



OPEN ACCESS

EDITED BY

Loulou Kosmala,
Université Paris-Est Créteil Val de Marne,
France

REVIEWED BY

Renia Lopez,
Hong Kong Polytechnic University, Hong
Kong SAR, China
Gabriele Fassina,
Polytechnic University of Milan, Italy

*CORRESPONDENCE

Nils F. Tolksdorf
✉ tolksdorf@idf.uni-heidelberg.de

RECEIVED 20 June 2025
REVISED 28 April 2026
ACCEPTED 29 April 2026
PUBLISHED 29 June 2026

CITATION

Tolksdorf NF, Rohlfsing KJ, Grimminger A
and Sekine K (2026) Social robots as a
method for cross-cultural gesture
elicitation in children.
Front. Commun. 11:1651149.
doi: 10.3389/fcomm.2026.1651149

COPYRIGHT

© 2026 Tolksdorf, Rohlfsing, Grimminger
and Sekine. This is an open-access
article distributed under the terms of the
[Creative Commons Attribution License
\(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or
reproduction in other forums is
permitted, provided the original author(s)
and the copyright owner(s) are credited
and that the original publication in this
journal is cited, in accordance with
accepted academic practice. No use,
distribution or reproduction is permitted
which does not comply with these
terms.

Social robots as a method for cross-cultural gesture elicitation in children

Nils F. Tolksdorf^{1*}, Katharina J. Rohlfsing², Angela Grimminger²
and Kazuki Sekine³

¹Institute of German as a Foreign Language Philology, Heidelberg University, Heidelberg, Germany, ²Psycholinguistics, Paderborn University, Paderborn, Germany, ³Faculty of Human Sciences, Waseda University, Saitama, Japan

The study of children's gestures from a cross-cultural perspective has attracted considerable interest in recent years. At the same time, it poses methodological challenges, such as ensuring cross-cultural comparability and culturally sensitive elicitation procedures. In this methods paper, we address the methodological challenges of eliciting gestures from children in structured yet naturalistic interactions by proposing a set of theory-informed design principles for using social robots as interactive partners in standardized, task-oriented interactions across cultural contexts. Building on an interactionist perspective grounded in theoretical considerations from pragmatic frames that integrate task, partner, and interaction models, we present a scripted, role-based paradigm to which the design principles were applied. In this paradigm, a humanoid robot adopts the role of a novice learner and engages preschool-aged children in dialogue to explain familiar, everyday actions, including prompting corrective behavior from children when the robot systematically makes errors in recalling previously demonstrated actions. We propose procedures intended to support cultural adaptability while maintaining comparable interactional contingencies across cultural environments. A pilot deployment in two locations (Germany and Japan) is reported to demonstrate feasibility of the developed methodical approach. Descriptive observations indicate that the paradigm can be implemented across cultural groups and can elicit representational gestures during children's explanations. We discuss how social robots may further contribute to research on children's multimodal communicative behavior in interaction, along with methodical refinements needed for future studies. Overall, the developed design demonstrates high experimental control and applicability, offering promising principles for future research on children's gestures in interaction with social robots.

KEYWORDS

children's communication, cross-cultural comparison, dialogue design, gesture elicitation, gesture production, social robots

1 Introduction

Research on communicative development has increasingly highlighted its multimodal nature, broadening the scope beyond vocal behaviors to encompass nonverbal modalities such as eye gaze, facial expressions, and co-speech gestures (e.g., [Rohlfsing et al., 2019](#) for an overview). Within this line of research, gestures have been shown to be temporally and semantically synchronized with language production and comprehension serving multiple

interpersonal and intrapersonal functions in communicative practices (Kendon, 2004; McNeill, 1992; Wagner et al., 2014). In addition, it has been pointed out that young children represent events gesturally in different ways than adults (Sekine et al., 2018). This finding highlights the importance of considering gestures when interpreting children's communicative behaviors, particularly as they are still developing the ability to express events in more decontextualized and linguistically explicit ways.

Recognizing communication as a multimodal process, some studies highlight potential universal patterns in early gesture development, such as early pointing behavior across diverse cultural groups (Liszkowski et al., 2012). At the same time, a growing body of evidence suggests that cultural differences in children's gesture use emerge early, affecting both how often children gesture (Marentette et al., 2016; Salomo and Liszkowski, 2013) and how their gestures are formed (Iverson et al., 2008; Pettenati et al., 2012). These studies recognize communication as an inherently social act and deeply embedded in cultural practices. Consequently, cultural conventions are reflected in gestural behaviors (Kita, 2009, Kita and Özyürek, 2003).

Notably, empirically exploring gestures across cultures entails methodological challenges. These primarily concern the representativeness, controllability, and comparability of the procedures used to elicit gestures in children. While we will return to these challenges in more detail in Section 3, it is important to note that creating similar interactional settings for participants from different cultures is complex because interpersonal interaction is inherently contingent: How an interaction unfolds depends not only on the task but also on partner-related and situational factors, such as moment-to-moment responsiveness, relationship dynamics, and culturally formed expectations. Consequently, human interaction partners, including experimenters, have difficulty reproducing identical behaviors (e.g., timing, feedback, gaze, or subtle backchannels) across participants, even when following the same script (Fischer and Prondzinska, 2020). As a consequence, although instructions and stimuli remain constant, the same task in different cultural settings can alter the opportunities and demands for producing gestures. To control for such confounds, procedures aimed at eliciting children's communicative behavior often rely on highly constrained elicitation paradigms, such as video-based or cartoon-based tasks, in which dialogical interaction is deliberately limited (Gullberg, 2008; Holler, 2013). Yet this methodological control comes at a cost, because the degree of dialogical engagement itself, such as whether the task involves dialogue or monologue, can affect children's gestural and communicative behaviors independently of the intended manipulations (in adults: Bavelas, 1994; Beattie and Aboudan, 1994; Wray and Norbury, 2018; proposed for children: Sparaci and Volterra, 2025), and materials such as cartoons are not only semantically different from natural settings, but may also be interpreted differently depending on the cultural setting.

Against the background with the challenges, there is a need for methodological approaches that support systematic, culturally sensitive comparisons of children's communicative behavior in dialogical settings. In this work, we propose a set of design principles grounded in an interactionist perspective (presented in Section 4.1) to guide the use of social robots as standardized

interaction partners in cross-cultural research on children's gesture and speech production. Specifically, our goals are to contribute (i) a theory-informed approach to designing interactions that elicit children's gestural behaviors within structured, repeatable sequences and (ii) concrete principles for implementing this approach with a social robot while preserving key interactional contingencies. In the following section, we discuss potential methodological advantages of social robots, which make them useful research tools—their capacity for reproducible multimodal behavior (Breazeal, 2003; Rohlfsing et al., 2024; Skantze et al., 2019; Tolksdorf, 2024), their flexibility in taking on different dialogical roles (Lighthart et al., 2020; Rohlfsing et al., 2022a), and their acceptance by children as embodied social actors and trustworthy informants (Oranç and Küntay, 2020). While social robots do not fully replicate the authenticity of human-to-human communication, these affordances allow researchers to study children's behavior across a wide range of interaction scenarios that can be difficult to achieve with human partners. On this basis, we present a robot-based elicitation paradigm for studying preschoolers' gestural behavior across cultural environments.

2 Social robots as interaction partners and methodical tool

In recent years, a growing body of research has provided evidence for the use of social robots in supporting children's learning across various domains—including education, healthcare, and entertainment (Peter and van Straten, 2024; Rohlfsing et al., 2024). Compared to other digital technologies, such as tablets, social robots offer unique capabilities for enriching interaction through multimodal communicative signals, including gaze, gesture, and body posture (Konijn et al., 2020; Vogt et al., 2017; Vulchanova et al., 2017). Thus, social robots provide an embodied interactive platform that enables dialogical, socially embedded interactions (Tolksdorf and Schaffrath, 2026; Vollmer et al., 2016). In line with this, studies have shown that children tend to conform to social robots and view them as trustworthy sources of information (Belpaeme et al., 2018). Current applications of social robots are diverse, ranging from promoting language development (Conti et al., 2019; Kory Westlund et al., 2017; Vogt et al., 2019) and computational thinking (Ioannou and Makridou, 2018) to supporting socio-emotional communication skills (Chen et al., 2020; Leite et al., 2017).

Due to their embodiment and the ability to communicate through multiple modalities, social robots can involve children in interactions that simulate human-like interactions (Kory Westlund et al., 2017). In fact, a large body of research suggests that social signals produced by robots are interpreted in similar ways to those produced by humans (Ham et al., 2015). Building on this, emerging research has begun to explore a dual purpose of social robots: not only as potential social learning partners, but also as a methodical tool for studying language learning processes and interactive behavior in children (Rohlfsing et al., 2022a; Tolksdorf, 2024; Tolksdorf et al., 2022).

Social robots may be particularly valuable for studying children's communicative behavior across cultures. Unlike human experimenters, robots can reproduce identical sequences

of multimodal behavior across participants and experimental trials (Jouen et al., 2025). This high level of control and replicability can be especially advantageous in cross-cultural research, where applying comparable elicitation procedures is crucial. In particular, social robots can be programmed to deliver similar, yet linguistically and culturally appropriate, verbal prompts, accompanied by comparable gestural movements, to different cultural groups, thereby reducing variability in elicitation procedures (Fischer and Prondzinska, 2020; Hanulíková et al., 2026). As a result, social robots allow researchers to standardize interaction parameters and enhance the replicability and internal validity of cross-cultural comparisons (Gunes et al., 2022).

Beyond enhancing controllability and replicability, social robots allow for systematic manipulation of interactional variables such as their appearance, gender, and a range of multimodal behaviors, including gestures, the timing of backchannel signals, as well as gaze shifts—which are difficult for human experimenters to standardize across participants. This level of control bears potential to novel experimental designs aimed at investigating how specific multimodal features of an interaction partner influence communicative behavior across cultures. In a similar vein, an underexplored potential lies in the robot's ability to adopt diverse dialogical and social roles during interaction. Along these lines, social robots can be programmed not only to take on the role of a tutor or peer, but also as a novice or a learner, adapting the role configurations to the tasks and content of a given educational or experimental setting (Rohlfing et al., 2022a). This is particularly promising in light of the fact that many studies on cross-cultural differences in gestural behavior rely on asymmetrical interaction structures: Typically, one participant takes a more active role in managing the task or has privileged access to task-relevant information, whereas the other assumes a complementary role, coordinating with and responding to the partner's contributions. Such setups often result in interactions that focus on one individual's performance and are characterized by limited dialogical exchange, even when there are no explicit interactional restrictions (Holler, 2013). The possibility of programming social robots to fulfill different social and dialogical roles thus offers a unique opportunity to manipulate interactional demands systematically and study how children adapt their communicative behavior in response. By varying the robot's role within the interaction, researchers can further investigate how different interactional configurations may influence gesture use and other communicative strategies across cultures.

