# Design of Social Features for Robot-mediated Cross-cultural Interaction

Sara Cooper
sara.cooper@jp.honda-ri.com
Honda Research Institute Japan
Wako, Japan

Randy Gomez
r.gomez@jp.honda-ri.com
Honda Research Institute Japan
Wako, Japan

Deborah Szapiro
deborah.szapiro@uts.edu.au
University Technology Sydney
Sydney, Australia

Luis Merino
lmercab@upo.es
Universidad Pablo de Olavide
Sevilla, Spain

## ABSTRACT

In this paper, we design a hierarchical set of socio-emotional features to understand the effect of cross-cultural mediation facilitated by a social robot between two remote groups of schoolchildren in Japan and in Australia. We also equip the robot with behaviors that maximize children's participation and stimulate socio-emotional interaction which in turn increases the chance of observing the designed features. Differences were observed between the cultures (AUS and JP), suggesting the need of future adaptation of the framework to different cultures as the pilot is expanded to more countries.

## CCS CONCEPTS

• **Human-centered computing → HCI design and evaluation methods**; **Empirical studies in HCI**.

## KEYWORDS

human-robot interaction, guidelines, socio-emotional learning

**ACM Reference Format:**
Sara Cooper, Randy Gomez, Deborah Szapiro, and Luis Merino. 2024. Design of Social Features for Robot-mediated Cross-cultural Interaction. In *Companion of the 2024 ACM/IEEE International Conference on Human-Robot Interaction (HRI '24 Companion), March 11–14, 2024, Boulder, CO, USA.* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3610978.3640597

## 1 INTRODUCTION

Schools nowadays embrace the exciting opportunity to nurture a diverse student body, each with its unique abilities and motivations for learning. In today's educational landscape, there is a strong emphasis on fostering essential social-emotional skills such as empathy and decision-making [5], improving both academic performance and overall well-being. Hence, it is important to factor in the notion of social-emotional learning (SEL) to the design of cutting-edge educational technologies, such as robots in the classrooms [6]. This trend has also gained attention from influential organizations like UNICEF [30] and OECD [4].

Educational robots may promote SEL and enhance learning outcomes in tasks like language learning, by offering a physical presence lacking in traditional learning technologies [3], or facilitating learning problem-solving skills and engagement [21]. As social mediators, they can use their gaze direction and speech abilities to promote equal social participation [13], or increase social awareness [27]. For example, Moxie by Embodied [1, 18] can engage through conversations and storytelling to build children's confidence and enhance both emotional and cognitive skills.

We detect and interpret socio-emotional cues to understand emotions and intentions, which helps us adjust our role in the interaction [20]. However, there are two main challenges: (1) lack of sufficient definition and mapping [2] between low-level cues that robots are usually programmed to capture (face detection, speech text) to higher-level features (engagement, empathy) (2) Need of indicators to evaluate interaction improvement from a more socio-emotional perspective.

To address this gap, this paper proposes Haru as a social cross-cultural mediator robot, taking as a basis its original UNICEF pilot study [31]. In past works, Haru has been used to promote child-team dynamics [15], as well as a facilitator to balance a multi-party interaction [8]. Specifically, it outlines the pilot study conducted between Japan and Australia and shares initial findings, aiming to identify socio-emotional features and design relevant surveys in collaboration with behavioral experts. For this, Haru adopts different roles to enhance socio-emotional features. At a later stage of the project, the robot will be taught to automatically detect the identified features and adapt its behavior accordingly, as well as escalate the study to 12 different schools around the world (USA, Ukraine, Italy, China...) to check on generalized and nuanced features.

## 2 BACKGROUND

### 2.1 Socio-emotional learning (SEL)

SEL positively impacts students across age groups, enhancing emotional detection, problem-solving, decision-making, attitudes, school bonding, and academic performance [23]. Different frameworks of SEL define different social competencies.

