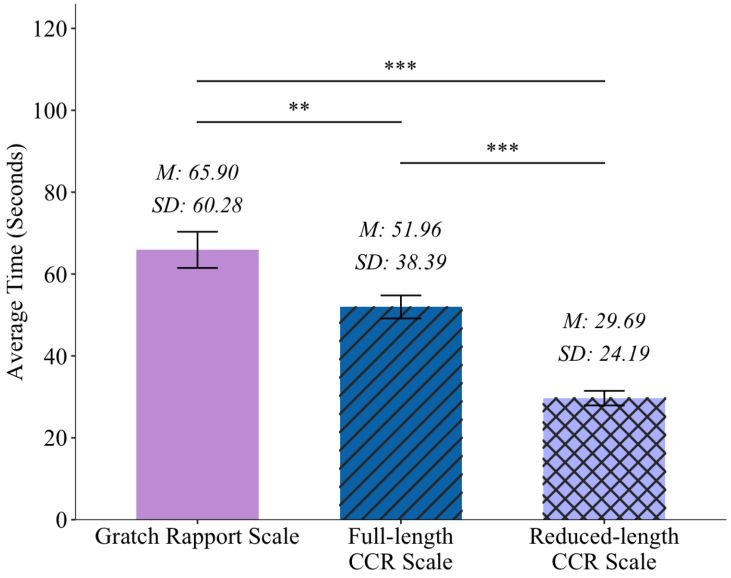

Fig. 3. Participants spent significantly less time on the reduced-length CCR scale compared to both the full-length CCR scale and the Gratch Rapport Scale. They also spent significantly less time on the full-length CCR scale compared to the Gratch Rapport Scale. (**), (***) denote $p < 0.01$, and $p < 0.001$, respectively. Error bars show one standard error from the mean.

reduced-length CCR scale relates strongly enough to the full-length scale to warrant confidence in its usage. However, for research projects where rapport is a primary focus, we recommend using the full-length CCR scale, which provides a more comprehensive assessment of the two measures. We further discuss the tradeoff between using the reduced-length CCR scale and the full-length CCR in Section 6.3.

5.5.2 (B2) Conduct Content Analysis. Using the data we collected from HRI experts in Study 4 detailed in Section 4, we performed content validation on the items in the full-length scale and kept a total of eight items in the reduced-length scale (four items in each factor) that were rated as fundamental to the concept of rapport by the HRI experts. Since we retained both factors in the reduced-length scale and our parent overlap (as detailed in step B5) is 77%, we are confident that, even with the reduction in the number of scale items, our reduced-length CCR scale still ensures sufficient content coverage of both the connection and coordination factors.

5.5.3 (B3) Administer to a New Sample. We ran Study 5 ($N = 186$) in order to administer the reduced-length CCR scale on an independent sample, as well as the full-length CCR scale, convergent scale, and divergent scale to assess the validity of the reduced-length CCR scale.

5.5.4 (B4) Show Reliability. We found that the reduced-length CCR scale has high internal reliability ($\alpha = 0.91$, $\omega_{total} = 0.94$, $\lambda_2 = 0.91$). The two factors in the reduced-length scale are also highly reliable: connection ($\alpha = 0.84$, $\omega_{total} = 0.88$) and coordination ($\alpha = 0.86$, $\omega_{total} = 0.89$). These values of reliability are all considered excellent; typical suggestions are for at least 0.70 or 0.80 [19, 85].

5.5.5 (B5) Calculate Parent Overlap. We found that the reduced-length CCR scale has a strong correlation with its parent: the full-length CCR scale ($R = 0.88$, $p < 0.001$). The reduced-length

Table 5. Means, Standard Deviations, *T*-Test Statistics, p Values, and Effect Sizes of the Measures in Study 5

	Responsive Robot		Unresponsive Robot		$t_{(184)}$	p	Cohen's d
	M	SD	M	SD			
Reduced-length CCR Scale	3.84	0.72	3.41	0.93	3.52	$< 0.001^{***}$	0.51
Reduced-length Connection Factor	3.74	0.81	3.29	0.97	3.49	$< 0.001^{***}$	0.51
Reduced-length Coordination Factor	3.94	0.72	3.54	0.99	3.13	$< 0.002^{**}$	0.46
Full-length CCR Scale	3.66	0.83	3.25	0.98	3.12	0.002^{**}	0.46
Full-length Connection Factor	3.46	0.93	3.03	1.07	2.94	$< 0.004^{**}$	0.43
Full-length Coordination Factor	3.86	0.80	3.46	0.97	3.06	0.003^{**}	0.45
Gratch Rapport Scale (Convergent Scale)	3.66	0.69	3.23	0.80	3.90	$< 0.001^{***}$	0.57
Perceived Danger Scale (Divergent Scale)	2.17	1.00	2.19	0.91	−0.10	0.917	−0.02

(**) denotes $p < 0.01$ and (***) denotes $p < 0.001$.

CCR scale also captures 77% ($R^2 = 77\%$) of variability in the full-length CCR scale. Additionally, the reduced-length scale has a very similar inter-item correlation ($\bar{r} = 0.57$) compared to the full-length CCR scale ($\bar{r} = 0.62$). Because the inter-item correlations are so similar, we can assume that the coverage of the scale is similar, as inter-item correlation reflects how broadly or narrowly the items measure the construct. When a scale is shortened, the goal is to cover a comparable breadth: the small difference between the two scales suggests that we succeeded.

5.5.6 (B6) Demonstrate Factor Structure. We also performed CFA on Study 5 data and found the reduced-length CCR scale has a good fit based on CFI, TLI, and SRMR: $CFI = 0.996$ (≥ 0.95), $TLI = 0.995$ (≥ 0.95), and $SRMR = 0.050$ (≤ 0.08). In the meantime, the RMSEA only indicates a moderate fit [65]: $RMSEA = 0.088$ (≤ 0.08). This result provides additional evidence that the reduced-length CCR scale maintains the clear two-factor structure of the full-length CCR scale.

5.5.7 (B7) Validate Short Form. We validated the reduced-length CCR scale by examining its convergent validity and divergent validity. We found that the reduced-length CCR scale has a strong correlation with the Gratch Rapport Scale ($R = 0.79$, $p < 0.001$), which supports its convergent validity. We also found that the reduced-length CCR scale has a weak correlation with the Perceived Danger Scale ($R = -0.05$, $p = 0.492$), confirming its divergent validity. These results suggest that the reduced-length CCR scale is capturing the same construct (rapport) as other scales seeking to measure the same construct (convergent validity) but not being so broad that it measures everything associated with HRI (divergent validity).