Yet, it is important to acknowledge the current technical limitations of social robots, which lack the temporal and semantic contingencies that characterize natural human interaction (Rohlfing et al., 2024). In fact, the way robots' interaction with children is currently implemented is limited with respect to smoothness of turn-timing (Tykhonenko et al., 2026). Crucially, however, evidence from human–robot interaction suggests that both children and adults tend to evaluate robots less normatively than other humans and are less likely to sanction or negatively interpret their behavioral peculiarities (Beraldo et al., 2019). This difference in social judgment holds methodical relevance because it can potentially buffer experimental interactions against confounding factors that commonly arise when responsiveness is necessarily constrained.

More specifically, in tightly controlled paradigms—where the partner may withhold clarification, provide limited emotional feedback, or refrain from interrupting—such restrictions can be experienced as interactional violations with human partners and may elicit irritation or disengagement in culturally specific ways (Lieven and Stoll, 2013). With a robotic partner, however, limited responsiveness can be attributed to the technological nature of the agent rather than to deviations from social interactional norms (Leite et al., 2013). In cross-cultural comparisons, this attribution shift may increase the likelihood that participants across cultural groups interpret the partner's constrained behavior in more similar terms, thereby reducing a risk that culturally variable expectations about interpersonal behavior confound observed multimodal behavior. In this way, social robots can be particularly compatible with controlled experimental designs. They can therefore provide a form of interaction in which the constraints required for standardization may be more readily tolerated by children, while still allowing for dialogical engagement. In sum, social robots therefore provide a promising methodical tool for investigating children's multimodal communicative behavior under conditions that require high experimental control, including in cross-cultural developmental research.

3 Methodical challenges in cross-cultural gesture research

An illustrative example of a typical methodical approach to investigate cross-cultural differences in children's gestural behavior is provided by Colletta et al. (2015), who employ a storytelling paradigm that is widely used in gesture research. In their study, children aged 5 and 10 from France, the United States, and Italy watched a wordless cartoon and were then asked to retell the story to an experimenter acting as an adult listener. The findings revealed that Italian children used more gestures during their narration than their French counterparts, who in turn gestured more than the U.S. children. While these results provide valuable insights into cross-cultural differences in children's gestural behavior, the study does not report further details about how the experimenter's multimodal behavior was controlled during the task. For example, it remains unclear whether the specific instructions given to the children, the extent and nature of the experimenter's contingent verbal and nonverbal responses, or the absence of such responses were consistently implemented across cultural contexts. Similar methods have also been used in other cross-cultural studies of gesture (e.g., Gullberg, 1998; Iverson et al., 2008; Sekine and Furuyama, 2010). Such limited reporting of experimenter behavior leaves open the possibility that variability in the interactional setting may have influenced the children's use of gestures, highlighting the broader challenge of ensuring consistency in experimenter behavior across cultural environments.

Methodical comparable challenges arise in cross-cultural gesture research with adult participants, where the similar paradigms are widely used. We briefly consider one such study to illustrate that the issues of standardization discussed above are not specific to children but extend to the broader use of narrative-based elicitation across cultures. Nicoladis et al. (2018), for example, employed a similar paradigm in a study comparing

participants across four cultural populations (Mandarin, Hindi, French, Spanish). Here, participants narrated a Pink Panther cartoon after watching the clip. Supporting the notion that gestural behavior can differ across cultures (see also [Graziano and Gullberg, 2024](#)), [Nicoladis et al. \(2018\)](#) found that Chinese and Hindi speakers gestured less frequently than French and Spanish speakers. The authors attributed this difference to group-related differences in storytelling style. Participants from Asian cultures were more likely to tell evaluative stories, whereas the participants who spoke Romance languages tended to use a chronicle style, which was associated with more gestures. Although using standardized video stimuli such as cartoons can provide a high degree of control over the visual content presented across cultures, the interpretation and familiarity of the depicted events may still vary between participants from different backgrounds ([Serafini, 2017](#)). Furthermore, these paradigms rest on the assumption that the broader social and interactional context is comparably experienced across cultural groups.

A particularly complex methodical issue in such approaches concerns the role of the human experimenter who typically acts as the participant's interaction partner during the task. Achieving control over this variable across cultures is inherently difficult, as even subtle variations in an experimenter's verbal or nonverbal behavior, or even their appearance or gender (e.g., [Charlesworth et al., 2019](#)), can influence both the frequency and forms of participants' communicative behaviors. This limitation is particularly salient given that co-speech gestures are highly context-dependent and multifunctional ([Wagner et al., 2014](#)), being shaped by the moment-to-moment dynamics of the interaction ([Holler et al., 2018](#)). While certain experimenter behaviors, such as scripted questions or instructions, can be reliably standardized across trials, other behaviors, such as facial expressions, gaze, intonation, or timing of feedback, are much more difficult to control and may be interpreted differently depending on cultural norms. Alongside story retelling tasks, picture-naming tasks are among the most widely employed paradigms in the cross-cultural study of children's gesture uses, reflecting a prevailing emphasis on structured elicitation formats and pointing to the potential value of expanding the range of interactional contexts used in this field. For example, [Marentette et al. \(2016\)](#) compared spontaneous representational gesture use in 2-year-old Italian and English-speaking Canadian children in a picture-naming task with 40 standardized images. They found higher co-speech gesture frequency in Italian children compared to English-speaking children. Similarly, [Cattani et al. \(2019\)](#) compared two-year-olds from Italy, the UK, and Australia and found that Italian children used more representational gestures than the other cultural groups, while pointing gestures were more frequent among Australian and Italian children than among British children. Moreover, several cross-cultural studies on gesture use have relied on the same task and report differences between cultural groups (e.g., [Huttunen et al., 2013](#); [So, 2010](#)). By contrast, other studies did not find differences between different cultural groups: For example, [Pettenati et al. \(2012\)](#) found no significant differences in co-speech gesture frequency between Italian and Japanese toddlers. Such inconsistencies may, in part, reflect subtle variation in how experimenters interact with children or deliver prompts—factors

that are difficult to standardize fully across cultures, even with carefully controlled procedures. Thus, achieving comparable interactional settings across cultures is complicated by natural intra- and interpersonal variability. In this regard, social robots offer a clear methodical advantage, because their multimodal behavior, including timing, gaze, verbal feedback, can be precisely specified and replicated. This enables a more controlled and consistent interactional frame across participants and cultures.

In addition to the role of the human experimenter, another issue is the task. In this vein, much of the existing empirical evidence on cross-cultural and cross-linguistic multimodal behavior is based on methodical approaches and elicitation procedures that rely on a limited number of interactive settings, primarily focused on tasks such as picture-naming or narrative retelling. While these paradigms have contributed valuable insights, this scope of elicitation procedures may also affect the range of gesture types that can be observed. In this vein, it can be argued that referential gestures, for example, are more likely to be elicited in naming tasks ([Samuelsson et al., 2025](#)), whereas more conversational settings may give rise to a broader spectrum of gestures, including pragmatic gestures, that is, those involved in discourse marking, interaction management, or emphasis ([Rohrer et al., 2020](#); [Ter Bekke et al., 2024a, b](#)). In fact, as [Efron \(1972\)](#) demonstrated, some cultural groups, such as Italian speakers, may tend to use more referential gestures to illustrate the content of their speech, while others, such as Yiddish speakers, may rely more on pragmatic gestures to structure the discourse. This underscores that gesture use is crucially shaped by the communicative demands of the interaction.

Taken together, existing cross-cultural work on children's co-speech gestures faces the recurring methodological challenges of achieving controllability of the experimenter's behavior across cultural environments and maintaining comparability of the interactional setting participants experience. In this way, we argued that the programmable and embodied nature of social robots addresses these challenges offering a promising, innovative methodical tool that supports repeatable, multimodal behavior across participants and cultural settings. At the same time, integrating social robots into a broader repertoire of experimental paradigms—including interactive, task-oriented designs—may enable researchers to gain deeper insights into the cultural, cognitive, and linguistic processes that shape multimodal communication in childhood.

In what follows, we outline developed design principles for how social robots can be employed to advance methodical practices in cross-cultural gesture research and potentially contribute to observing patterns of children's gestural behavior that are difficult to elicit using conventional approaches.

4 Design principles

Below, we describe our design principles for using a social robot to explore children's multimodal behavior in different cultural environments (Germany and Japan). We particularly focus on the task design for child-robot interaction with preschoolers to elicit their co-speech gestures in different cultures.

4.1 The interaction design: models guiding the methodology

For the interaction design, we adopted an interactionist view as the theoretical foundation for the principles. According to this view, a dialogue unfolds between participants acting jointly toward a goal (Rohlfing et al., 2016). We consider three aspects of the design that need to be examined in more detail:

- Task Model: The pragmatic circumstances that guide the sequences of actions toward a goal.
- Partner Model: The conception of the partner and whether they can contribute to the joint goal.
- Interaction Model: The meaningful way of exchange that both partners agree on.