The well-known CASEL framework [7, 10] classifies 5 social competencies: self-awareness, self-management, social awareness, relationship skills, and responsible decision-making, aligning with OECD's emotion management [4], teamwork, and goal achievement model. The paper focuses on Social Awareness and Relationship Skills, detailed in its technical report [10, 11]. Social awareness


*HRI '24 Companion, March 11–14, 2024, Boulder, CO, USA*

ACM ISBN 979-8-4007-0323-2/24/03…$15.00
https://doi.org/10.1145/3610978.3640597

involves understanding diverse backgrounds, norms, and ethical behavior. Relationship skills encompass positive interactions, communication adaptation, collaboration, conflict resolution, and seeking assistance. The paper explores these competencies within a cross-cultural mediator robot scenario, aiming to enhance children's social awareness and relationship skills within the SEL framework.

## 2.2 Social feature classification

Affective computing introduced social signal processing [25] extracting emotions and features from audio, text, or gaze inputs. Verbal and non-verbal cues, classified by literature into social cues, behavioral signals, and social signals [12], convey attention, emotions, and intentions in human interaction. Works detecting social features from videos focus on detecting student engagement [19] or creating large-scale datasets for group interaction [20].

However, interpreting social signals like openness and engagement is complex for software. Especially, challenges arise in evaluating behavioral cues in group interactions compared to dyadic communication [20], hindering the understanding of social signals. Issues include limited datasets, difficulties measuring cues, and lack of rules for mapping cues to higher-level features [24].

In summary, defining the relationship between social cues is crucial for robots to automatically detect and adapt behavior, improving interaction.

# 3 METHODS

## 3.1 Social robot mediator pilot study

The scenario involves, in each of the two locations, 3 children, one Haru robot, two screens for video conference and audiovisual content, and one iPad for text and audio input for each child, seen in Figure 1. A teacher has a tablet to monitor the flow of the interaction, resulting in a Shared Control with Human Initiative.

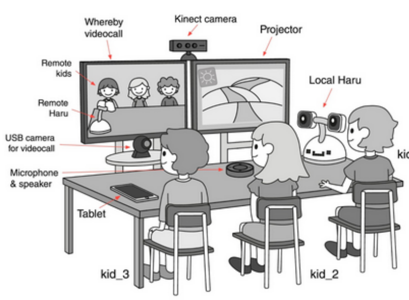


**Figure 1: Set-up for the encouraging mediator pilot**

The topic of this scenario was to discuss schools: how they go to school, what they eat, and what sports they play. Haru was programmed to adopt 8 different roles to promote different socio-emotional features, with the main ones being:

- **Equitable turn-taking:** Haru promotes inclusivity by facilitating turn-taking among children, using explicit prompts to decide who speaks next i.e. by showing a photo of each kid and gazing at the respective kid to ask a question [28].
- **Explainer:** Haru guides the kids through the process of each activity through prompts and clearly articulated guidance, supporting group interdependence and promoting group equity [29].
- **An Ice Breaker:** As children can be shy when talking in groups, Haru cracks jokes to break the ice when a long pause or no reply is detected from children, overcoming shyness [22].

The others consist of being a trusted ally, a validator, an immersive entertainer, a robot that sparks topics for conversation and provides a safe space for children.

The pilot starts with Haru's introduction and initiation of the call, syncing with the remote robot. In the videoconference, children introduce themselves, addressing privacy concerns along with information handling. Following Haru's explanation, children create avatars on iPads. Haru then guides the purpose of the meeting, where participants upload and tag relevant pictures. To break the ice, Haru shares a light joke. The interactive session involves discussions based on uploaded pictures, with Haru prompting each child. The session concludes with reflections and gratitude in the farewell.

Three trial interaction sessions were conducted for two months. Twelve participants were recruited: 7 males in Japan (avg. age 14.85), from Jiyugaoka High School all-boys school (Tokyo), and 4 females plus 1 male in Australia (avg. age 15), from Tempe High School (Sydney). Japanese children with a minimum English proficiency were recruited. Participants were grouped in threes by country, ensuring everyone interacted with the system at least once. To adhere to ethical standards in social robot design, the project aligns with the Sustainable Development Goals (SDGs) in UNICEF's AI policy [30] and NHMRC ethical conduct [17]. Ethics approval was secured from each school region.

## 3.2 Definition of features by behavioral experts

Two behavioral experts identified and analyzed social features within this scenario, classifying them into low-, mid- and high-level categories. Based on CASEL wheel described in Section 2.1 and its technical report [10], we focus on **high-level features** Social Awareness and Relationship Skills. From here, 7 **mid-level** features were identified including active listening, empathy, acknowledgment, and social engagement.