After collapsing data from the three interaction contexts into a single group, we used *t*-tests to examine the ratings of the reduced-length CCR scale between the responsive and unresponsive experimental conditions to assess the reduced-length CCR scale's known-groups validity. As shown in Figure 4 and Table 5, we found that participants in the responsive condition ($M = 3.84$, SD = 0.72) perceived significantly greater rapport than those in the unresponsive condition ($M = 3.41$, SD = 0.93, $t_{(184)} = 3.52$, $d = 0.51$, $p < 0.001$). In addition to participants perceiving greater rapport with the responsive robot than the unresponsive robot using the reduced-length CCR scale, the same significant effect was also observed using the full-length CCR scale, the individual dimensions of connection and coordination from both the full-length CCR scale and the reduced-length CCR scale, and the convergent scale (see Table 5 for the full statistics). This ability of the reduced-length CCR scale to detect a statistically significant difference in rapport between the responsive and unresponsive conditions (known-groups validity) provides additional support for its strong construct validity.

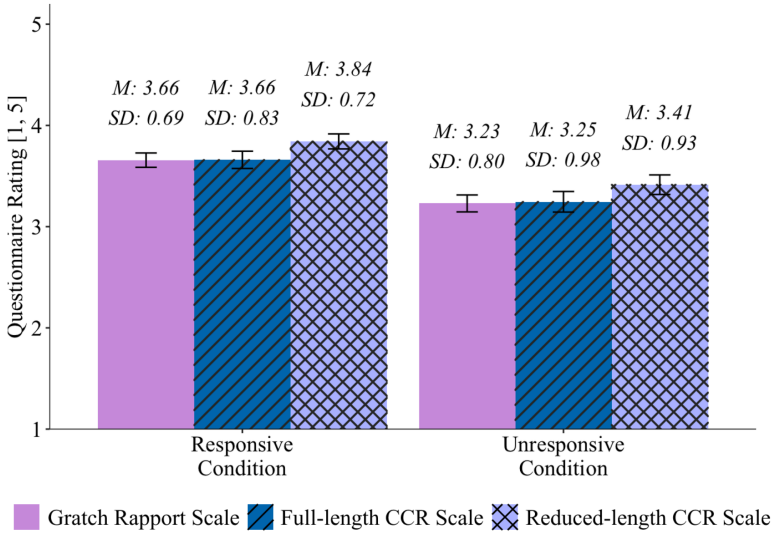

Fig. 4. Participants' responses for the reduced-length CCR scale, the full-length CCR scale, and the Gratch Rapport Scale between the responsive condition and the unresponsive condition. Error bars show one standard error from the mean.

In Figure 4, it is worth noting that participant ratings using the reduced-length CCR scale are slightly higher than both the full-length CCR scale and the Gratch Rapport Scale for both responsive and unresponsive conditions in Study 5. While ratings with the reduced-length CCR scale may be marginally higher, we do not expect these elevated ratings to have an impact on the reduced-length CCR scale's ability to measure rapport, as we were able to demonstrate its known-groups validity through its ability to detect a significant difference in rapport between the two conditions.

5.5.8 (B8) Assess Classification Accuracy. Since our reduced-length CCR scale is a Likert scale that is not used to classify categories, this step does not apply to our scale.

5.5.9 (B9) Justify Omitted Subfacets. Our reduced-length CCR scale does not omit any factors or subfacets from the full-length CCR scale, so this step is not applicable for our purpose.

5.6 Additional Validation of the Full-Length CCR Scale from Study 5 Data

While the primary purpose of Study 5 was to validate the new reduced-length CCR scale, Study 5 data also provide additional evidence (beyond that provided in Studies 1–3) that validates the full-length CCR scale:

- *Reliability:* The full-length CCR scale demonstrates high internal reliability ($\alpha = 0.97$, $\omega_{total} = 0.97$, $\lambda_2 = 0.97$) as well as each of its two factors individually: connection factor ($\alpha = 0.96$, $\omega_{total} = 0.97$) and coordination factor ($\alpha = 0.90$, $\omega_{total} = 0.94$).
- *Construct Validity:* The full-length CCR scale shows convergent validity through a strong positive correlation with the Gratch Rapport Scale ($R = 0.81$, $p < 0.001$), divergent validity through a weak correlation with the Perceived Danger Scale ($R = -0.05$, $p = 0.537$), and known-groups validity demonstrated by participants rating the videos in the responsive condition as significantly higher in rapport using the full-length CCR scale than the videos in the unresponsive condition (visualized in Figure 4 and detailed statistics found in Table 5).

— *Time Savings Compared with the Gratch Rapport Scale:* Participants spent significantly less time on the full-length CCR scale ($M = 51.96s$, SD = 38.39s) compared to the Gratch Rapport Scale ($M = 65.90s$, SD = 60.28s, $t_{(185)} = -2.84$, $d = 0.28$, p = 0.005; Figure 3).

This additional validation of the full-length CCR scale using the data from Study 5 provides strong support for the use of the full-length CCR scale to measure human–robot rapport and its additional time-saving benefits compared to the Gratch Rapport Scale, whose prior version [38] is the most widely cited rapport scale used to measure human–robot rapport before the publication of the full-length CCR scale.

5.7 The Influence of Robot Responsiveness on Human–Robot Rapport across Different Settings

In line with our prior findings from Study 3 in Lin et al. [60] (summarized in Section 2.4), our results from Study 5 provide further evidence that robot responsiveness has a strong positive relationship with human–robot rapport: the greater a robot's responsiveness, the greater the perception of human–robot rapport. As reported in Section 5.5.7 and displayed in Figure 4 and Table 5, participants rated the human–robot rapport significantly higher in the responsive robot condition than the unresponsive robot condition using both the full-length ($t_{(184)} = 3.12$, $d = 0.46$, p = 0.002) and reduced-length ($t_{(184)} = 3.52$, $d = 0.51$, p < 0.001) CCR scales.

It is also worth noting that both the connection factor and the coordination factor contribute to the perceived increase in human–robot rapport as a result of the robot's responsiveness. As shown in Table 5, the connection and coordination factors for both the full-length and reduced-length CCR scales are reported as significantly higher in the responsive robot condition than the unresponsive robot condition. Lin et al. [60] also found the same effect, where both the connection and coordination factors are rated significantly higher after participants engage with a responsive as opposed to an unresponsive robot. These results together indicate that a robot's responsiveness shapes both the perceptions of human–robot connection and human–robot coordination, rather than manipulating just one of these factors, to result in an overall change in the perceived human–robot rapport.