For the interaction design, we first conceptualized these aspects in separation before considering their intertwining:

The idea of a Task Model is based on the assumption that in a dialogue, partners act jointly toward a goal. The concept of a goal is basic to human actions in the sense that actions are organized toward it (Csibra and Gergely, 2007, p. 60) and already young infants are sensitive to such a structure. A Task Model, thus, requires a mental representation of desired states, “where states are broadly construed as outcomes, events, or processes” (Austin and Vancouver, 1996, p. 338). This representation seems especially important in an asymmetric interaction, in which one less and one more knowledgeable partner acts jointly. In such a case, Wood et al. (1976) assign more responsibility for the joint goal to the more knowledgeable partner. They require the “expert” to have a model of the “task or problem and how it may be completed” (ibid., p. 97). Clearly, a Task Model intersects with a Partner Model because to recognize the role of an expert, the knowledge about what the partner can do (as a part of a concrete Partner Model) is necessary. In this regard, by informing about whether a partner can contribute to the joint goal to arrive at it, a Partner Model yields responsibilities in an interaction (Rohlfing et al., 2022b). In addition, the knowledge about what and how a partner can contribute to the joint goal is also informative for the Interaction Model. In talking to children, caregivers modify their interactive behaviors in a manifold way (see, e.g., Rohlfing et al., 2022b). These modifications result in a complex task being broken down into smaller units (e.g., Rohlfing et al., 2006) or gestural behaviors specifying parts of difficult tasks (Grimminger et al., 2010; Wray and Norbury, 2018). It is widely assumed that learners benefit from such modifications in their learning (Rohlfing et al., 2022b). In this sense, the Partner Model intersects with the Interaction Model, giving rise to such modifications. The latter informs about successful means of interaction which the partners use and on which they continue to agree as the interaction unfolds in order to arrive at the joint goal. Thus, it is possible that the interaction starts with means (e.g., verbal behaviors) that the partners need to adjust during the process (e.g., use more pointing gestures to specify crucial contents). This adjustment is important to emphasize because an ongoing interaction allows all Models to be adjusted driving their adaptive nature.

Although the models are considered separately in the design phase, they influence each other when applied in dialogue. In

our case, we decided that the goal of the designed interaction between the child and the robot is to convey knowledge about everyday actions to the robot, that is, to explain how to perform them correctly. Hereby, the Partner Model was crucial: With a cover story (see dialogue script available as [Supplementary Material](#)), the robot was introduced as a less knowledgeable partner, who wants to learn more about human actions. As suggested in Rohlfing et al. (2022b), the Partner Model is crucial to determine the dialogical role driving dialogical behaviors of a person. For example, it makes a difference whether a child is supposed to act as a learner or an expert. In our task, the robot asked for help revealing itself as a novice learner. At the same time, it made the children the informed “experts” in this task. Rohlfing et al. (2022a) emphasize that the design of the dialogical roles is crucial for children’s productive behaviors consisting of perceptual, cognitive, and dialogical capabilities that are applied in a dialogue. Despite a clear conception of the roles and the goal, i.e., what needs to be achieved in an interaction, the nature of the goal can influence the interaction itself. Overall, with the interaction, we aimed to elicit co-speech gestures in children’s interactional behaviors. From the literature, we derived some criteria about the stimulus: Participants gesture more when they have to recall sensorimotor knowledge (e.g., Hostetter and Alibali, 2008; Kita et al., 2017), after watching a video (Hostetter and Hopkins, 2002), and as a repair (Boden et al., 2024). Accordingly, for the Task Model, we designed a task that first allowed the children to perceive actions and generate sensorimotor knowledge from watching a video. Then, they recalled this knowledge when interacting with a communicative partner. Due to their role, the children were motivated to carefully observe the presented actions so that they could inform their interaction partner about them. Because of their adaptive nature, the intertwining of the Models, once applied in a dialogue, can give rise to specific dynamics. Therefore, it is necessary to simulate and critically reflect on the partner’s specific behavior in terms of collaboration. For instance, if the robot misinterprets actions repeatedly, children may deem the explanation insufficient and may discontinue providing it (Interaction Model).

4.2 Overview of the study design

4.2.1 Cross-cultural experimental design and operationalization of key interactional strategies

Building on the interactionist framework and the three intertwined Models introduced in the previous section—the Task Model, Partner Model, and Interaction Model—this study was designed to translate these conceptual principles into a structured yet ecologically valid experimental paradigm across cultural settings. The experimental paradigm involved preschool-aged children engaging in task-oriented dialogue with an interactive partner. Specifically, the Partner Model was operationalized by introducing a programmable humanoid robot as a novice learner. By doing so, the child was positioned as the teacher or expert. The approach of reversing the “classical” roles assigned to children in (language) learning studies and its application in child–robot–interaction is inspired by the work of

Tanaka and colleagues (Tanaka and Kimura, 2010, Tanaka and Matsuzoe, 2012) who introduced a learning-by-teaching paradigm for preschool-aged children. They argue that this not only serves as “indirect practice [...] for the children” (Tanaka and Matsuzoe, 2012, p. 80) but is also highly motivating for them. In our experimental design, the social robot introduced itself by initiating a dialogue with the child. After a short warm-up phase, during which the child guessed which animal the robot was imitating through its gestures, the robot asked the child to teach it new things, providing the following background: “I was just built, and I want to learn from you. I want to know about things that aren’t animals. Can you tell me a little more about them? [pause for child to answer] Ok then, I would like you to teach me how to [action]” (the full dialogue script is available as [Supplementary Material](#)). Thus, the robot acts as a willing learner of new actions who needs the child’s explanations.

In line with the Task Model, the interaction was centered around a shared pragmatic goal: enabling the child to convey knowledge about six everyday actions (e.g., making a paper airplane), thereby creating opportunities for spontaneous explanatory speech and gesture production. Each of the six trials followed a consistent interactional sequence illustrated in [Figure 1](#). After the child explained an action, the robot displays of partial understanding on an external monitor (a kind of a “thinking monitor”). For example, it was crumpling the paper instead of folding it when making a paper airplane (see [Figure 2](#), step 6). In this way, the child was convincingly given the role of a more knowledgeable partner, in line with the intended asymmetrical instructional roles.

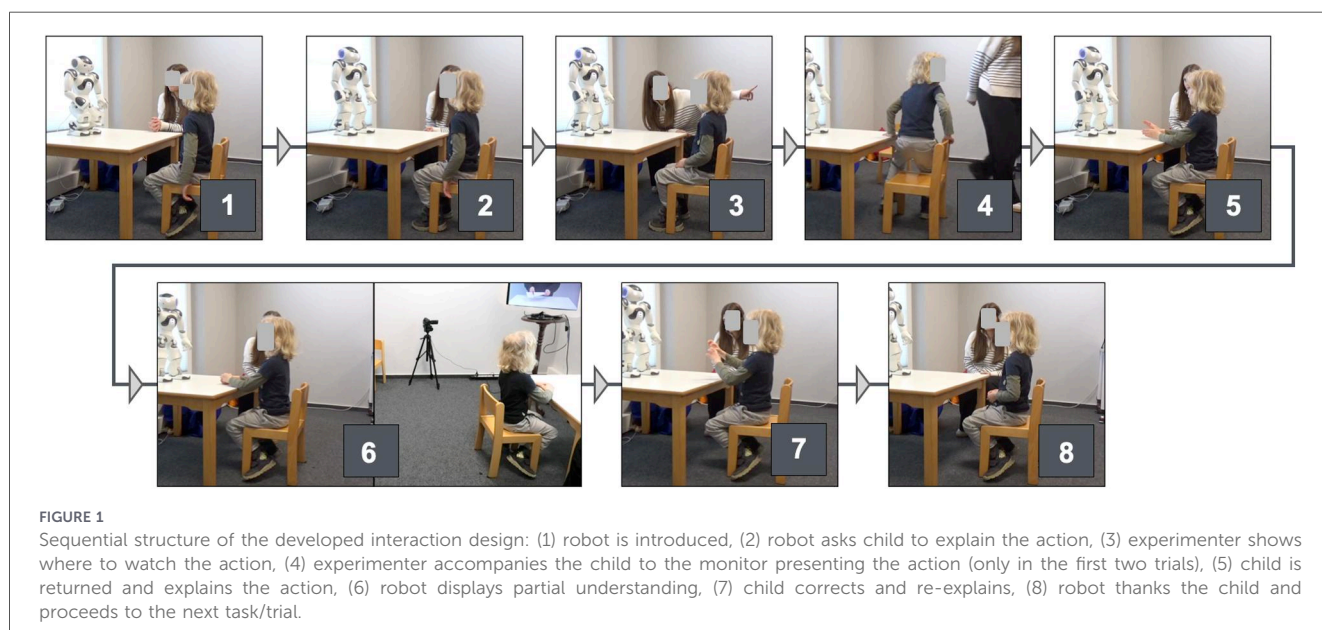
The choice of this sequence was motivated by prior findings: A previous study with preschool-aged children by Boden et al. (2024) investigated a similar task, in which children reacted to a hand puppet acting in a noncanonical manner. For example, the puppet was eating with a spoon but holding the spoon on the wrong end. When describing these actions to their caregivers, children often shared their judgments about the actions. The

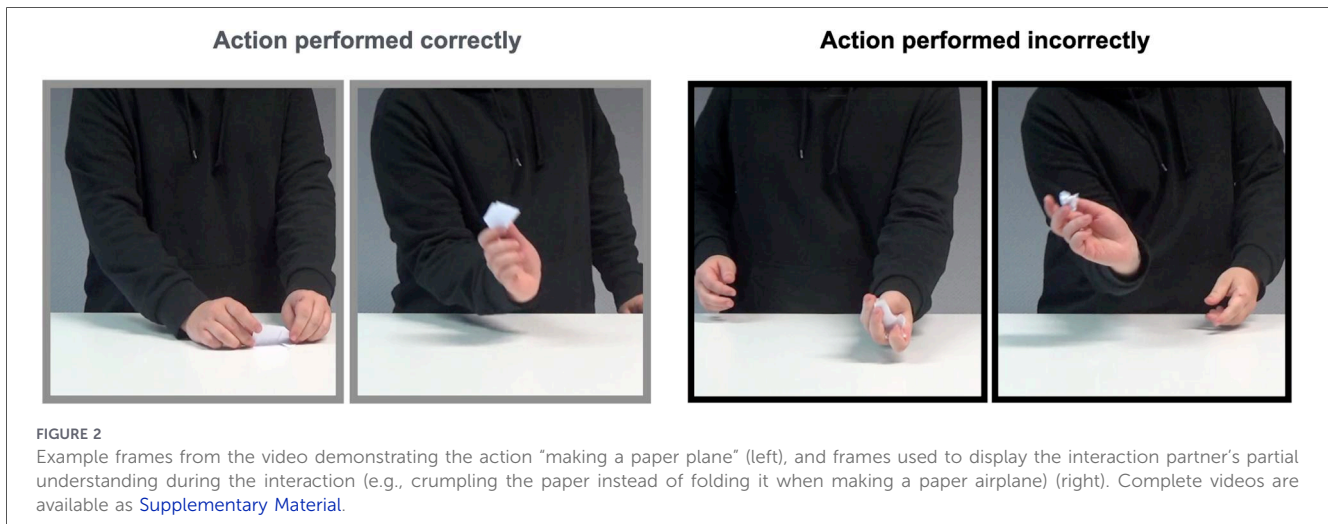
authors report that in comparison to a narration or an explanation, this task elicited co-speech gestures more frequently (Boden et al., 2024). In our design, the partial-understanding displayed on the thinking monitor adds an element of surprise around the robot’s understanding and supports the Partner Model of the robot being a learner who does not fully memorize the action explanation. To ensure that the child understood the thinking monitor as a representation of the robot’s understanding, it was visually framed by a thought bubble pointing toward the robot. After the child’s further explanation, the robot thanked them and proceeded to the next trial.