As for **low-level features**, four main features have been identified that facilitate the detection of the higher features described in the previous section: speech, gaze, body and hand gestures, and facial expressions.

Based on this information, this paper proposes a design framework of socio-emotional features, summarised in Table 1, that a social mediator robot like Haru can help promote. Moreover, it proposes a 5-point Likert Scale survey that participants may fill out to correlate with the features. With this table, we also aim to motivate researchers to detect combinations of gaze, speech, and gestures leading to more complex features and enabling the evaluation of improvement in interaction.

## 3.3 Video annotation

Videos from 2 cameras of the 20.3 min long interaction were annotated in parallel with the design of socio-emotional features, to identify features from Section 3.2, following similar steps as [19]

**Table 1: Proposed design framework of socio-emotional features**

| Mid-level feature | Definition | Measurement methods | | | | |
|---|---|---|---|---|---|---|
| | | Survey | Low-level feature | | | |
| | | Likert scale (1-5) | Speech | Gaze | Facial expression | Hand/body gestures |
| Active listening | Ability to absorb information shared by Haru, peer, and reflecting on them through questions and by presence | "I was attentive to the speaker while he/she was speaking" | 1. "Yeah, ok, true, right" 2 times during speaker speaking (Haru, peers) 2. "Do you mean…?" 2 times during speaker speaking 3. Mirror 1 word or phrase from speaker ("table tennis") | 1. Eye contact with speaker for 2 seconds | 1. Smile 2 times to speaker 2. Mirror facial expression of speaker | 1. Nod 2 times during speaker speaking 2. Change posture to lean in to the speaker |
| Empathy | Show the ability to understand and share feelings and perspectives of others | "I tried to understand how my classmates were feeling during our conversation and considered their perspectives." | 1. Empathetic language 2 times: oh that is so cool! That is so pretty, such a shame, oh that is yummy, you take as much as an hour to go to school? | 1. Eye contact with speaker for 2 seconds | 1. Smile for 2 seconds to peers or Haru 2. Mirror facial emotion of Haru or peers | 1. Nod 2 times to same peer or Haru |
| Turn-taking | Effectively participates in discussion, promoting others to engage equally | (no need to survey) | 1. Provide at least 2 times 1-second pause between phrases when talking 2. "what do you think…?", "which avatar did you choose"? Asking once 3. At least 1 follow-up question to peers "do you usually eat inside or outside?" | 1.When talking, switch gaze once from two different peers or peer to Haru 2. When listening, alternate gaze from speaker to a non-speaker | 1. Smile for 2 seconds to a non-speaker when talking | 1. Nod once to a non-speaker 2. Lean back 3. Open palm or point once to a non-speaker |
| Proactiveness | Inclination and readiness of children to actively engage in the conversation by either asking insightful follow-up questions or providing comprehensive answers to open-ended questions. | "I actively participated in the conversation by either asking follow-up questions or being the first to answer open-ended questions." | 1. Initiate asking a question or a comment to peer: "what did you choose?", "what do you do at school festivals?" | 1. Initiate eye contact with peer once to invite them to converse | 1. Smile for 2 seconds to anyone in the group | 1. Being the first to wave or nod |
| Social engagement | Active involvement, attention, and interaction during the conversation | "I actively engaged with my peers,and responded in a friendly and cooperative manner." | 1. Respond once to peer or robot comments "Oh that is yummy!" 2. Answer questions asked: "I like swimming" 3. Comments during open discussion, such as when watching welcome video or avatar creation | 1. Eye contact of 2 seconds with a same peer or Haru | 1. Smile for 2 seconds at any point in the discussion 2. Laugh once at any point in the discussion 3. Mirror smile or laughter | 1. Nod twice 2. Lean forward during group discussion 3. Move hands with speech |
| Acknowledgement | Actively pay attention and show responses to what the speaker is saying | "I acknowledged their contributions by nodding, smiling, or providing verbal feedback when appropriate." | 1. Acknowledgement responses like "yeah", "wow", "ok", "that is good" at least twice | 1. Eye contact of 2 seconds with focus of attention (speaker, projector) | 1. Smile for 2 seconds at speaker when appropriate | 1. Nod or wave twice 2. Mirror a nod or wave from peer 3. Lean forward to speaker |
| Openness | Ability to consider different viewpoints, share their thoughts openly | "I actively shared my opinions openly and encouraged others to share their own" | 1. A phrase expressing an opinion: "I like pizza", "I think Haru would be a flower" | 1. Eye contact of 2 seconds with a same peer or Haru, two times | 1. Smile at least once when talking about an opinion 2. Expression of interest when listening | 1. Open palms 2. Use hand gestures with speech |