5.8 Reduced-Length CCR Scale Validation Discussion

We adhered to the steps from B1 to B9 in Section 3.2 to validate the reduced-length 8-item CCR scale. We found that the reduced-length scale yields a time savings of 42.86% compared with the full-length scale, shows a clear two-factor structure, and demonstrates construct validity (convergent validity, divergent validity, known-groups validity). This reduced-length scale is also highly reliable and correlates strongly with the full-length scale. Even though the reduced-length CCR scale shows slightly elevated rapport scores compared to the full-length CCR scale and the Gratch Rapport Scale in Study 5 (shown in Figure 4), we do not anticipate that the elevated scores would affect the usability of the reduced-length CCR scale, given that it has already demonstrated strong construct validity in its ability to show a statistically significant difference in rapport between the responsive and unresponsive conditions. While validating the reduced-length CCR scale, we also further validated the full-length CCR scale, which also shows high internal reliability and construct validity as well as time savings compared with the Gratch Rapport Scale. In concert with our prior results of Lin et al. [60], we have also found further strong empirical evidence that increasing a robot's responsiveness also increases perceptions of human–robot rapport, both for rapport overall and for each individual factor (connection and coordination).

Box 1. Full-Length 18-Item CCR Scale Usage

Full-Length 18-Item CCR Scale Administration

Use the following prompt in your questionnaire: "Rate how much you think the following was present in the interaction." Then, have participants rate each of the 18 items below on a five-point Likert scale: (1) Strongly Disagree, (2) Disagree, (3) Neither Agree nor Disagree, (4) Agree, (5) Strongly Agree. The full-length CCR scale can be administered on paper or online. We also recommend randomizing the 18 items and separating them equally into two pages to reduce respondent fatigue (i.e., having 9 items on one page at once).

Full-Length 18-Item CCR Scale

	Strongly Disagree	Disagree	Neither Agree nor Disagree	Agree	Strongly Agree
Warmth	○	○	○	○	○
Empathy	○	○	○	○	○
Friendliness	○	○	○	○	○
Sympathy	○	○	○	○	○
Closeness	○	○	○	○	○
Positivity	○	○	○	○	○
Liking each other	○	○	○	○	○
Enthusiasm	○	○	○	○	○
Respect	○	○	○	○	○
Getting along	○	○	○	○	○
Excitement	○	○	○	○	○
Connection	○	○	○	○	○
Coordination	○	○	○	○	○
Focus	○	○	○	○	○
Attentiveness	○	○	○	○	○
Smooth flow	○	○	○	○	○
Equal participation	○	○	○	○	○
Engagement	○	○	○	○	○

Full-Length 18-Item CCR Scale Scoring

The full-length CCR scale can capture a score for each individual factor (connection and coordination) and rapport as an overall construct:

(1) **Connection Factor:** Average the participant's ratings for the 12 items within the connection factor (Warmth, Empathy, Friendliness, Sympathy, Closeness, Positivity, Liking each other, Enthusiasm, Respect, Getting along, Excitement, Connection).

(2) **Coordination Factor:** Average the participant's ratings for the 6 items within the coordination factor (Coordination, Focus, Attentiveness, Smooth flow, Equal participation, Engagement).

(3) **Rapport:** Average the scores from (1) Connection Factor and (2) Coordination Factor together to get one final score for rapport.

Box 2. Reduced-Length 8-Item CCR Scale Usage

Reduced-Length 8-Item CCR Scale Administration

Use the following prompt in your questionnaire: "Rate how much you think the following was present in the interaction." Then, have participants rate each of the 8 items below on a five-point Likert scale: (1) Strongly Disagree, (2) Disagree, (3) Neither Agree nor Disagree, (4) Agree, (5) Strongly Agree. The reduced-length CCR scale can be administered on paper or online. We recommend randomizing the display order of the items for each participant when administering this scale (e.g., using "Statement Randomization" in Qualtrics).

Reduced-Length 8-Item CCR Scale

	Strongly Disagree	Disagree	Neither Agree nor Disagree	Agree	Strongly Agree
Connection	○	○	○	○	○
Getting along	○	○	○	○	○
Empathy	○	○	○	○	○
Respect	○	○	○	○	○
Coordination	○	○	○	○	○
Engagement	○	○	○	○	○
Smooth flow	○	○	○	○	○
Attentiveness	○	○	○	○	○

Reduced-Length 8-Item CCR Scale Scoring

The reduced-length CCR scale can capture a score for each individual factor (connection and coordination) and rapport as an overall construct:

(1) **Connection Factor:** Average the participant's ratings for the 4 items within the connection factor (Connection, Getting along, Empathy, Respect).
(2) **Coordination Factor:** Average the participant's ratings for the 4 items within the coordination factor (Coordination, Engagement, Smooth flow, Attentiveness).
(3) **Rapport:** Average the scores from (1) Connection Factor and (2) Coordination Factor together to get one final score for rapport.

6 General Discussion

In this work, we describe how we developed and validated a reduced-length version of the CCR scale that contains 8 items with 4 items in each of its two factors (connection and coordination). In this section, we discuss the primary strengths of the reduced-length CCR scale and the process we used to produce it. We also discuss scale-shortening in HRI more broadly, the use of the full-length vs. reduced-length CCR scales, and the connection between robot responsiveness and rapport. The detailed usage of the full-length CCR scale is shown in Box 1, while the detailed usage of the reduced-length CCR scale is illustrated in Box 2.

6.1 The Reduced-Length CCR Scale as a Validated Measure to Capture Human–Robot Rapport

As detailed in Section 4—Development of the Reduced-Length CCR Scale (Study 4) and Section 5—Validation of the Reduced-Length CCR Scale (Study 5), the work that we present in this article provides substantial evidence that the reduced-length CCR scale is a validated metric to measure

human–robot rapport. Here, we highlight three aspects of the reduced-length CCR scale's development and validation that showcase the strength of our scale-shortening approach and the reduced-length CCR scale itself.

6.1.1 Scale Validity: How Can I Be Sure That the Reduced-Length CCR Scale Accurately Measures Human–Robot Rapport? The steps we followed to validate the reduced-length CCR scale ("Methodological Steps During Short-Form Validation" by Smith et al. [89] detailed in Section 3.2) provided us with evidence that the reduced-length CCR scale meets several expectations to ensure an accurate measurement of the construct of rapport. First, we showed that the reduced-length CCR scale has *high reliability*: the items within the scale are cohesive and related to the same fundamental construct. Next, we demonstrated that the reduced-length CCR scale is *highly related to its parent*—the full-length CCR scale. We were able to show that the reduced-length CCR scale has both a strong correlation with the full-length CCR scale and the same 2-factor structure. Finally, we found that the reduced-length CCR scale has high *construct validity*, meaning that it reliably measures the construct of rapport. The reduced-length CCR scale has both a strong correlation with a different scale that also measures rapport (convergent validity) and an extremely weak correlation with a different scale that measures an unrelated construct (divergent validity). We also showed that the reduced-length CCR scale can differentiate between HRIs that are known to have different degrees of human–robot rapport, as evidenced by participants in Study 5 rating human–robot rapport higher with a responsive robot compared with an unresponsive robot (known-groups validity). These together show that the reduced-length CCR scale has a high degree of psychometric validation.

