Received: 2 September 2019 | Revised: 6 August 2020 | Accepted: 17 August 2020

DOI: 10.1111/jcal.12497

ARTICLE

Journal of Computer Assisted Learning WILEY

A toy or a friend? Children's anthropomorphic beliefs about robots and how these relate to second-language word learning

Rianne van den Berghe[1] | Mirjam de Haas[2] | Ora Oudgenoeg-Paz[1] | Emiel Krahmer[3] | Josje Verhagen[4] | Paul Vogt[2] | Bram Willemsen[3] | Jan de Wit[3] | Paul Leseman[1]

[1]Department of Special Education, Utrecht University, Utrecht, The Netherlands

[2]Department of Cognitive Science and Artificial Intelligence, Tilburg University, Tilburg, The Netherlands

[3]Department of Communication and Cognition, Tilburg University, Tilburg, The Netherlands

[4]Amsterdam Center for Language and Communication, University of Amsterdam, Amsterdam, The Netherlands

Correspondence

Mirjam de Haas, Department of Cognitive Science and Artificial Intelligence, Tilburg University, Tilburg, The Netherlands.
Email: mirjam.dehaas@uvt.nl

Funding information

Horizon 2020 Framework Programme, Grant/Award Number: 688014

Abstract

This study investigates the degree to which children anthropomorphize a robot tutor and whether this anthropomorphism relates to their vocabulary learning in a second-language (L2) tutoring intervention. With this aim, an anthropomorphism questionnaire was administered to 5-year-old children (N = 104) twice: prior to and following a seven-session L2 vocabulary training with a humanoid robot. On average, children tended to anthropomorphize the robot prior to and after the lessons to a similar degree, but many children changed their attributed anthropomorphic features. Boys anthropomorphized the robot less after the lessons than girls. Moreover, there was a weak but significant positive correlation between anthropomorphism as measured before the lessons and scores on a word-knowledge post-test administered the day after the last lesson. There was also a weak but significant positive correlation between the change in anthropomorphism over time and scores on a word-knowledge post-test administered approximately 2 weeks after the last lesson. Our results underscore the need to manage children's expectations in robot-assisted education. Also, future research could explore adaptations to individual children's expectations in child-robot interactions.

KEYWORDS

anthropomorphism, child-robot interaction, educational robots, robot tutoring, second-language learning

1 | INTRODUCTION

1.1 | Anthropomorphism

When interacting with a social robot, people have a tendency to attribute human forms, characteristics and/or behaviours to the robot. This phenomenon is called anthropomorphism (Bartneck, Kulić, Croft, & Zoghbi, 2009). People do not only anthropomorphize robots, but also many other non-human entities, such as animals, toys, and machines (Caporael, 1986), and presumably this helps them to understand and gain control over their environment (Duffy, 2003; Waytz et al., 2010). Anthropomorphism can be a useful mechanism in human-robot interaction (Duffy, 2003; Fink, 2012), because people evaluate robots more positively, collaborate better with them, and empathize more with robots that are more human-like or display more human-like behaviour than with robots that are less human-like

Rianne van den Berghe and Mirjam de Haas had equal contributions.

(Breazeal, Kidd, Thomaz, Hoffman, & Berlin, 2005; Eyssel, Kuchenbrandt, Hegel, & Ruiter, 2012; Hegel, Krach, Kircher, Wrede, & Sagerer, 2008; Moon et al., 2014; Riek, Rabinowitch, Chakrabarti, & Robinson, 2009). In this article, we set out to study the degree to which children anthropomorphize a humanoid robot, how children's anthropomorphic beliefs about the robot may change after multiple interactions with the robot, and whether children's anthropomorphistic perception of the robot and word knowledge after a second-language (L2) vocabulary training are related.

The degree to which people anthropomorphize a robot is affected by the robot's appearance and behaviours (DiSalvo, Gemperle, Forlizzi, & Kiesler, 2002; Phillips, Zhao, Ullman, & Malle, 2018; Tung, 2016). For example, people are more likely to anthropomorphize robots that have a torso, a skin, or appear to have gender (Phillips et al., 2018). Robot movement in general has also been found to increase human-likeness ratings (Tung, 2016). More specifically, using co-speech gestures has been found to increase anthropomorphism, and the use of social gaze to increase life-likeness (Salem, Eyssel, Rohlfing, Kopp, & Joublin, 2013; Zaga, de Vries, Li, Truong, & Evers, 2017).