and [14], but extending from low- to higher level features. For this, video coding tool ELAN [33] was, developed by Max Planck Institute for Psycholinguistics. ELAN enables simultaneous tracking of gaze, speech transcription, and non-verbal cues. Annotations followed the structure of Table 1, utilizing high-, mid-, and low-level labels. For instance, under the low-level feature "Gaze," annotations specified Child1: LookAtRobot, Child2: LookAtPeer, Child3: LookAtPeer.

## 4 ANALYSIS OF SOCIAL FEATURES

This section illustrates the application of the design framework (Table 1) in analyzing group interactions within the social mediator robot scenario. Focusing on the example of Haru promoting equitable turn-taking through discussion, a scenario is presented where Haru displays a photo, prompting children to elaborate and ask related questions to each other.

### 4.1 Low-level feature analysis

Figure 2 shows an annotation of when Haru adopts the **Equitable turn-taker role**, and the children are encouraged to discuss their photos. In this particular snapshot, an Australian kid asks Japan "Do you live close to the beach much?", using hand gestures as well. Peers exhibit shared attention and active listening. After a brief response from Japan, another Australian child follows up, eliciting smiles. Japanese children respond more elaborately, fostering increased participation, verbal feedback, and joint gazing.

In Figure 3, common features include maintaining eye contact with the focus (Haru, peers, or the remote) and nodding. Australian children exhibit higher engagement, smiling more, nodding, and providing empathetic responses. Conversely, Japanese children show fewer social cues, sometimes looking down. Upon the robot's initiation, children ask follow-up queries. The question-response

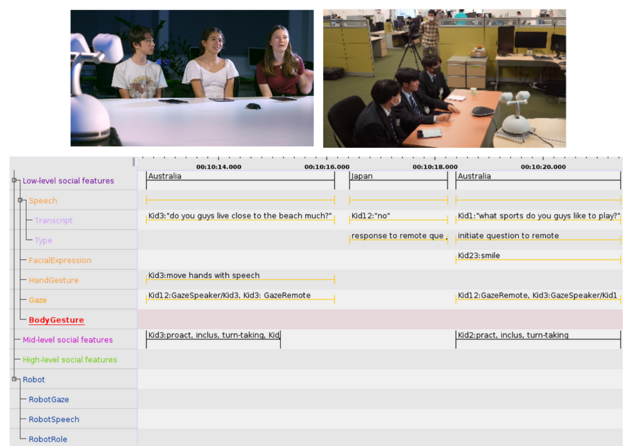


**Figure 2: ELAN annotation during turn-taking**

count reflects initiative and turn-taking. Japanese children offer more comments, possibly due to Australian children's proactive questioning.

In this scenario, examining the correlation between features in Japan and Australia revealed key findings (Table 2):

- **Higher Initiative in Questioning and Active Listening:** Individuals initiating questions demonstrate active listening behaviors, fostering attentive responses.
- **Increased Focused Gaze, Facial Expressions, and Empathy:** Gazing at the focus correlates with positive facial expressions and empathetic responses, indicating encouragement, emotional connection, and mutual understanding.