In this way, the design followed an interactionist perspective on a joint task construction (Rohlfing et al., 2016), emphasizing both the pragmatic framing of the interaction (the goal) and the sequential structure of actions. Importantly, this role assignment was critical in shaping the child’s expectations of the interaction and aligning their dialogical behaviors with the communicative goals of the task. The robot’s behavior was carefully scripted and timed (operated by a second experimenter, see [Section 4.2.3](#)) to ensure consistent interactional contingencies across children and cultures. This repetition of structured yet dynamic exchanges provided a high degree of experimental control while preserving key features of spontaneous communication. This allowed us to observe gesture use not only as part of initial explanations but also as a means of clarification or repair.

4.2.2 Cultural sensitivity in materials and procedures

In designing a paradigm that is deployable across cultural settings, particular attention was paid to the development of the task content, materials, and the communicative roles assigned within the interaction. The core instructional task—explaining how to do familiar everyday actions (e.g., making a paper airplane)—was selected for its cultural neutrality and accessibility to preschool children in both Germany and Japan (see [Figure 2](#)). Similarly, the specific





everyday actions were extensively discussed and carefully chosen to allow for cross-cultural applicability.

Vocabulary and linguistic framing of the dialogue was similar for both cultures; however, small adjustments were made in each language version to ensure age-appropriate comprehension without introducing culturally loaded terms. Moreover, the decision to frame the child as the informed expert and the robot as the novice learner was based on prior research demonstrating children's readiness to take on their role as an explainer across cultures (Jurkat et al., 2022). This asymmetry was designed to elicit a natural pedagogical motivation from children without relying on culturally specific classroom conventions. Finally, in order to ensure comparability across cultural environments, care was taken to avoid interactional norms that might favor one cultural group over another (e.g., high-context indirectness vs. explicit correction styles).

4.2.3 Employed social robot

To ensure methodical consistency, we utilized the same robotic platform across cultural groups: the humanoid Nao robot, developed by SoftBank Robotics (see Figure 1). This type of social robot is the most widely used robotic platform in the field of child–robot interaction (Belpaeme et al., 2018). It has also proven to be an accepted interaction partner in various cultural settings (Belpaeme et al., 2018). The 58-cm-tall robot has 25 degrees of freedom and is capable of executing a variety of expressive behaviors. All behaviors were implemented using *Choregraphe*, a visual programming environment developed for the Nao platform. Drawing on prior work (Rohlfing et al., 2022a; Tolksdorf, 2024; Tolksdorf et al., 2024), we designed the robot's dialogical behavior to be both multimodal and context-sensitive. To maintain a physical presence throughout the interaction, the robot continuously displayed its built-in "breathing" animation—subtle, periodic movements of the arms, legs, and neck joints that prevent the robot from appearing static and enhance its embodied naturalness (Tolksdorf, 2024).

In addition to this passive animation, we programmed the social robot to produce specific multimodal behaviors during key moments of the interaction sequence described in Section 4.2.1 and illustrated in Figure 1. At the partial-understanding slot, the robot produced the utterance "OK! This is what I understood how to [e.g., make a paper

airplane]" accompanied by a reduced iconic gesture of the target action, followed by "Is that correct?" with a pointing gesture and a head turn toward the thinking monitor, which simultaneously displayed the incorrect version of the action. The iconic gesture was adjusted to reflect the action being explained; the full dialogue is provided in the Supplementary Material.

To enhance the naturalness while preserving standardization, we implemented a semi-autonomous control approach (Maure and Bruno, 2025) in which the robot's behavior was initiated at predefined points in the interaction via teleoperation by a human operator. This enabled consistent presentation of scripted dialogue while allowing for contingent timing and smooth turn-taking across participants (Vogt et al., 2017). Specifically, the robot's multimodal actions followed a fixed dialogue script and pre-specified gestures (see Supplementary Material). To minimize variability within and across cultural environments, the operator's role was limited to selecting from a set of scripted behaviors that advanced the interaction along the same sequence in every trial, maintaining the core contingencies (request to explain → child explanation → display of partial misunderstanding → opportunity for repair). In this way, teleoperation was used to ensure that designed behaviors occurred at the appropriate times, not to modify what the robot expressed. Operators from both cultural environments were trained by practicing cues and responses in interactions with other lab members to further ensure consistent robot behavior.

4.2.4 Pilot sample and conditions

To assess the feasibility of the designed interaction paradigm, we conducted a small pilot deployment in both cultural settings. Participants included nine preschoolers. The German sample included five children ($Mage = 5.13$ years, $SD = 0.25$; 2 girls, 3 boys) and the Japanese sample included four children ($Mage = 5.15$ years, $SD = 0.70$; 1 girl, 3 boys); overall $Mage$ was 5.14 years ($SD = 0.50$). All participants were monolingual speakers of the local language, and no children had prior experience interacting with social robots. The pilot study was conducted in accordance with institutional guidelines and local ethical and legal regulations. Given its focus on methodical development and feasibility, the pilot work was covered by the general child–robot

study procedures reviewed and approved by the Ethical Committee of Paderborn University (05/2025). Prior to participation, written informed consent was obtained from all legal guardians, and children provided age-appropriate assent.

The pilot was designed to evaluate procedural aspects (e.g., children’s comprehension of the task, completion of trials, and whether the interaction reliably created opportunities for gesture production), not to support inferential tests of cross-cultural differences.

Each child completed six trials and was assigned to one of two counterbalanced orders to include a non-robot comparison partner within the same task structure: (i) Robot-first: three trials with the humanoid robot as novice partner, followed by three trials with a hand puppet; (ii) Puppet-first: the reverse order. The puppet condition was included to support future work on partner effects under otherwise matched interactional contingencies. However, given the scope of this methods paper, we do not analyze or interpret this condition further, as our focus is on specifying and documenting the robot-based elicitation approach for cross-cultural gesture research, as well as illustrating its feasibility and implementation across sites.

4.2.5 Developed experimental setup

The experimental setup was designed to be child-oriented and culturally adaptable, with a clear spatial arrangement to support face-to-face interaction between the child and the interactive partner. As illustrated in Figure 3, the child was seated in front of the robot. Additionally, an experimenter was positioned at a 90-degree angle to the right of the child. Before the child interacted with the robot, the experimenter greeted the child and their caregiver, accompanied them to their seats, briefly introduced the

robot as well as the thinking monitor (next to the NAO robot) displaying the robot’s understanding, and explained that the robot needed the child’s help. During the first two trials, the experimenter showed and encouraged the child to do the following: when the robot asked if the child could teach an action, such as making a paper airplane, the child could first watch the corresponding action on the monitor on the right side of the experimental room and then return to the robot. High-definition video and audio recordings captured the child’s multimodal responses. A second experimenter, acting as the robot operator responsible for the teleoperation of the robot, remained in the background and ensured that the videos were displayed on the monitors.

5 Pilot observations and feasibility of the developed methodical approach

The interactive design principles guiding our methodical approach were developed to elicit children’s multimodal communication behavior, particularly their gestural behavior, in structured yet naturalistic interactions. In our pilot deployment, our goal was to assess feasibility of cross-site implementation and to illustrate the types of behaviors the paradigm can yield. Accordingly, all pilot results are reported descriptively and are interpreted as indicators of feasibility (e.g., whether children produced spontaneous co-speech gestures during intended elicitation at instructional slots and whether the robot’s displayed partial understanding created observable opportunities for a repair, i.e., modified second explanation). Below, we report on two main observations: (a) a descriptive overview of the distribution of representational gestures produced by children across cultures and action contexts when acting as experts and