6.1.2 Item Selection: How Can I Trust That the Items Chosen for the Reduced-Length Scale Represent the Best 8 Items to Measure Rapport? To avoid potential bias from the research team in the reduced-length CCR scale item-selection process, we made two key decisions that guided our scale-reduction process. First, to select which items would be included in the reduced-length version of the CCR scale, we sought the input of external HRI experts ($N = 30$) to rate the quality of the scale items and then followed published guidelines for scale shortening [26, 89, 104] to use these expert ratings to construct a reduced-length scale. Second, we adopted and followed "A Priori Steps in Rigorous Short-Form Development" by Smith et al. [89] detailed in Section 3.1. To follow these steps, we defined our objectives in reducing the scale (A2): preserving the two-factor structure, including a similar number of items within each factor, and saving time for participants filling out the questionnaire. Following these steps also involved clarifying how we selected the items for the reduced-length scale (A2) and computing estimates of its reliability (A3), parent overlap (A4), validity correlations (A5), time saved (A7), and validity lost (A7). In sum, our reduced-length scale was constructed and validated by following published scale-shortening guidelines [26, 89, 104] and leveraging expert opinions of external HRI researchers, minimizing any potential bias of the research team in the item-selection process.

6.1.3 Scale Benefits: What Are the Benefits of the Reduced-Length CCR Scale Compared with the Full-Length Version? The primary benefit of using the reduced-length CCR scale is that it takes less time for participants to complete, without losing much in terms of the scale's ability to measure rapport. As shown in Figure 3, participants took an average of 30 seconds to complete the reduced-length CCR scale compared to an average of 52 seconds for the full-length CCR scale.

6.2 On Scale Reduction in HRI

The field of HRI relies heavily on the use of multiple scales to understand the complex interactions between humans and robots. However, this approach often requires participants to complete

numerous scales, which can lead to fatigue and decreased response rates [45, 62]. To address this issue, it is essential for HRI researchers to strive for excellent measurements that are not only reliable and valid but also concise and efficient. This can be achieved by following established guidelines for scale creation and reduction [26, 89, 104].

In Lin et al. [60] and this article, we provide a comprehensive example of how to create and evaluate scales in a systematic and scientifically appropriate manner. Our research demonstrates a strong method for scale development, from initial creation to validation, and showcases a roadmap for future researchers to follow. Furthermore, we highlight the importance of shortening scales in a principled way, rather than relying on ad hoc methods that can compromise the validity of the original scale. By providing guidelines and an example of how to reduce scales while maintaining their psychometric properties, our article aims to promote best practices in HRI research and encourage the development of more efficient and effective measurement tools.

By following established guidelines for scale creation and reduction, HRI researchers can ensure that their measurements are not only reliable and valid but also concise and participant-friendly. This is particularly important in HRI research, where complex constructs and multi-dimensional phenomena are often measured. Our article provides a valuable resource for researchers seeking to create and shorten scales in a way that is grounded in scientific theory and practice, and we hope that it will contribute to the development of more rigorous and effective measurement tools in the field of HRI.

6.3 When Should I Use the Full-Length vs. Reduced-Length CCR Scale?

Now that we have demonstrated a validated reduced-length version of the CCR scale, the question naturally arises: *now that two versions of the CCR scale exist (full-length and reduced-length), which should I use in my research study?* We would answer this question by posing and responding to the following two questions:

Q1—How important is the measurement of rapport to the research?
Q2—How time-sensitive is the administration of the questionnaire?

The more important the construct of rapport is to a researcher's central research question (Q1), the more we recommend the administration of the full-length CCR scale. We have shown that the reduced-length CCR scale captures 77% of the variability in the full-length scale, indicating that it does broadly cover the construct of rapport measured in the full-length version but not as fully. The full-length CCR scale is still the best way to capture the full extent of the construct of rapport.

While in an ideal world, the full-length CCR scale would always be administered to assess rapport, the realities of participant survey fatigue make more concise and efficient scales a necessity (Q2). Especially for online studies, repeated measurements using the same scale, and questionnaires measuring many different constructs, administering a scale with a longer length risks lower-quality data. Therefore, the more time-sensitive a questionnaire's administration becomes for the researcher, the more strongly we recommend the use of the reduced-length CCR scale as opposed to the full-length CCR scale.

Taken together, if rapport is the central construct of the research, we recommend administering the full-length CCR scale as it offers the most comprehensive measurement of rapport. In cases where time is a constraint, we recommend using the reduced-length CCR scale since it serves as an effective alternative with high internal reliability.

6.4 The Relationship between Robot Responsiveness and Rapport

In addition to constructing and validating a reduced-length version of the CCR scale, the work presented in this article also provides further strong evidence that robot responsiveness leads to

greater perceptions of human–robot rapport. Our results from Study 5, detailed in Section 5.7, show that participants viewed a greater degree of human–robot rapport in videos where the robot was designed to be responsive as opposed to unresponsive. These results, combined with those from Study 3 in Lin et al. [60] and Study 1 in Birnbaum et al. [18], demonstrate a firm connection between robot responsiveness and human–robot rapport that has now been demonstrated from both first-person [18, 60] and third-person (Study 5) perspectives and from both one narrow context (a person disclosing current stressors to a robot [18, 60]) and a broader set of HRI-relevant contexts (Study 5: healthcare, education, collaboration). This combination of empirical evidence provides very strong support for the positive influence of robot responsiveness on human–robot rapport.

Additionally, the results from both Study 5 in this article and Study 3 in Lin et al. [60] show that robot responsiveness influences both the connection and coordination factors within the CCR scale. The HRIs in these studies were viewed as having significantly more connection and coordination than a HRI with an unresponsive robot. This finding highlights one of the benefits of our rapport scale: we now have the capability to determine more about the theoretical structure of rapport and how to build different aspects of rapport better. It could be that different features of an interaction increase (or decrease) the different dimensions of rapport. For example, did each of the specific behaviors expressed by the robot (e.g., speech delays, eye contact, warm communication style) impact only the coordination factor, only the connection factor, or both? How might each minute of robot behavior contribute individually to each factor of rapport? We encourage HRI researchers to leverage the 2D nature of this scale to further tease apart robot behaviors and interaction strategies that can impact the multi-dimensional and multi-faceted nature of rapport.

7 Limitations

Even though we have devoted a significant amount of effort to make our reduced-length CCR scale generalizable (e.g., validating it across three different interaction contexts with three different robot embodiments), some limitations in our work should be discussed. In our online studies to assess the full-length and reduced-length CCR scales (Studies 1, 2, 5), we recruited participants who are native English speakers based in the United States. Future work should test our full-length and reduced-length CCR scales with a more diverse set of populations (e.g., participants who are not native English speakers or those who are not based in the United States) to strengthen our scales' generalizability. Also, the reduced-length CCR scale is only validated from a third-person perspective in an online setting in Study 5. Future research could further validate this scale from a first-person perspective in an in-person study, even though its parent, the full-length CCR scale, is already validated in-person in Study 3. In addition, while the items in the reduced-length CCR scale can be applied to robots that communicate verbally and nonverbally, we only validated it with verbal robots in Study 5. Future work is required to further validate the reduced-length CCR scale with robots that do not produce any sound (e.g., robotic arms and assistive feeding robots).