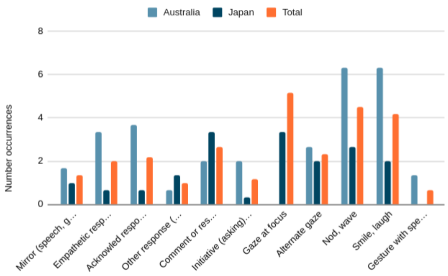


**Figure 3: Main low-level social features in turn-taking**

**Table 2: Pearson correlation of low-level features in turn-taking (blue = higher correl)**

| | Nodding | Initiative (asking) peer or remote | Other comments | Gaze at focus | Facial expression(smile, laugh) | Gestures with speech | Mirroring | Empathetic speech |
|---|---|---|---|---|---|---|---|---|
| Nodding | 1.00 | | | | | | | |
| Initiative (asking) peer or remote | **0.95** | 1.00 | | | | | | |
| Other comments | -0.17 | 0.13 | 1.00 | | | | | |
| Gaze at focus | 0.69 | 0.54 | -0.51 | 1.00 | | | | |
| Facial expression (smile, laugh) | 0.71 | 0.61 | -0.26 | **0.85** | 1.00 | | | |
| Gestures with speech | **0.89** | 0.75 | -0.43 | 0.56 | 0.63 | 1.00 | | |
| Mirroring | 0.00 | -0.18 | -0.31 | -0.18 | 0.05 | 0.44 | 1.00 | |
| Empathetic speech | 0.59 | 0.50 | -0.31 | **0.93** | **0.92** | 0.43 | -0.20 | 1.00 |

## 4.2 Mid-level social feature analysis

This analysis aimed to assess the frequency of mid-level social features elicited by the robot, identifying the most common ones and associated low-level features. During the given Equitable turn-taking scenario, as seen in Figure 4, acknowledgment and social engagement predominated followed by active listening in discussion and shared attention in listening. In contrast to other robot roles analyzed, turn-taking, empathetic responses towards peers, and openness are elevated in this scenario, facilitated by the robot's role in encouraging kids to share thoughts and exchange ideas.

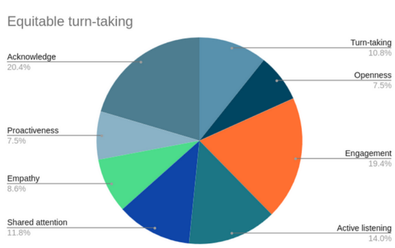


**Figure 4: Frequent mid-level features in turn-taking**

73% of instances showed social engagement behaviors, characterized by focused gaze at the subject (e.g., speaker, Haru, peer, or remote, 68.2% of times) and verbal feedback (82%), often encompassing questions or responses, correlating with proactiveness.

Additionally, turn-taking behaviors, where children turned to or asked another peer questions, featured alternating gaze (76% of times). Japanese participants frequently looked at peers before responding, while Australian children engaged in mutual gaze during discussions on topics like avatars.

# 5 DISCUSSION

This research identifies socio-emotional features for evaluating Socio-Emotional Learning (SEL) through collaboration with behavioral experts, as outlined in Table 1. Haru the robot exhibits potential in fostering socio-emotional development such as social engagement or acknowledgement, particularly in the Equitable Turn-taking role, highlighting correlations between question initiative and non-verbal cues like gazing and nodding. Additionally, more empathetic responses were observed, possibly attributed to Haru not being perceived as too human-like, which could be a positive indicator of the robot's ability to encourage empathy towards other children [26].

Cultural differences between Japanese and Australian children in social expressions are evident. Japanese children, less facially expressive, use nods and concise verbal responses for acknowledgements, reducing turn-taking and proactiveness. Australians, with a more direct communication style, show more smiles and active listening. Noteworthy features include Japanese children looking at peers before responding, emphasizing consensus. Considering these cultural nuances is crucial for measuring socio-emotional features and adapting robot behavior across cultures [13, 16, 32].

Study limitations include potential language barriers for Japanese children and a focus on feature identification rather than automated detection (addressed in future work), hence why only two cultures were studied so far. Only Equitable Turn-taking results are presented here, with additional findings in future papers.

Recognizing the universality of social cues, researchers can apply proposed surveys, robot roles, and feature mapping (Table 1) in diverse human-robot interaction scenarios like storytelling, virtual classrooms or emotional support, as well as use the features as indicators for interaction evaluation. Future pilots will extend to the other 12 participating schools [9] to assess cultural differences once automatic feature acquisition is developed.

# 6 CONCLUSIONS

A robust socio-emotional development is crucial to be able to make meaningful connections. This paper proposes the design of a hierarchical socio-emotional features framework, together with behavioral experts, by analyzing a scenario between children in Japan and Australia as it is mediated by a robot. The robot takes roles such as Explainer or Equitable Turn-Taker to promote exhibition of features defined including social engagement, openness or proactiveness, therefore showing potential of such a robot to enhance socio-emotional interaction. Cultural differences have been detected between Australia and Japan which suggests the need for cultural adaptation of the framework. The identified features as well as the associated surveys can be used by fellow researchers in scenarios such as virtual classrooms to measure socio-emotional interaction. The next step involves developing a system to automatically detect these features and adapt robot behaviour accordingly.