However, people do not all anthropomorphize robots to the same degree. One of the reasons for these individual differences is that people use their own experiences in rationalizing the actions of an object and in reasoning about its mental states (Epley, Waytz, Akalis, & Cacioppo, 2008; Epley, Waytz, & Cacioppo, 2007; Lemaignan, Fink, & Dillenbourg, 2014), and may thus ascribe different mental states to objects depending on their own experiences. Thus, in human-robot interaction, the degree to which people anthropomorphize robots likely does not only depend on the type of robot used and the behaviour the robot displays, but also on the specific characteristics and experiences of the person interacting with the robot.

While most robot research on anthropomorphism has focused on adults (see Fink, 2012 for a review), children of all ages have been found to anthropomorphize robots as well (Beran, Ramirez-Serrano, Kuzyk, Fior, & Nugent, 2011; Kahn, Gary, & Shen, 2013; Lemaignan, Fink, Mondada, & Dillenbourg, 2015; Monaco, Mich, Ceol, & Potrich, 2018). Younger children (up to 12 years old) are more likely than older children to anthropomorphize robots (Beran et al., 2011; Kahn et al., 2012; van Straten, Peter, & Kühne, 2019). They experience more enjoyment and are less sensitive to the robot's style of interaction than older children (van Straten et al., 2019), which may relate to a higher degree of anthropomorphism. In particular, younger children are more likely to assign cognitive and affective beliefs to robots than older children, such as the ability to remember people and understand people's feelings (Beran et al., 2011). However, even preschool children attribute few biological properties to robots (Jipson & Gelman, 2007) and already understand that robots are something in between living beings and mechanical artifacts (Kory-Westlund & Breazeal, 2019). In a meta-analysis by Van Straten et al. (2019), a robot's responsiveness and role were the strongest predictor of children's closeness to a robot but the predictors for trust were not consistent. Also, this meta-analysis showed that boys feel more close to a robot with the same gender but girls are not affected by the gender of the robot.

1.2 | Changes in anthropomorphism

Previous research indicates that children's perceptions or expectations of robots can change over time. Children value a robot's properties differently depending on their experience with robots (Obaid, Barendregt, Alves-Oliveira, Paiva, & Fjeld, 2015; Sciutti, Rea, & Sandini, 2014). Before interacting with a robot, children attribute more importance to a robot's shape (e.g., having a head or arms) than its sensory and motor properties (e.g., the ability to feel or move). After having interacted with a robot, they value its sensory and motor properties more and its shape less than before (Sciutti et al., 2014). While this study did not specifically investigate anthropomorphism, it does suggest that sensory and motor properties, which can be linked to anthropomorphism, may become more important over time when children's experience with robots increases.

Bernstein and Crowley (2008) asked children between four and seven to evaluate different entities (including two robots) on livingness and intelligence. Children who had had little experience with robots, judged the robot more often as living than children who had had more experience with robots. Moreover, children who had had experience with robots were more likely to distinguish robots from other entities that they already knew (e.g., things that are living) and consider robots as intelligent, albeit in a unique ‘robot intelligence’ manner.

In contrast, a study by Kory-Westlund, Martinez, Archie, Das, and Breazeal (2016) did not find changes in anthropomorphism. A robot was framed either as a social agent or a machine by using either inclusive language and second-person pronouns or third-person pronouns and the word ‘robot’. In this study, children between ages three and seven played a sorting game with the robot. The degree to which they anthropomorphised the robot was assessed through a questionnaire administered both before and after the game. The study did not show an effect of framing on children's anthropomorphism, and there was no difference in the degree to which children anthropomorphized the robot before or after the game.

It is not clear from this study whether children's anthropomorphism is indeed unaffected by their interaction with the robot, or whether one interaction session was not enough to change their degree of anthropomorphism. On the one hand, people might attribute cognitive and social abilities to robots that robots cannot meet (Dautenhahn, 2004), which is particularly a problem for repeated interactions (Leite, Martinho, & Paiva, 2013). On this idea, the longer people would interact with robots, the more likely it should be that the robot falls short of these expectations, which would negatively affect people's tendency to anthropomorphize the robot. Evidence for this idea comes from a previous study with children in which explicitly informing children on the robot's lack of psychological abilities (e.g., self-consciousness, social cognition) led to lower anthropomorphism and trust (van Straten, Peter, Kühne, & Barco, 2020). It is also in line with a proposed model on the dynamics of anthropomorphism (Lemaignan, Fink, & Dillenbourg, 2014; Lemaignan, Fink, Dillenbourg, & Braboszcz, 2014). In this model, people are most likely to anthropomorphize a robot when first encountering it, because of their expectations about the robot and because the robot's behaviour may seem unpredictable and complex. Upon getting acquainted with a robot, people build a mental model to predict the robot's behaviour, and as the accuracy of this model increases, the robot is considered more machine-like than human-like, and anthropomorphic tendencies decrease.