8 Conclusion

In this work, we developed and validated a reduced-length version of the CCR scale. Using HRI expert ratings of the scale items, we shortened the full-length CCR scale of 18 items to a reduced-length version containing 8 items, with 4 items in each of its two factors (connection and coordination). We then validated the reduced-length CCR scale by conducting an online experiment where participants viewed videos of HRIs that varied in robot responsiveness and evaluated those videos using the reduced-length and full-length versions of the CCR scale. Our results demonstrate the validity of the reduced-length CCR scale to measure human–robot rapport, additional evidence of the validity of the full-length CCR scale, and strong empirical evidence demonstrating the connection between robot responsiveness and rapport.

Acknowledgments

We thank Tsz Long Leung, Hannah Dinner, and Rachel Liu for their research assistance.

References

[1] 2024. The American Heritage Dictionary of the English Language. Retrieved from https://www.ahdictionary.com/
[2] 2024. Cambridge Dictionary. Retrieved from https://dictionary.cambridge.org/
[3] 2024. The Chambers Dictionary. Retrieved from https://chambers.co.uk/
[4] 2024. Collins English Dictionary. Retrieved from https://www.collinsdictionary.com/
[5] 2024. Dictionary.com. Retrieved from https://www.dictionary.com/
[6] 2024. Google's English Dictionary (Oxford Languages). Retrieved from https://www.google.com/
[7] 2024. The Merriam-Webster Dictionary. Retrieved from https://www.merriam-webster.com/
[8] 2024. Oxford English Dictionary. Retrieved from https://www.oed.com/
[9] 2024. Oxford Learner's Dictionaries. Retrieved from https://www.oxfordlearnersdictionaries.com/
[10] 2024. Vocabulary.com. Retrieved from https://www.vocabulary.com/
[11] Allison Abbe and Susan E. Brandon. 2013. The role of rapport in investigative interviewing: A review. *Journal of Investigative Psychology and Offender Profiling* 10, 3 (2013), 237–249.
[12] Alexander Mois Aroyo, Francesco Rea, Giulio Sandini, and Alessandra Sciutti. 2018. Trust and social engineering in human robot interaction: Will a robot make you disclose sensitive information, conform to its recommendations or gamble? *IEEE Robotics and Automation Letters* 3, 4 (2018), 3701–3708.
[13] Minja Axelsson, Micol Spitale, and Hatice Gunes. 2024. "Oh, sorry, I think I interrupted you": designing repair strategies for robotic longitudinal well-being coaching. In *Proceedings of the 2024 ACM/IEEE International Conference on Human-Robot Interaction*, 13–22.
[14] Alexandria Bellas, Stefawn Perrin, Brandon Malone, Kaytlin Rogers, Gale Lucas, Elizabeth Phillips, Chad Tossell, and Ewart de Visser. 2020. Rapport building with social robots as a method for improving mission debriefing in human-robot teams. In *2020 Systems and Information Engineering Design Symposium (SIEDS)*. IEEE, 160–163.
[15] Frank J. Bernieri. 1988. Coordinated movement and rapport in teacher-student interactions. *Journal of Nonverbal Behavior* 12, 2 (1988), 120–138.
[16] Frank J. Bernieri, John S. Gillis, Janet M. Davis, and Jon E. Grahe. 1996. Dyad rapport and the accuracy of its judgment across situations: A lens model analysis. *Journal of Personality and Social Psychology* 71, 1 (1996), 110.
[17] Shreyas Bhat, Joseph B. Lyons, Cong Shi, and X. Jessie Yang. 2024. Evaluating the impact of personalized value alignment in human-robot interaction: Insights into trust and team performance outcomes. In *Proceedings of the 2024 ACM/IEEE International Conference on Human-Robot Interaction*, 32–41.
[18] Gurit E. Birnbaum, Moran Mizrahi, Guy Hoffman, Harry T. Reis, Eli J. Finkel, and Omri Sass. 2016. What robots can teach us about intimacy: The reassuring effects of robot responsiveness to human disclosure. *Computers in Human Behavior* 63 (2016), 416–423.
[19] Godfred O. Boateng, Torsten B. Neilands, Edward A. Frongillo, Hugo R. Melgar-Quiñonez, and Sera L. Young. 2018. Best practices for developing and validating scales for health, social, and behavioral research: A primer. *Frontiers in Public Health* 6 (2018), 149.
[20] Anya Bouzida, Alyssa Kubota, Dagoberto Cruz-Sandoval, Elizabeth W. Twamley, and Laurel D. Riek. 2024. CARMEN: A cognitively assistive robot for personalized neurorehabilitation at home. In *Proceedings of the 2024 ACM/IEEE International Conference on Human-Robot Interaction*, 55–64.
[21] Shane R. Brady. 2015. Utilizing and adapting the delphi method for use in qualitative research. *International Journal of Qualitative Methods* 14, 5 (2015), 1609406915621381.
[22] Vivienne Bihe Chi and Bertram F. Malle. 2024. Interactive Human-Robot teaching recovers and builds trust, even with imperfect learners. In *Proceedings of the 2024 ACM/IEEE International Conference on Human-Robot Interaction*, 127–136.
[23] Galen R. Collins. 2020. Improving human–robot interactions in hospitality settings. *International Hospitality Review* 34, 1 (Apr. 2020), 61–79. DOI: https://doi.org/10.1108/IHR-09-2019-0019
[24] Sawyer Collins, Kenna Baugus Henkel, Zachary Henkel, Casey C. Bennett, Cedomir Stanojevic, Jennifer A. Piatt, Cindy L. Bethel, and Selma Sabanović. 2024. "An emotional support animal, without the animal": Design guidelines for a social robot to address symptoms of depression. In *Proceedings of the 2024 ACM/IEEE International Conference on Human-Robot Interaction*, 147–156.
[25] Colleen Cook, Fred Heath, and Russel L. Thompson. 2000. A meta-analysis of response rates in web-or internet-based surveys. *Educational and Psychological Measurement* 60, 6 (2000), 821–836.
[26] Jose M. Cortina, Zitong Sheng, Sheila K. Keener, Kathleen R. Keeler, Leah K. Grubb, Neal Schmitt, Scott Tonidandel, Karoline M. Summerville, Eric D. Heggestad, and George C. Banks. 2020. From alpha to omega and beyond! a look at

the past, present, and (possible) future of psychometric soundness in the journal of applied psychology. *The Journal of Applied Psychology* 105, 12 (2020), 1351.

[27] Anna B. Costello and Jason Osborne. 2019. Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis. *Practical Assessment, Research, and Evaluation* 10, 1 (2019), 7.