# 7 ACKNOWLEDGMENTS

We are thankful to Vicky Charisi (European Commision-JRC), Steven Vosloo (UNICEF), Tomoko Imai (Jiyugaoka H.S. Tokyo), and participating schools [9].

# REFERENCES


[1] Evan Ackerman. 2020. Meet Moxie, a Social Robot That Helps Kids With Social-Emotional Learning. https://spectrum.ieee.org/moxie-a-social-robot-for-childhood-development. [Online; accessed 7-Sept-2023].

[2] Mehrabian Albert. 1967. Inference of attitudes from nonverbal communication in two channels. *Journal of consulting psychology* 31 (1967), 3.

[3] Tony Belpaeme, James Kennedy, Aditi Ramachandran, Brian Scassellati, and Fumihide Tanaka. 2018. Social robots for education: A review. *Science robotics* 3, 21 (2018), eaat5954.

[4] Tony Belpaeme and Fumihide Tanaka. 2021. Social robots as educators. In *OECD Digital Education Outlook 2021 Pushing the Frontiers with Artificial Intelligence, Blockchain and Robots: Pushing the Frontiers with Artificial Intelligence, Blockchain and Robots*. OECD Publishing Paris, 143.

[5] Peter L Benson. 2006. *All kids are our kids: What communities must do to raise caring and responsible children and adolescents*. Jossey-Bass.

[6] Robert Wm Blum, Heather P Libbey, John H Bishop, and Matthew Bishop. 2004. School connectedness—Strengthening health and education outcomes for teenagers. *Journal of School Health* 74, 7 (2004), 231–235.

[7] CASEL. 2021. *What Does the Research Say?* https://casel.org/fundamentals-of-sel/what-does-the-research-say/ Accessed on: accessed 26-Sept-2023.

[8] Jouh Yeong Chew and Keisuke Nakamura. 2023. Who to Teach a Robot to Facilitate Multi-party Social Interactions?. In *Companion of the 2023 ACM/IEEE International Conference on Human-Robot Interaction*. 127–131.

[9] Socially Intelligent Robotics Consortium. 2018. Embodied Mediator for Cross Cultural Understanding among Children across the World. https://mypersonalrobots.org/pilot. [Online; accessed 7-Sept-2023].

[10] Jennifer L DePaoli, Matthew N Atwell, and John Bridgeland. 2017. Ready to Lead: A National Principal Survey on How Social and Emotional Learning Can Prepare Children and Transform Schools. A Report for CASEL. *Civic Enterprises* (2017).

[11] European Union. 2018. *Strengthening Social and Emotional Education as a core curricular area across the EU*. https://nesetweb.eu/wp-content/uploads/AR3_Full-Report.pdf Accessed on: accessed 26-Sept-2023.

[12] Stephen M Fiore, Travis J Wiltshire, Emilio JC Lobato, Florian G Jentsch, Wesley H Huang, and Benjamin Axelrod. 2013. Toward understanding social cues and signals in human–robot interaction: effects of robot gaze and proxemic behavior. *Frontiers in psychology* 4 (2013), 859.

[13] Sarah Gillet, Katie Winkle, Giulia Belgiovine, and Iolanda Leite. 2022. Ice-breakers, turn-takers and fun-makers: Exploring robots for groups with teenagers. In *2022 31st IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*. IEEE, 1474–1481.

[14] Manuel Giuliani, Nicole Mirnig, Gerald Stollnberger, Susanne Stadler, Roland Buchner, and Manfred Tscheligi. 2015. Systematic analysis of video data from different human–robot interaction studies: a categorization of social signals during error situations. *Frontiers in psychology* 6 (2015), 931.

[15] Randy Gomez, Deborah Szapiro, Luis Merino, and Keisuke Nakamura. 2020. A holistic approach in designing tabletop robot's expressivity. In *2020 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 1970–1976.