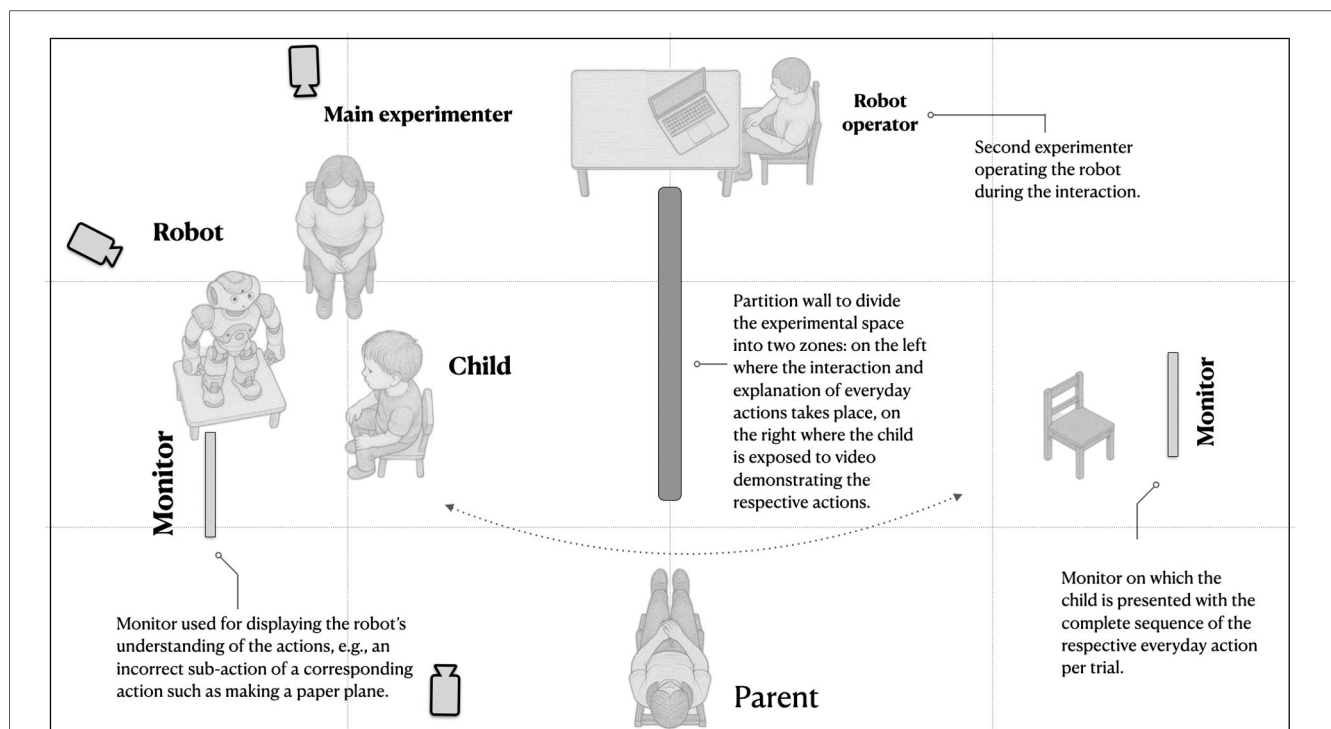


FIGURE 3 Schematic representation of the developed design aiming to elicit the child’s gestural communicative behavior during interaction.

explaining the presented actions to a partner; and (b) observations of corrective gestural behavior in response to a partner’s perceived partial understanding.

5.1 Distribution of iconic gestures across cultures and task contexts

Descriptive results suggest that the developed method successfully elicited iconic gestures across both cultural groups and while explaining different everyday actions, as illustrated in Figure 4. Mean frequencies of iconic gestures are further represented in Table 1.

Importantly, although there was noticeable variability in the children’s gestural behavior depending on the action being explained, all of the children at both sites produced representational gestures during their explanations. This indicates that the developed interaction successfully elicited gestural behaviors. Due to the limited scope of the pilot study, we do not further interpret differences in descriptive means across single actions or countries. Instead, we consider these observations to be preliminary illustrations of behavioral patterns and task-dependent variability that may be examined in future, adequately powered studies.

5.2 Corrective gestural behavior in response to displayed partial understanding

A further feasibility consideration was whether the robot’s scripted partial understanding created opportunities for children

to repair or clarify their explanations. In line with the developed interaction design, these partial-understanding displays were intentionally implemented to provide opportunities for children to modify previously explained actions. We defined repair initiation as any verbal clarification or re-explanation of a sub-step of an action and/or the addition of a gesture, that occurred after the display of a partial or incorrect understanding (e.g., re-enacting a fold in the paper airplane task). Overall, in the pilot, we observed that all children could recognize their interaction partner’s incorrect understanding, and most children used gestures to address it. However, their responses were often focused on a specific sub-action (e.g., re-explaining just one fold in the paper airplane action) rather than reiterating the entire

TABLE 1 Descriptive data of mean occurrences of iconic gestures across cultural groups (JP, Japan; GER, Germany) and actions.

Action	JP		GER	
	M	(SD)	M	(SD)
Read a book	2.0	2.8	2.0	1.0
Prepare pickles	1.0	2.0	1.6	1.5
Pour water in a glass	1.3	1.5	1.6	3.0
Make a juice	1.3	1.0	1.4	1.5
Making a paper airplane	2.8	2.5	3.4	3.4
Cut out shape with scissors	2.3	2.6	1.0	1.4
Total per participant	10.5	8.3	11.0	10.7

The bottom row reports the mean (and SD) of total gestures produced per participant across all six actions.

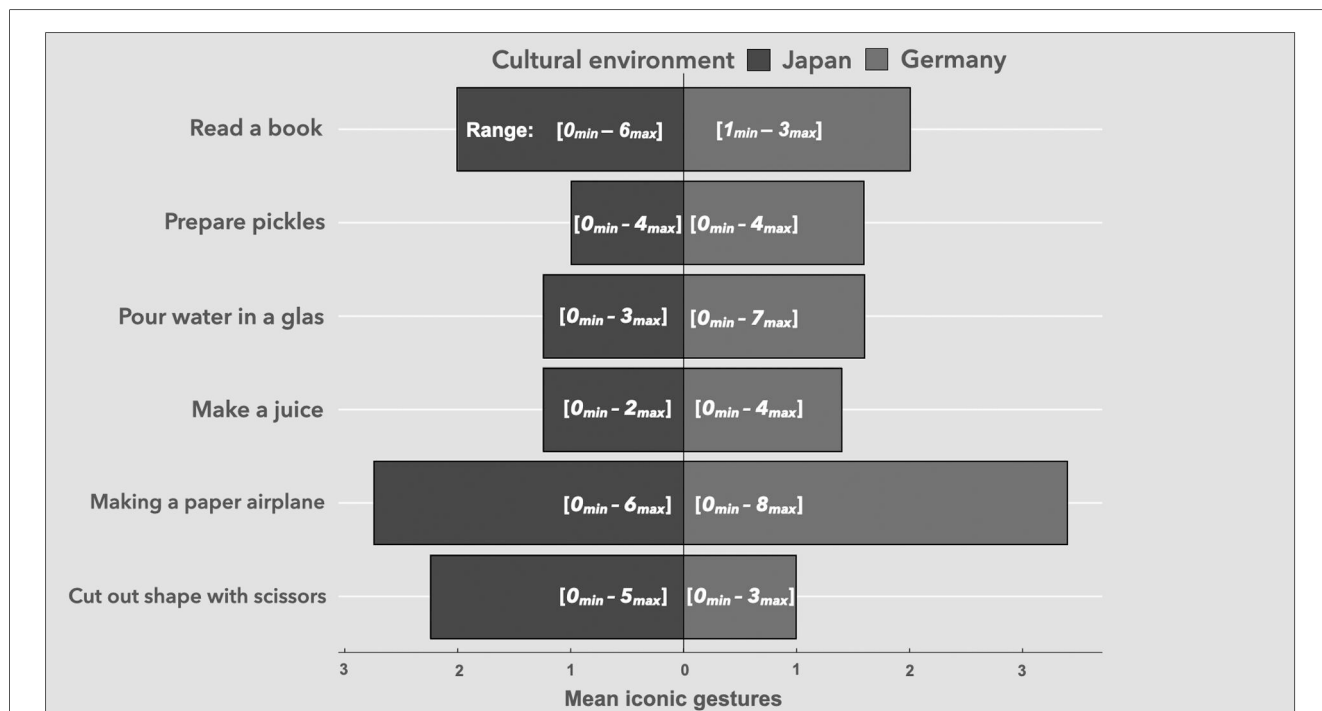


FIGURE 4 Overview of the distribution of children’s iconic gestures across cultures and action settings. Bar length represents the mean number of gestures produced for each action, shown separately for the Japanese (dark grey) and German (light grey) samples. Values in brackets indicate the observed range (minimum–maximum) across children within each cell.

action sequence. This suggests that their corrective behavior was targeted and sensitive to what the robot should already know. We also noted that children's corrective behavior diminished over the course of the interaction. In some cases, children initially attempted to clarify incorrectly understood actions but later discontinued doing so. This may align with the adaptive nature of the established Interaction Model: repeated unsuccessful attempts to clarify may have caused children to question the efficacy of their explanations, leading them to withhold further corrections. We will elaborate on possible interpretation of this behavior in the Discussion.

In addition to these action-related behaviors, the pilot recordings illustrate that our method can also shed light on some details on gestural behaviors. For example, children frequently used a small gesture space, often keeping their gestures close to the body (see Figure 5).

Self-adapters, such as touching the face or clothing, were also noted, which may indicate information processing and speech production and tend to appear in the presence of novel environmental factors, for example an unknown interaction partner (Allen and Honeycutt, 1997; Ekman and Friesen, 1969). Although not explicitly intended as part of the design, a small number of children also used deictic gestures—such as pointing to the monitor where the incorrect understanding was displayed—to guide the partner's attention during corrections. These instances seem to reflect how some children flexibly employed available resources when they explicated their reference.

Taken together, the pilot observations indicate that the proposed method elicits various gestural behaviors within a structured dialogical sequence. In addition, beyond demonstrating feasibility, the design also allows to examine children's multimodal behavior when an interaction partner signals partial understanding.

6 Discussion

In this work, we proposed and piloted a theory-informed methodical approach for eliciting preschool-aged children's multimodal communicative behaviors, particularly their co-speech gestures, using a social robot as a standardized interaction partner in Japan and Germany—countries that differ largely in their culture. Guided by an interactionist theoretical perspective, the approach specifies a structured dialogical

paradigm in which children adopt an expert role and the robot fulfills a novice-learner role, creating repeatable opportunities for explanation and repair under conditions of high degree of experimental control and cultural comparability. Our pilot study demonstrated the feasibility of implementing the procedure and provided preliminary observations of children's behaviors that the paradigm can yield. The cross-cultural applicability of the proposed approach is supported by the initial findings that all participating children in both cultural groups produced co-speech gestures during their explanations and engaged in the designed interaction, although the frequency of gestures varied across actions and children.

A key strength of the presented approach lies in its capacity to initiate a dialogue and interaction with children while ensuring a high degree of experimental control. By scripting the robot's multimodal behavior and maintaining consistent interactional conditions across trials (i.e., the different actions), we were able to standardize the child's dialogical experience and to elicit spontaneous communicative behaviors at the same time. This level of control bears potential in cross-cultural gesture research where consistent elicitation procedures are essential to minimize confounds and enhance data comparability. The robot's ability to maintain repeatable yet dynamic interactional patterns created an interactional setting that would be difficult to achieve with a human partner. By assigning the child the role of an expert responsible for teaching the robot, we established a dialogical structure that has been shown to promote rich communicative behavior in children (Boden et al., 2024). Specifically, the applied role configuration operationalized all three aspects of our design principles (Section 4.1): the Partner Model framed the robot as a less knowledgeable learner in pragmatic need of the child's expertise. The Task Model embedded a clear, goal-directed instructional sequence designed to support knowledge transfer and mutual understanding. The Interaction Model shaped the exchange around cooperative turn-taking and appropriate communicative means. The systematically embedded instances in which the robot displayed partially understanding further strengthened these intertwined Models.