[28] Jean Duncombe and Julie Jessop. 2002. Doing rapport' and the ethics of 'faking Friendship'. In *Ethics in Qualitative Research.* Melanie Mauthner, Maxine Birch, Julie Jessop, Tina Miller, Jean Duncombe and Julie Jesso (Eds.), SAGE Publications Ltd, 108–122. DOI: https://doi.org/10.4135/9781849209090.n6

[29] Phil Edwards, Ian Roberts, Mike Clarke, Carolyn DiGuiseppi, Sarah Pratap, Reinhard Wentz, and Irene Kwan. 2002. Increasing response rates to postal questionnaires: Systematic review. *BMJ* 324, 7347 (2002), 1183.

[30] Gudrun Eisele, Hugo Vachon, Ginette Lafit, Peter Kuppens, Marlies Houben, Inez Myin-Germeys, and Wolfgang Viechtbauer. 2022. The effects of sampling frequency and questionnaire length on perceived burden, compliance, and careless responding in experience sampling data in a student population. *Assessment* 29, 2 (2022), 136–151.

[31] Wendy English, Merryn Gott, and Jackie Robinson. 2022. The meaning of rapport for patients, families, and healthcare professionals: A scoping review. *Patient Education and Counseling* 105, 1 (2022), 2–14.

[32] Weimiao Fan and Zheng Yan. 2010. Factors affecting response rates of the web survey: A systematic review. *Computers in Human Behavior* 26, 2 (2010), 132–139.

[33] Marlena R. Fraune, Benjamin C. Oisted, Catherine E. Sembrowski, Kathryn A. Gates, Margaret M. Krupp, and Selma Šabanović. 2020. Effects of robot-human versus robot-robot behavior and entitativity on anthropomorphism and willingness to interact. *Computers in Human Behavior* 105 (2020), 106220.

[34] Brandi N. Frisby and Matthew M. Martin. 2010. Instructor–student and student–student rapport in the classroom. *Communication Education* 59, 2 (2010), 146–164.

[35] Luis A. Fuente, Hannah Ierardi, Michael Pilling, and Nigel T. Crook. 2015. Influence of upper body pose mirroring in human-robot interaction. In *Proceedings of the 7th International Conference on Social Robotics (ICSR '15)*. Springer, 214–223.

[36] Rachel Gockley, Allison Bruce, Jodi Forlizzi, Marek Michalowski, Anne Mundell, Stephanie Rosenthal, Brennan Sellner, Reid Simmons, Kevin Snipes, Alan C. Schultz, et al. 2005. Designing robots for long-term social interaction. In *2005 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, 1338–1343.

[37] Jonathan Gratch, David DeVault, Gale M. Lucas, and Stacy Marsella. 2015. Negotiation as a challenge problem for virtual humans. In *Proceedings of the 15th International Conference on Intelligent Virtual Agents (IVA '15)*. Springer, 201–215.

[38] Jonathan Gratch, Ning Wang, Jillian Gerten, Edward Fast, and Robin Duffy. 2007. Creating rapport with virtual agents. In *Proceedings of the 7th International Conference on Intelligent Virtual Agents (IVA '07)*. Springer, 125–138.

[39] Jonathan Gratch, Anya Okhmatovskaia, Francois Lamothe, Stacy Marsella, Mathieu Morales, R. J. van der Werf, and Louis-Philippe Morency. 2006. Virtual rapport. In *6th International Conference on Intelligent Virtual Agents (IVA '06).* Springer, 14–27.

[40] Dwayne D. Gremler and Kevin P. Gwinner. 2000. Customer-employee rapport in service relationships. *Journal of Service Research* 3, 1 (2000), 82–104.

[41] Dwayne D. Gremler and Kevin P. Gwinner. 2008. Rapport-building behaviors used by retail employees. *Journal of Retailing* 84, 3 (2008), 308–324.

[42] Omer Gvirsman and Goren Gordon. 2024. Effect of social robot's role and behavior on parent-toddler interaction. In *Proceedings of the 2024 ACM/IEEE International Conference on Human-Robot Interaction*, 222–230.

[43] Michael Hagenow, Emmanuel Senft, Robert Radwin, Michael Gleicher, Michael Zinn, and Bilge Mutlu. 2024. A system for human-robot teaming through end-user programming and shared autonomy. In *Proceedings of the 2024 ACM/IEEE International Conference on Human-Robot Interaction*, 231–239.

[44] Denise Hebesberger, Tobias Koertner, Christoph Gisinger, and Jürgen Pripfl. 2017. A long-term autonomous robot at a care hospital: A mixed methods study on social acceptance and experiences of staff and older adults. *International Journal of Social Robotics* 9, 3 (2017), 417–429.

[45] Michael Hoerger. 2010. Participant dropout as a function of survey length in internet-mediated university studies: Implications for study design and voluntary participation in psychological research. *Cyberpsychology, Behavior and Social Networking* 13, 6 (2010), 697–700.

[46] Long-Jing Hsu, Philip B. Stafford, Weslie Khoo, Manasi Swaminathan, Kyrie Jig Amon, Hiroki Sato, Katherine M. Tsui, David J. Crandall, and Selma Sabanović. 2024. "Give it time:" Longitudinal panels scaffold older adults' learning and robot co-design. In *Proceedings of the 2024 ACM/IEEE International Conference on Human-Robot Interaction*, 283–292.

[47] Lixing Huang, Louis-Philippe Morency, and Jonathan Gratch. 2011. Virtual rapport 2.0. In *Proceedings of the 10th International Conference on Intelligent Virtual Agents (IVA '11)*. Springer, 68–79.

[48] Bahar Irfan, Aditi Ramachandran, Samuel Spaulding, Dylan F. Glas, Iolanda Leite, and Kheng Lee Koay. 2019. Personalization in long-term human-robot interaction. In *2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. IEEE, 685–686.

[49] Rajat Kumar Jenamani, Daniel Stabile, Ziang Liu, Abrar Anwar, Katherine Dimitropoulou, and Tapomayukh Bhattacharjee. 2024. Feel the bite: Robot-assisted inside-mouth bite transfer using robust mouth perception and physical interaction-aware control. In *Proceedings of the 2024 ACM/IEEE International Conference on Human-Robot Interaction*, 313–322.

[50] Oliver P. John, Eileen M. Donahue, and Robert L. Kentle. 1991. *Big Five Inventory (BFI)*. American Psychological Association. DOI: https://doi.org/10.1037/t07550-000

[51] Hyunsu Kim, Kevin Kam Fung So, and Jochen Wirtz. 2022. Service robots: Applying social exchange theory to better understand human–robot interactions. *Tourism Management* 92 (2022), 104537.

[52] Yunkyung Kim and Bilge Mutlu. 2014. How social distance shapes human–robot interaction. *International Journal of Human-Computer Studies* 72, 12 (2014), 783–795.