[16] William B Gudykunst, Yuko Matsumoto, Stella Ting-Toomey, Tsukasa Nishida, Kwangsu Kim, and Sam Heyman. 1996. The influence of cultural individualism-collectivism, self construals, and individual values on communication styles across cultures. *Human communication research* 22, 4 (1996), 510–543.

[17] National Health and Medical Research Council. 2007. National Statement on Ethical Conduct in Human Research. https://www.nhmrc.gov.au/about-us/publications/national-statement-ethical-conduct-human-research-2007-updated-2018#block-views-block-file-attachments-content-block-1. [Online; accessed 7-Sept-2023].

[18] Nikki Hurst, Caitlyn Clabaugh, Rachel Baynes, Jeff Cohn, Donna Mitroff, and Stefan Scherer. 2020. Social and emotional skills training with embodied moxie. *arXiv preprint arXiv:2004.12962* (2020).

[19] Minsu Jang, Dae-Ha Lee, Jaehong Kim, and Youngjo Cho. 2013. Identifying principal social signals in private student-teacher interactions for robot-enhanced education. In *2013 IEEE RO-MAN*. IEEE, 621–626.

[20] Hanbyul Joo. 2019. *Sensing, Measuring, and Modeling Social Signals in Nonverbal Communication*. Ph. D. Dissertation. Carnegie Mellon University Pittsburgh, PA, USA.

[21] James Kennedy, Paul Baxter, and Tony Belpaeme. 2015. Comparing robot embodiments in a guided discovery learning interaction with children. *International Journal of Social Robotics* 7 (2015), 293–308.

[22] Jill F Kilanowski. 2012. Breaking the ice: a pre-intervention strategy to engage research participants. *Journal of Pediatric Health Care* 26, 3 (2012), 209–212.

[23] Christina F Mondi, Alison Giovanelli, and Arthur J Reynolds. 2021. Fostering socio-emotional learning through early childhood intervention. *International Journal of Child Care and Education Policy* 15, 1 (2021), 1–43.

[24] Behnaz Nojavanasghari, Tadas Baltrušaitis, Charles E Hughes, and Louis-Philippe Morency. 2016. Emoreact: a multimodal approach and dataset for recognizing emotional responses in children. In *Proceedings of the 18th acm international conference on multimodal interaction*. 137–144.

[25] Soujanya Poria, Erik Cambria, Rajiv Bajpai, and Amir Hussain. 2017. A review of affective computing: From unimodal analysis to multimodal fusion. *Information fusion* 37 (2017), 98–125.

[26] Laurel D Riek, Tal-Chen Rabinowitch, Bhismadev Chakrabarti, and Peter Robinson. 2009. How anthropomorphism affects empathy toward robots. In *Proceedings of the 4th ACM/IEEE international conference on Human robot interaction*. 245–246.

[27] Eleanor Sandry, Randy Gomez, and Keisuke Nakamura. 2021. Art, design and communication theory in creating the communicative social robot 'haru'. *Frontiers in Robotics and AI* 8 (2021), 577107.

[28] Hamish Tennent, Solace Shen, and Malte Jung. 2019. Micbot: A peripheral robotic object to shape conversational dynamics and team performance. In *2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. IEEE, 133–142.

[29] Elli J Theobald, Sarah L Eddy, Daniel Z Grunspan, Benjamin L Wiggins, and Alison J Crowe. 2017. Student perception of group dynamics predicts individual performance: Comfort and equity matter. *PloS one* 12, 7 (2017), e0181336.

[30] UNICEF. 2020. Policy Guidance on AI for Children. https://www.unicef.org/globalinsight/reports/policy-guidance-ai-children. [Online; accessed 7-Sept-2023].

[31] UNICEF. 2021. Honda Research Institute Japan European Commission, Joint Research Centre. https://www.unicef.org/globalinsight/media/2206/file. [Online; accessed 7-Sept-2023].

[32] Shota Uono and Jari K Hietanen. 2015. Eye contact perception in the west and east: A cross-cultural study. *Plos one* 10, 2 (2015), e0118094.

[33] Peter Wittenburg, Hennie Brugman, Albert Russel, Alex Klassmann, and Han Sloetjes. 2006. ELAN: A professional framework for multimodality research. In *5th international conference on language resources and evaluation (LREC 2006)*. 1556–1559.