Further, our study was insightful for future methodical refinement of the here proposed approach. In this respect, we noted that the partial understanding that the robot demonstrated as a fixed response strategy can lead to crucial changes in the Partner Model that children hold about the robot: If the robot barely demonstrates an understanding, children might consider it



FIGURE 5

Example of the gestural behavior of a child explaining the sequence of actions involved in making a paper airplane.

as less capable of learning and reduce their effort in explaining and the use of corrective gestures. Our initial observation suggests that the repeated lack of acknowledgment or improvement in the robot's understanding may have disrupted the expected interactive dynamics. This observation stresses the importance of aligning the robot's behavior with both the child's role expectations and the perceived progression of mutual understanding. To address this issue, we propose refining the current dialogue design by introducing variations in the robot's responses that would reflect a learner's progress. For instance, if a child provides a successful correction, the robot could subsequently demonstrate the correct action on the monitor and provide verbal feedback indicating improved understanding (e.g., "Oh, now I see how to do it!"). Such behavior might allow children to recognize how their explanations affect the unfolding interaction. This could potentially increase their engagement by confirming their role as competent explainers. In addition, this adaptation would also enable the investigation of how children monitor and respond to their partner's learning trajectory, a dimension closely tied to the Partner Model and relevant for understanding how communicative behavior unfolds in asymmetric instructional settings (Agrawal et al., 2023).

One limitation to this approach is that children did not perform actions together with the robot. Thus, the current paradigm does not capture situations in which dialogue is directly coordinated with ongoing joint action. Future work that includes such joint action may require an additional Situation Model to link what is said to what is being done at each point in the interaction. This would require a tight coordination between the dialogue and actions, as illustrated by Singh et al. (2025).

Another limitation pertains to the technical realization of the design principles. In the current implementation, the robot's behavior was controlled via a semi-autonomous teleoperation setup. Although the multimodal behaviors of the robot followed a fixed script, and the operator's role was limited to triggering predefined actions at specific points in the sequence of the interaction, teleoperation can introduce variability related to the experimenter (e.g., response timing) and, in cross-cultural deployments, the possibility of cultural-specific operator effects (Gamboa et al., 2025). Future iterations therefore could benefit from increased levels of automation. For example, automated detection of relevant child behaviors could trigger the next scripted robot action in a more uniform manner across settings. Currently, however, fully automated solutions are not feasible. Despite rapid progress in child speech recognition and vision-language models (Janssens et al., 2025), the real-time interpretation of children's variable speech and spontaneous gestures remains limited. For this reason, a semi-autonomous approach offers a reliable compromise, supporting contingency while maintaining control over the dialogical structure and content.

Thus, whereas the here proposed interactional dynamics appeared to be effective in prompting children to revise or clarify their explanations, further research needs to be conducted to clarify what behaviors of the partner engage children in interaction and when they disengage. For this, a delicate balance between the partner demonstrating understanding and misunderstanding is required to keep children engaged. The achieved results and the open questions make the paradigm suitable as a methodical tool for not only

eliciting representational gestures during a child's explanations, but also for examining how children flexibly adapt their multimodal communicative means in response to incorrect understandings, in line with the adaptive dynamics postulated by the interactionist framework.

In addition, beyond cross-cultural applications, the presented standardized elicitation paradigm may also be useful in distinguishing potential sources of variability in children's multimodal behavior. For instance, when studying gesture production in heterogeneous developmental groups, an established comparable task structure and interactional contingencies can help distinguish effects attributable to the elicitation context from those related to developmental differences. Thus, the present robot-based paradigm could serve as a methodological baseline for future work explicitly examining different, heterogeneous developmental groups under matched interactional conditions.

Moreover, a promising direction for future work involves manipulating children's prior knowledge of the actions. Informal feedback from caregivers during the pilot study revealed that some children were already familiar with specific actions, such as making a paper airplane, and showed an increased use of gestures when explaining these actions. This suggests that action-specific familiarity could be related to children's multimodal explanations. In future iterations of the method, we plan to introduce a pre-lab practice phase in which children prepare selected actions at home with their caregivers. This could allow investigating the effects of prior action experience contributing to strengthening children's role as an expert, their better understanding of the actions, and, thus, promoting more elaborate explanations during subsequent laboratory interactions. In this vein, previous studies have shown that children's gesture use increases when they have opportunities to encode and reflect on actions before communicating them (Hostetter and Alibali, 2008; Wakefield et al., 2018).

7 Conclusion and future directions

In conclusion, the method presented here provides a new methodological approach for using social robots across cultural groups to elicit and investigate children's multimodal behavior. It also has broader implications for research with social robots in educational settings and the design of future child-robot interactions. While the presented pilot study was limited in scope and sample, it illustrates the potential of the proposed theory-informed, structured interaction design that can be implemented across two cultural environments and elicits spontaneous co-speech gestures during children's explanations within structured elicitation slots. Importantly, the presented method provides a level of control, repeatability, and adaptability that can address core methodological challenges in the field. At the same time, our initial observations point to the importance of refining the robot's behavior so that it aligns more closely with the child's expectations and the dynamics of the interaction, rather than providing a fixed response strategy. Future research will need to systematically investigate how such adjustments, e.g., displaying a progress in understanding, affect children's behavior.

In sum, we see the developed methodological approach not as a replacement for existing paradigms, but as a complementary tool

that can expand the experimental repertoire available to researchers studying children's communicative development across cultures. By enabling dynamic yet controlled interactions, social robots may contribute to a more nuanced understanding of how multimodal behaviors emerge and adapt within social and cultural interactional settings.

Data availability statement

Based on the consent given by the caregivers for the future reuse of their data by other researchers, the datasets reported as part of the pilot study are not readily available and will not be made publicly available. Requests to access the datasets should be directed to Katharina J. Rohlfing (katharina.rohlfing@upb.de).

Ethics statement

The studies involving humans were approved by the Ethical Committee of Paderborn University (05/2025). The studies were conducted in accordance with the local legislation and institutional requirements. The pilot study was conducted in accordance with institutional guidelines and local ethical and legal regulations. Given its focus on methodical development and feasibility, the pilot work was covered by the general child-robot study procedures reviewed and approved by the Ethical Committee of Paderborn University (05/2025). Written informed consent for participation in this study was provided by the participants' legal guardians.

Author contributions

NT: Conceptualization, Methodology, Investigation, Writing – original draft, Writing – review & editing. KR: Conceptualization, Methodology, Investigation, Resources, Project administration, Funding acquisition, Writing – original draft, Writing – review & editing. AG: Conceptualization, Methodology, Investigation, Writing – original draft, Writing – review & editing. KS: Conceptualization, Methodology, Investigation, Resources, Project administration, Funding acquisition, Writing – original draft, Writing – review & editing.

Funding

The author(s) declared that financial support was received for this work and/or its publication. The research was supported by the Deutscher Akademischer Austauschdienst e.V. (DAAD).

References

Agrawal, A., Liu, J., Bodur, K., Favre, B., and Fourtassi, A. (2023). Development of multimodal turn coordination in conversations: evidence for adult-like behavior in middle childhood. In *Proceedings of the Annual Meeting of the Cognitive Science Society*, 3275–3282. Available online at: <https://escholarship.org/uc/item/6pw6b3x9> (Accessed June 20, 2025).

Acknowledgments

We gratefully thank all the children and parents who participated in this study. We would also like to thank the students who contributed to this project: Koki Honda, Yoshika Takaki, Jonna Wilkens, and Tim Schaffhauser. In addition, we acknowledge support for the publication cost by the Open Access Publication Fund of Paderborn University. Finally, the design of the study benefitted immensely from discussions within the working group “gesture studies in children” with Paula Marentette, Elena Nicoladis, Bayimkhanim Bahmani.

Conflict of interest

The author(s) declared that this work was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The author AG declared that they were an editorial board member of Frontiers, at the time of submission. This had no impact on the peer review process and the final decision.

Generative AI statement

The author(s) declared that generative AI was not used in the creation of this manuscript.

Any alternative text (alt text) provided alongside figures in this article has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fcomm.2026.1651149/full#supplementary-material>

Allen, T. H., and Honeycutt, J. M. (1997). Planning, imagined interaction, and the nonverbal display of anxiety. *Commun. Res.* 24, 64–82. doi: 10.1177/009365097024001003

Austin, J. T., and Vancouver, J. B. (1996). Goal constructs in psychology: structure, process, and content. *Psychol. Bull.* 120, 338–375. doi: 10.1037/0033-2909.120.3.338