[53] Rachel Kirby, Jodi Forlizzi, and Reid Simmons. 2010. Affective social robots. *Robotics and Autonomous Systems* 58, 3 (2010), 322–332.

[54] Paula D. Koppel, K. Park Hye-Young, Leila S. Ledbetter, Emily J. Wang, Lesley C. Rink, and Jennie C. De Gagne. 2024. Rapport between nurses and adult patients with cancer in ambulatory oncology care settings: A scoping review. *International Journal of Nursing Studies* 149 (2024), 104611. DOI: https://doi.org/10.1016/j.ijnurstu.2023.104611

[55] Jacqueline M. Kory-Westlund. 2023. 2023. Implications of children's social, emotional, and relational interactions with robots for human–robot empathy. In *Conversations on Empathy: Interdisciplinary Perspectives on Empathy, Imagination and Othering*. Routledge, 256–279.

[56] Jacqueline M. Kory-Westlund and Cynthia Breazeal. 2019. A long-term study of young children's rapport, social emulation, and language learning with a peer-like robot playmate in preschool. *Frontiers in Robotics and AI* 6, (2019), 81.

[57] Matthew J. Leach. 2005. Rapport: A key to treatment success. *Complementary Therapies in Clinical Practice* 11, 4 (2005), 262–265.

[58] Min Kyung Lee, Jodi Forlizzi, Sara Kiesler, Paul Rybski, John Antanitis, and Sarun Savetsila. 2012. Personalization in HRI: A longitudinal field experiment. In *Proceedings of the 7th Annual ACM/IEEE International Conference on Human-Robot Interaction*, 319–326.

[59] Mike E. U. Ligthart, Simone M. De Droog, Marianne Bossema, Lamia Elloumi, Mirjam de Haas, Matthijs H. J. Smakman, Koen V. Hindriks, and Somaya Ben Allouch. 2024. Back to school-sustaining recurring child-robot educational interactions after a long break. In *Proceedings of the 2024 ACM/IEEE International Conference on Human-Robot Interaction*, 433–442.

[60] Ting-Han Lin, Hannah Dinner, Tsz Long Leung, Bilge Mutlu, J. Gregory Trafton, and Sarah Sebo. 2025. Connection-coordination rapport (CCR) scale: A dual-factor scale to measure human-robot rapport. In *Proceedings of the 2025 ACM/IEEE International Conference on Human-Robot Interaction (HRI '25)*. IEEE Press, 869–879.

[61] Ting-Han Lin, Spencer Ng, and Sarah Sebo. 2022. Benefits of an interactive robot character in immersive puzzle games. In *2022 31st IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*. IEEE, 37–44.

[62] Mingnan Liu and Laura Wronski. 2018. Examining completion rates in web surveys via over 25,000 real-world surveys. *Social Science Computer Review* 36, 1 (2018), 116–124.

[63] Nichola Lubold, Erin Walker, and Heather Pon-Barry. 2021. Effects of adapting to user pitch on rapport perception, behavior, and state with a social robotic learning companion. *User Modeling and User-Adapted Interaction* 31 (2021), 35–73.

[64] Nichola Lubold, Erin Walker, Heather Pon-Barry, and Amy Ogan. 2019. Comfort with robots influences rapport with a social, entraining teachable robot. In *Proceedings of the 20th International Conference on Artificial Intelligence in Education (AIED '19), Part I*. Springer, 231–243.

[65] Robert C. MacCallum, Michael W. Browne, and Hazuki M. Sugawara. 1996. Power analysis and determination of sample size for covariance structure modeling. *Psychological Methods* 1, 2 (1996), 130.

[66] Robert C. MacCallum, Keith F. Widaman, Shaobo Zhang, and Sehee Hong. 1999. Sample size in factor analysis. *Psychological Methods* 4, 1 (1999), 84.

[67] Rishabh Madan, Skyler Valdez, David Kim, Sujie Fang, Luoyan Zhong, Diego T. Virtue, and Tapomayukh Bhattacharjee. 2024. RABBIT: A robot-assisted bed bathing system with multimodal perception and integrated compliance. In *Proceedings of the 2024 ACM/IEEE International Conference on Human-Robot Interaction*, 472–481.

[68] D. Betsy McCoach, Robert K. Gable, and John P. Madura. 2013. *Instrument Development in the Affective Domain*. Vol. 10. Springer.