- Bavelas, J. B. (1994). Gestures as part of speech: methodological implications. *Res. Lang. Soc. Interact.* 27, 201–221. doi: 10.1207/s15327973rlsi2703_3
- Beattie, G., and Aboudan, R. (1994). Gestures, pauses and speech: an experimental investigation of the effects of changing social context on their precise temporal relationships. *Semiotica* 99, 239–272. doi: 10.1515/semi-1994-993-402
- Belpaeme, T., Kennedy, J., Ramachandran, A., Scassellati, B., and Tanaka, F. (2018). Social robots for education: a review. *Sci. Robot.* 3, eaat5954. doi: 10.1126/scirobotics.aat5954
- Beraldo, G., Menegatti, E., De Tommasi, V., Mancin, R., and Benini, F. (2019). A preliminary investigation of using humanoid social robots as non-pharmacological techniques with children. In *2019 IEEE International Conference on Advanced Robotics and its Social Impacts (ARSO)*, Beijing, China: IEEE, 393–400. doi: 10.1109/ARSO46408.2019.8948760
- Boden, U. J., Kern, F., Koutalidis, S., Abramov, O., Nemeth, A., Kopp, S., et al. (2024). Four-year-olds' visuospatial cognitive abilities and their relation to observer-viewpoint gestures across three communicative tasks. *Pragmat. Cogn.* 31, 49–96. doi: 10.1075/pc.00040.roh
- Breazeal, C. (2003). Toward sociable robots. *Rob. Auton. Syst.* 42, 167–175. doi: 10.1016/S0921-8890(02)00373-1
- Cattani, A., Floccia, C., Kidd, E., Pettenati, P., Onofrio, D., and Volterra, V. (2019). Gestures and words in naming: evidence from crosslinguistic and crosscultural comparison. *Lang. Learn.* 69, 709–746. doi: 10.1111/lang.12346
- Charlesworth, T. E. S., Hudson, S. T. J., Cogsdill, E. J., Spelke, E. S., and Banaji, M. R. (2019). Children use targets' facial appearance to guide and predict social behavior. *Dev. Psychol.* 55, 1400–1413. doi: 10.1037/dev0000734
- Chen, H., Park, H. W., and Breazeal, C. (2020). Teaching and learning with children: impact of reciprocal peer learning with a social robot on children's learning and emotive engagement. *Comput. Educ.* 150, 103836. doi: 10.1016/j.compedu.2020.103836
- Colletta, J.-M., Guidetti, M., Capirci, O., Cristilli, C., Demir, O. E., Kunene-Nicolas, R. N., et al. (2015). Effects of age and language on co-speech gesture production: an investigation of French, American, and Italian children's narratives. *J. Child Lang.* 42, 122–145. doi: 10.1017/S0305000913000585
- Conti, D., Cirasa, C., Di Nuovo, S., and Di Nuovo, A. (2019). "Robot, tell me a tale!": a social robot as tool for teachers in kindergarten. *Interact. Stud.* 20, 1–16. doi: 10.1075/is.18024.con
- Csibra, G., and Gergely, G. (2007). Obsessed with goals': functions and mechanisms of teleological interpretation of actions in humans. *Acta Psychol. (Amst)* 124, 60–78. doi: 10.1016/j.actpsy.2006.09.007
- Efron, D. (1972). *Gesture, Race and Culture*. Den Haag: Mouton and Co.
- Ekman, P., and Friesen, W. V. (1969). Nonverbal leakage and clues to deception. *Psychiatry* 32, 88–106. doi: 10.1080/00332747.1969.11023575
- Fischer, K., and Prondzinska, A. D. (2020). Experimental contrastive pragmatics using robots. *Contrastive Pragmat.* 1, 82–107. doi: 10.1163/26660393-BJA10004
- Gamboa, M., Thunberg, S., Alves-Oliveira, P., and Loerakker, M. B. (2025). "We are the robots: tapping into the lived experiences of Wizards of Oz," in *Proceedings of the Extended Abstracts of the CHI Conference on Human Factors in Computing Systems (Yokohama: ACM)*, 1–8. doi: 10.1145/3706599.3720149
- Graziano, M., and Gullberg, M. (2024). Providing evidence for a well-worn stereotype: italians and Swedes do gesture differently. *Front. Commun.* 9, 1314120. doi: 10.3389/fcomm.2024.1314120
- Grimminger, A., Rohlfing, K. J., and Steneken, P. (2010). Children's lexical skills and task demands affect gestural behavior in mothers of late-talking children and children with typical language development. *Gesture* 10, 251–278. doi: 10.1075/gest.10.2-3.07gri
- Gullberg, M. (1998). *Gesture as a Communication Strategy in Second Language Discourse: A Study of Learners of French and Swedish*. Lund: Lund Univ. Press.
- Gullberg, M. (2008). "Gestures and second language acquisition," in *Handbook of Cognitive Linguistics and Second Language Acquisition*, eds. P. Robinson, and N. C. Ellis (London: Routledge), 286–315.
- Gunes, H., Broz, F., Crawford, C. S., der Pütten, A. R., Strait, M., and Riek, L. (2022). Reproducibility in human-robot interaction: furthering the science of HRI. *Curr. Robot. Rep.* 3, 281–292. doi: 10.1007/s43154-022-00094-5
- Ham, J., Cuijpers, R. H., and Cabibihan, J.-J. (2015). Combining robotic persuasive strategies: the persuasive power of a storytelling robot that uses gazing and gestures. *Int. J. Soc. Robot.* 7, 479–487. doi: 10.1007/s12369-015-0280-4
- Hanulíková, A., Tolksdorf, N. F., and Kapp, S. (2026). Robot speech: how variability matters for child-robot interactions. *Front. Robot. AI* 12, 1725423. doi: 10.3389/frobot.2025.1725423
- Holler, J. (2013). "Experimental methods in co-speech gesture research," in *Handbooks of Linguistics and Communication Science (HSK) 38/1*, eds. C. Müller, A. Cienki, E. Fricke, S. Ladewig, D. McNeill, and S. Tessendorf (Berlin: De Gruyter), 837–857. doi: 10.1515/9783110261318.837
- Holler, J., Kendrick, K. H., and Levinson, S. C. (2018). Processing language in face-to-face conversation: questions with gestures get faster responses. *Psychon. Bull. Rev.* 25, 1900–1908. doi: 10.3758/s13423-017-1363-z
- Hostetter, A. B., and Alibali, M. W. (2008). Visible embodiment: gestures as simulated action. *Psychon. Bull. Rev.* 15, 495–514. doi: 10.3758/PBR.15.3.495
- Hostetter, A. B., and Hopkins, W. D. (2002). The effect of thought structure on the production of lexical movements. *Brain Lang.* 82, 22–29. doi: 10.1016/S0093-934X(02)00009-3
- Huttunen, K. H., Pine, K. J., Thurnham, A. J., and Khan, C. (2013). The changing role of gesture in linguistic development: a developmental trajectory and a cross-cultural comparison between British and Finnish children. *J. Psycholinguist. Res.* 42, 81–101. doi: 10.1007/s10936-012-9205-7
- Ioannou, A., and Makridou, E. (2018). Exploring the potentials of educational robotics in the development of computational thinking: a summary of current research and practical proposal for future work. *Educ. Inf. Technol.* 23, 2531–2544. doi: 10.1007/s10639-018-9729-z
- Iverson, J. M., Capirci, O., Volterra, V., and Goldin-Meadow, S. (2008). Learning to talk in a gesture-rich world: early communication in Italian vs. American children. *First Lang.* 28, 164–181. doi: 10.1177/0142723707087736
- Janssens, R., Verhelst, E., Abbo, G. A., Ren, Q., Bernal, M. J. P., and Belpaeme, T. (2025). "Child speech recognition in human-robot interaction: problem solved?," in *Social Robotics*, eds. O. Palinko, L. Bodenhausen, J.-J. Cabibihan, K. Fischer, S. Šabanovic, K. Winkle, et al. (Singapore: Springer Nature Singapore), 476–486. doi: 10.1007/978-981-96-3519-1_43
- Jouen, A.-L., Matsunaka, R., and Hiraki, K. (2025). Once upon a time... acquisition of second language vocabulary through robotic storytelling in classroom settings. *Int. J. Soc. Robot.* 17, 955–988. doi: 10.1007/s12369-025-01253-2
- Jurkat, S., Iza Simba, N. B., Hernández Chacón, L., Itakura, S., and Kärtner, J. (2022). Cultural similarities and differences in explaining Others' behavior in 4- to 9-year-old children from three cultural contexts. *J. Cross. Cult. Psychol.* 53, 659–682. doi: 10.1177/002202212211098423
- Kendon, A. (2004). *Gesture: Visible Action as Utterance*. 1st Edn. Cambridge: Cambridge University Press. doi: 10.1017/CBO9780511807572
- Kita, S. (2009). Cross-cultural variation of speech-accompanying gesture: a review. *Lang. Cognit. Process.* 24, 145–167. doi: 10.1080/01690960802586188
- Kita, S., Alibali, M. W., and Chu, M. (2017). How do gestures influence thinking and speaking? The gesture-for-conceptualization hypothesis. *Psychol. Rev.* 124, 245–266. doi: 10.1037/rev0000059
- Kita, S., and Özyürek, A. (2003). What does cross-linguistic variation in semantic coordination of speech and gesture reveal?: evidence for an interface representation of spatial thinking and speaking. *J. Mem. Lang.* 48, 16–32. doi: 10.1016/S0749-596X(02)00505-3
- Konijn, E. A., Smakman, M., and van den Berghe, R. (2020). "Use of robots in education," in *The International Encyclopedia of Media Psychology*, ed. J. Bulck (Hoboken, NJ: Wiley), 1–8. doi: 10.1002/9781119011071.iemp0318
- Kory Westlund, J. M., Dickens, L., Jeong, S., Harris, P. L., DeSteno, D., and Breazeal, C. L. (2017). Children use non-verbal cues to learn new words from robots as well as people. *Int. J. Child Comput. Interact.* 13, 1–9. doi: 10.1016/j.ijcci.2017.04.001
- Leite, I., McCoy, M., Lohani, M., Ullman, D., Salomons, N., Stokes, C., et al. (2017). Narratives with robots: the impact of interaction context and individual differences on story recall and emotional understanding. *Front. Robot. AI* 4, 29. doi: 10.3389/frobot.2017.00029
- Leite, I., Pereira, A., Mascarenhas, S., Martinho, C., Prada, R., and Paiva, A. (2013). The influence of empathy in human-robot relations. *Int. J. Hum. Comput. Stud.* 71, 250–260. doi: 10.1016/j.ijhcs.2012.09.005
- Lieven, E., and Stoll, S. (2013). Early communicative development in two cultures: a comparison of the communicative environments of children from two cultures. *Hum. Dev.* 56, 178–206. doi: 10.1159/000351073
- Lighthart, M. E. U., Neerincx, M. A., and Hindriks, K. V. (2020). Design patterns for an interactive storytelling robot to support children's engagement and agency. In *Proceedings of the 2020 ACM/IEEE International Conference on Human-Robot Interaction*, Cambridge United Kingdom: ACM, 409–418. doi: 10.1145/3319502.3374826
- Liszkowski, U., Brown, P., Callaghan, T., Takada, A., and De Vos, C. (2012). A prelinguistic gestural universal of human communication. *Cogn. Sci.* 36, 698–713. doi: 10.1111/j.1551-6709.2011.01228.x
- Marentette, P., Pettenati, P., Bello, A., and Volterra, V. (2016). Gesture and symbolic representation in Italian and English-speaking Canadian 2-year-olds. *Child Dev.* 87, 944–961. doi: 10.1111/cdev.12523
- Maure, R., and Bruno, B. (2025). Autonomy in socially assistive robotics: a systematic review. *Front. Robot. AI* 12, 1586473. doi: 10.3389/frobot.2025.1586473
- McNeill, D. (1992). *Hand and Mind: What Gestures Reveal About Thought*. Chicago: University of Chicago Press.
- Nicoladis, E., Nagpal, J., Marentette, P., and Hauer, B. (2018). Gesture frequency is linked to story-telling style: evidence from bilinguals. *Lang. Cogn.* 10, 641–664. doi: 10.1017/langcog.2018.25
- Oranç, C., and Küntay, A. C. (2020). Children's perception of social robots as a source of information across different domains of knowledge. *Cogn. Dev.* 54, 100875. doi: 10.1016/j.cogdev.2020.100875