[69] Derek McColl and Goldie Nejat. 2013. Meal-time with a socially assistive robot and older adults at a long-term care facility. *Journal of Human-Robot Interaction* 2, 1 (2013), 152–171.
[70] Lynden K. Miles, Louise K. Nind, and C. Neil Macrae. 2009. The rhythm of rapport: Interpersonal synchrony and social perception. *Journal of Experimental Social Psychology* 45, 3 (2009), 585–589.
[71] Steven M. Miller. 1952. The participant observer and over-rapport. *American Sociological Review* 17, 1 (1952), 97–99.
[72] Noriaki Mitsunaga, Christian Smith, Takayuki Kanda, Hiroshi Ishiguro, and Norihiro Hagita. 2008. Adapting robot behavior for human–robot interaction. *IEEE Transactions on Robotics* 24, 4 (2008), 911–916.
[73] Jaclyn Molan, Laura Saad, Eileen Roesler, J. Malcolm McCurry, Nathaniel Gyory, and J. Gregory Trafton. 2025. The perceived danger (PD) scale: Development and validation. In *Proceedings of the 2025 ACM/IEEE International Conference on Human-Robot Interaction (HRI '25)*. IEEE Press, 420–428.
[74] Jonathan Mumm and Bilge Mutlu. 2011. Human-robot proxemics: Physical and psychological distancing in human-robot interaction. In *Proceedings of the 6th International Conference on Human-Robot Interaction*, 331–338.
[75] David A. Neequaye and Erik Mac Giolla. 2022. The use of the term rapport in the investigative interviewing literature: A critical examination of definitions. *Meta-Psychology* 6 (2022).
[76] Spencer Ng, Ting-Han Lin, You Li, and Sarah Sebo. 2024. Role-playing with robot characters: Increasing user engagement through narrative and gameplay agency. In *Proceedings of the 2024 ACM/IEEE International Conference on Human-Robot Interaction*, 522–532.
[77] Tatsuya Nomura and Takayuki Kanda. 2016. Rapport–expectation with a robot scale. *International Journal of Social Robotics* 8 (2016), 21–30.
[78] Amy O'Connell, Ashveen Banga, Jennifer Ayissi, Nikki Yaminrafie, Ellen Ko, Andrew Le, Bailey Cislowski, and Maja Mataric. 2024. Design and evaluation of a socially assistive robot schoolwork companion for college students with ADHD. In *Proceedings of the 2024 ACM/IEEE International Conference on Human-Robot Interaction*, 533–541.
[79] Katarzyna Pasternak, Zishi Wu, Ubbo Visser, and Christine Lisetti. 2021. Let's be friends! A rapport-building 3D embodied conversational agent for the human support robot. arXiv:2103.04498. Retrieved from https://arxiv.org/abs/2103.04498
[80] Katarzyna Pasternak, Zishi Wu, Ubbo Visser, and Christine Lisetti. 2021. Towards building rapport with a human support robot. In *Robot World Cup*. Springer, 214–225.
[81] Hailian Qiu, Minglong Li, Boyang Shu, and Billy Bai. 2020. Enhancing hospitality experience with service robots: The mediating role of rapport building. *Journal of Hospitality Marketing and Management* 29, 3 (2020), 247–268.
[82] Beatrice Rammstedt and Oliver P. John. 2007. Measuring personality in one minute or less: A 10-item short version of the big five inventory in English and German. *Journal of Research in Personality* 41, 1 (2007), 203–212.
[83] William Revelle and Richard E. Zinbarg. 2009. Coefficients alpha, beta, omega, and the glb: Comments on sijtsma. *Psychometrika* 74 (2009), 145–154.
[84] Laurel D. Riek, Philip C. Paul, and Peter Robinson. 2010. When my robot smiles at me: Enabling human-robot rapport via real-time head gesture mimicry. *Journal on Multimodal User Interfaces* 3 (2010), 99–108.
[85] Laura Saad, Eileen Roesler, Elizabeth K. Phillips, and J. Gregory J. Trafton. 2026. Choosing the "perfect" scale: A primer to evaluate existing scales in HRI. *ACM Transactions on Human-Robot Interaction* 15, 2, Article 42 (2026), 29.
[86] Stela H. Seo, Keelin Griffin, James E. Young, Andrea Bunt, Susan Prentice, and Verónica Loureiro-Rodríguez. 2018. Investigating people's rapport building and hindering behaviors when working with a collaborative robot. *International Journal of Social Robotics* 10 (2018), 147–161.
[87] Amanda Sharkey and Noel Sharkey. 2011. Children, the elderly, and interactive robots. *IEEE Robotics and Automation Magazine* 18, 1 (2011), 32–38.
[88] Cory-Ann Smarr, Tracy L. Mitzner, Jenay M. Beer, Akanksha Prakash, Tiffany L. Chen, Charles C. Kemp, and Wendy A. Rogers. 2014. Domestic robots for older adults: Attitudes, preferences, and potential. *International Journal of Social Robotics* 6 (2014), 229–247.
[89] Gregory T. Smith, Denis M. McCarthy, and Kristen G. Anderson. 2000. On the sins of short-form development. *Psychological Assessment* 12, 1 (2000), 102.
[90] Bo Song, Meng Zhang, and Peipei Wu. 2022. Driven by technology or sociality? Use intention of service robots in hospitality from the human–robot interaction perspective. *International Journal of Hospitality Management* 106 (2022), 103278.
[91] Helen Spencer-Oatey. 2004. *Culturally Speaking: Managing Rapport through Talk across Cultures*. A&C Black.
[92] Helen Spencer-Oatey. 2005. (Im) politeness, face and perceptions of rapport: Unpackaging their bases and interrelationships. *Journal of Politeness Research* 1, 1 (2005), 95–119. DOI: https://doi.org/10.1515/jplr.2005.1.1.95
[93] Janis Stolzenwald and Paul Bremner. 2017. Gesture mimicry in social human-robot interaction. In *2017 26th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*. IEEE, 430–436.

[94] Dakota Sullivan, Nathan Thomas White, Andrew Schoen, and Bilge Mutlu. 2024. Making informed decisions: Supporting cobot integration considering business and worker preferences. In *Proceedings of the 2024 ACM/IEEE International Conference on Human-Robot Interaction*, 706–714.

[95] Angelique Taylor, Tauhid Tanjim, Huajie Cao, and Hee Rin Lee. 2024. Towards collaborative crash cart robots that support clinical teamwork. In *Proceedings of the 2024 ACM/IEEE International Conference on Human-Robot Interaction*, 715–724.

[96] Xiaoyi Tian, Nichola Lubold, Leah Friedman, and Erin Walker. 2020. Understanding rapport over multiple sessions with a social, teachable robot. In *Proceedings of the 21st International Conference on Artificial Intelligence in Education (AIED '20), Part II*. Springer, 318–323.

[97] Linda Tickle-Degnen and Robert Rosenthal. 1990. The nature of rapport and its nonverbal correlates. *Psychological Inquiry* 1, 4 (1990), 285–293.

[98] J. Gregory Trafton, Chelsea R. Frazier, Kevin Zish, J. Bio Branden, and J. Malcolm McCurry. 2023. The perception of agency: Scale reduction and construct validity. In *2023 32nd IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*. IEEE, 936–942.

[99] Yi-Shiuan Tung, Matthew B. Luebbers, Alessandro Roncone, and Bradley Hayes. 2024. Workspace optimization techniques to improve prediction of human motion during human-robot collaboration. In *Proceedings of the 2024 ACM/IEEE International Conference on Human-Robot Interaction*, 743–751.

[100] Alessandro Vercelli, Innocenzo Rainero, Ludovico Ciferri, Marina Boido, and Fabrizio Pirri. 2018. Robots in elderly care. *DigitCult-Scientific Journal on Digital Cultures* 2, 2 (2018), 37–50.

[101] Dave Walsh and Ray Bull. 2012. Examining rapport in investigative interviews with suspects: Does its building and maintenance work? *Journal of Police and Criminal Psychology* 27 (2012), 73–84.

[102] Rosalie H. Wang, Aishwarya Sudhama, Momotaz Begum, Rajibul Huq, and Alex Mihailidis. 2017. Robots to assist daily activities: Views of older adults with Alzheimer's disease and their caregivers. *International Psychogeriatrics* 29, 1 (2017), 67–79.

[103] Tina Yue, Alexandra E. Janiw, Aaron Huus, Salvador Aguiñaga, Megan Archer, Krista Hoefle, and Laurel D. Riek. 2012. Creating human-robot rapport with mobile sculpture. In *Proceedings of the 7th Annual ACM/IEEE International Conference on Human-Robot Interaction*, 223–224.

[104] Muhamad Saiful Bahri Yusoff. 2019. ABC of content validation and content validity index calculation. *Education in Medicine Journal* 11, 2 (2019), 49–54.

[105] Alex Wuqi Zhang, Ting-Han Lin, Xuan Zhao, and Sarah Sebo. 2023. Ice-breaking technology: Robots and computers can foster meaningful connections between strangers through in-person conversations. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, 1–14.

[106] Alex Wuqi Zhang, Rafael Queiroz, and Sarah Sebo. 2025. Balancing user control and perceived robot social agency through the design of end-user robot programming interfaces. In *2025 20th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. IEEE, 899–908.

Received 29 August 2025; revised 19 December 2025; accepted 13 January 2026