- Peter, J., and van Straten, C. L. V. (2024). "Social robots and children: a field in development," in *The De Gruyter Handbook of Robots in Society and Culture*, eds. L. Fortunati, and A. Edwards (Berlin: De Gruyter), 371–388. doi: 10.1515/9783110792270-020
- Pettenati, P., Sekine, K., Congestri, E., and Volterra, V. (2012). A comparative study on representational gestures in Italian and Japanese children. *J. Nonverbal. Behav.* 36, 149–164. doi: 10.1007/s10919-011-0127-0
- Rohlfing, K. J., Altvater-Mackensen, N., Caruana, N., Van Den Berghe, R., Bruno, B., Tolksdorf, N. F., et al. (2022a). Social/dialogical roles of social robots in supporting children's learning of language and literacy—a review and analysis of innovative roles. *Front. Robot. AI* 9, 971749. doi: 10.3389/frobt.2022.971749
- Rohlfing, K. J., Fritsch, J., Wrede, B., and Jungmann, T. (2006). How can multimodal cues from child-directed interaction reduce learning complexity in robots? *Adv. Robot.* 20, 1183–1199. doi: 10.1163/156855306778522532
- Rohlfing, K. J., Leonardi, G., Nomikou, I., Rączaszek-Leonardi, J., and Hüllermeier, E. (2019). Multimodal turn-taking: motivations, methodological challenges, and novel approaches. *IEEE Trans. Cogn. Dev. Syst.* 12 (2), 260–271. doi: 10.1109/TCDS.2019.2892991
- Rohlfing, K. J., Vollmer, A.-L., Fritsch, J., and Wrede, B. (2022b). Which "motionese" parameters change with children's age? Disentangling attention-getting from action-structuring modifications. *Front. Commun.* 7, 922405. doi: 10.3389/fcomm.2022.922405
- Rohlfing, K. J., Wildt, E., and Tolksdorf, N. F. (2024). Language learning with media and technology in (early) childhood. *Dzieciństwo. Liter. Kult.* 6, 35–69. doi: 10.32798/dlk.1376
- Rohlfing, K. J., Wrede, B., Vollmer, A.-L., and Oudeyer, P.-Y. (2016). An alternative to mapping a word onto a concept in language acquisition: pragmatic frames. *Front. Psychol.* 7, 470. doi: 10.3389/fpsyg.2016.00470
- Rohrer, P. L., Delais-Roussarie, E., and Prieto, P. (2020). Beat gestures for comprehension and recall: differential effects of language learners and native listeners. *Front. Psychol.* 11, 575929. doi: 10.3389/fpsyg.2020.575929
- Salomo, D., and Liszkowski, U. (2013). Sociocultural settings influence the emergence of prelinguistic deictic gestures. *Child Dev.* 84, 1296–1307. doi: 10.1111/cdev.12026
- Samuelsson, C., Marklund, U., Lyxell, B., and Danielsson, H. (2025). The use of gestures in book Reading and play situations related to vocabulary in young children with cochlear implants. *Clin. Linguist. Phon.* 39, 559–575. doi: 10.1080/02699206.2024.2449216
- Sekine, K., and Furuyama, N. (2010). Developmental change of discourse cohesion in speech and gestures among Japanese elementary school children. *Riv. Psicolinguist. Appl.* 10, 97–116.
- Sekine, K., Wood, C., and Kita, S. (2018). Gestural depiction of motion events in narrative increases symbolic distance with age. *Lang. Interact. Acquisit.* 9, 40–68. doi: 10.1075/lia.15020.sek
- Serafini, F. (2017). "Visual literacy," in *Oxford Research Encyclopedia of Education*, ed. K. Hytten (Oxford: Oxford University Press). doi: 10.1093/acrefore/9780190264093.013.19
- Singh, A., Wrede, B., Richter, B., Groß, A., and Rohlfing, K. J. (2025). Manners matter: action history guides attention and repair choices during interaction. In *IEEE International Conference on Development and Learning (ICDL)*, 1–8. doi: 10.1109/ICDL63968.2025.11204385
- Skantze, G., Gustafson, J., and Beskow, J. (2019). "Multimodal conversational interaction with robots," in *The Handbook of Multimodal-Multisensor Interfaces: Language Processing, Software, Commercialization, and Emerging Directions - Volume 3*, eds. S. Oviatt, B. Schuller, P. R. Cohen, D. Sonntag, G. Potamianos, A. Krüger (New York, NY: Association for Computing Machinery), 77–104. doi: 10.1145/3233795.3233799
- So, W. C. (2010). Cross-cultural transfer in gesture frequency in Chinese–English bilinguals. *Lang. Cognit. Process.* 25, 1335–1353. doi: 10.1080/01690961003694268
- Sparaci, L., and Volterra, V. (2025). Deconstructing language development: a commentary on Karadöller, Sümer and Özyürek. *First. Lang.* 45, 01427237251326988. doi: 10.1177/01427237251326988
- Tanaka, F., and Kimura, T. (2010). Care-receiving robot as a tool of teachers in child education. *Interact. Stud.* 11, 263–268. doi: 10.1075/is.11.2.14tan
- Tanaka, F., and Matsuzoe, S. (2012). Children teach a care-receiving robot to promote their learning: field experiments in a classroom for vocabulary learning. *J. Hum. Robot Interact.* 1(1), 78–95. doi: 10.5898/JHRI.1.1.Tanaka
- Ter Bekke, M., Drijvers, L., and Holler, J. (2024a). Gestures speed up responses to questions. *Lang. Cogn. Neurosci.* 39, 423–430. doi: 10.1080/23273798.2024.2314021
- Ter Bekke, M., Drijvers, L., and Holler, J. (2024b). Hand gestures have predictive potential during conversation: an investigation of the timing of gestures in relation to speech. *Cogn. Sci.* 48, e13407. doi: 10.1111/cogs.13407
- Tolksdorf, N. F. (2024). *Wortlernen mit sozialen Robotern: Der Einfluss einer systematischen Variation des pragmatischen Rahmens auf das langfristige Lernen morphologisch komplexer Wörter von Vorschulkindern, 1. Auflage*. Tübingen: Narr Francke Attempto. <https://www.narr.de/wortlernen-mit-sozialen-robotern-1145-1/>
- Tolksdorf, N. F., Hönemann, D., Viertel, F. E., and Rohlfing, K. J. (2022). Who is that?! does changing the robot as a learning companion impact preschoolers' language learning? In *Proceedings of the 2022 ACM/IEEE International Conference on Human-Robot Interaction*, Sapporo Hokkaido, Japan, 1069–1074. doi: 10.5555/3523760.3523937
- Tolksdorf, N. F., and Schaffrath, S. (2026). "Converging perspectives on social robots in early childhood education: educators' and developers' visions for using social robots and an approach to interaction design for co-constructed language learning," in *The Digital Society*, eds. M. Anastasiadis, Y. Peters, and C. Thimm (Wiesbaden: Springer Fachmedien Wiesbaden), 267–288. doi: 10.1007/978-3-658-49118-5_12
- Tolksdorf, N. F., Wildt, E., and Rohlfing, K. J. (2024). Preschoolers' interactions with social robots: investigating the potential for eliciting metatalk and critical technological thinking. In *Companion of the 2024 ACM/IEEE International Conference on Human-Robot Interaction*, Boulder, CO USA: ACM, 1053–1057. doi: 10.1145/3610978.3640654
- Tykhonenko, V., Tolksdorf, N. F., and Rohlfing, K. J. (2026). The relation between children's shyness and their contingent dialogical actions when reacting to a social robot's instructions. *Philos. Trans. R Soc. Lond. B Biol. Sci.* 381, 20240375. doi: 10.1098/rstb.2024.0375
- Vogt, P., de Haas, M., de Jong, C., Baxter, P., and Krahmer, E. (2017). Child-robot interactions for second language tutoring to preschool children. *Front. Hum. Neurosci.* 11, 1–6. doi: 10.3389/fnhum.2017.00073
- Vogt, P., van den Berghe, R., de Haas, M., Hoffman, L., Kanero, J., Mamus, E., et al. (2019). Second language tutoring using social robots: a large-scale study. In *2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, Daegu, Korea (South): IEEE, 497–505. doi: 10.1109/HRI.2019.8673077
- Vollmer, A.-L., Wrede, B., Rohlfing, K. J., and Oudeyer, P.-Y. (2016). Pragmatic frames for teaching and learning in human–robot interaction: review and challenges. *Front. Neurobot.* 10, 1–20. doi: 10.3389/fnbot.2016.00010
- Vulchanova, M., Baggio, G., Cangelosi, A., and Smith, L. (2017). Editorial: language development in the digital age. *Front. Hum. Neurosci.* 11, 1–6. doi: 10.3389/fnhum.2017.00447
- Wagner, P., Malisz, Z., and Kopp, S. (2014). Gesture and speech in interaction: an overview. *Speech Commun.* 57, 209–232. doi: 10.1016/j.specom.2013.09.008
- Wakefield, E., Novack, M. A., Congdon, E. L., Franconeri, S., and Goldin-Meadow, S. (2018). Gesture helps learners learn, but not merely by guiding their visual attention. *Dev. Sci.* 21, e12664. doi: 10.1111/desc.12664
- Wood, D., Bruner, J. S., and Ross, G. (1976). The role of tutoring in problem solving. *Child Psychol. Psychiatry* 17, 89–100. doi: 10.1111/j.1469-7610.1976.tb00381.x
- Wray, C., and Norbury, C. F. (2018). Parents modify gesture according to task demands and child language needs. *First Lang.* 38, 419–439. doi: 10.1177/0142723718761729