On the other hand, studies have found that children attributed more anthropomorphic or more positive judgments after having repeated interactions with a robot (Leite, Pereira, & Lehman, 2017; Michaelis & Mutlu, 2018). Michaelis and Mutlu (2018) had 10- to 12-year-old children participate in in-home guided reading activities with a robot, and found that more children attributed feelings, emotions and a personality to the robot after the 2-week study than before. Though not measuring anthropomorphism directly, Leite et al. (2017) focused on likeability and found that four- to ten-year-old children liked the robot more after having multiple conversations with it. The current study is aimed at further investigating changes in children's evaluations of a robot in terms of anthropomorphism after multiple interactions with this robot, and relating these evaluations to their learning outcomes in a vocabulary training.

1.3 | Anthropomorphism and learning

Education is one of the most widely used domains in which social robots are used. Robots can be used to support children's learning, and as such, complement teachers. One of the most often used applications is the use of a robot as a tutor, such that the robot and child together work through educational materials and the robot provides individual support to the child (Belpaeme, Kennedy, Ramachandran, Scassellati, & Tanaka, 2018). A robot can interact with the children in their physical, referential world. The robot's embodiment and its potential for social interactions to establish common ground is one of the advantages social robots in theory have over other forms of technology such as tablets (Belpaeme, Kennedy, et al., 2018). Physical robots indeed have generally been found to be more enjoyable and a preferred social partner compared to their virtual counterparts (Kidd, 2003; Pereira, Martinho, Leite, & Paiva, 2008). It is assumed that such robots are more natural conversational partners, and robot-assisted learning interactions may benefit from similar social behaviours as humans use in learning interactions, such as the use of gestures (de Nooijer, van Gog, Paas, & Zwaan, 2013; de Wit et al., 2018; Kelly, McDevitt, & Esch, 2009; Macedonia, Müller, & Friederici, 2011; Tellier, 2008; Verhagen, van den Berghe, Oudgenoeg-Paz, Küntay, & Leseman, 2019). Furthermore, children have been shown to be less anxious and more motivated when learning with a robot than without a robot (Alemi, Meghdari, Basiri, & Taheri, 2015). Finally, an advantage of a robot is that it can endlessly repeat tasks with individual children where a teacher has to pay attention to other children.

These advantages of robots in education may particularly benefit robot-assisted language learning, which is studied in this article. Robots can gesture, move around and manipulate objects, and by doing so, embed the language that they are teaching in the physical environment that they share with the learner. For example, robots can point to the objects they are naming or act out the meaning of a word. This embedding is known to be important for language learning (Barsalou, 2008; Hockema & Smith, 2009; Iverson, 2010; Oudgenoeg-Paz, Leseman, & Volman, 2015; Wellsby & Pexman, 2014). As a result, (second) language learning has often been studied in robot-assisted learning research (see Van den Berghe, Verhagen, Oudgenoeg-Paz, van der Ven, & Leseman, 2019, Kanero et al., 2018 for reviews). So far, results on the effectiveness of robots for language learning are mixed, however. In this paper, we further explore one of the factors that may, at least in part, explain the mixed findings in earlier work, but has received relatively little attention to date: anthropomorphism.

As discussed earlier, anthropomorphizing robots seems advantageous for human–robot interactions (Duffy, 2003; Fink, 2012), but it is not clear if and how anthropomorphism can affect robot-assisted (language) learning. Yet, the degree to which learners anthropomorphize robots may play an important role in learning situations too, as learning is first and foremost a social process (Vygotsky, 1978). Children who anthropomorphize the robot to a greater degree might interact with the robot in ways similar to how they would interact with peers. Peer learning has been shown to be beneficial to learning (see Topping, 2005 for a review), either directly through helping each other, or indirectly through enhancing motivation, confidence and enjoyment.

Anthropomorphism is related to social presence: 'the degree to which a user feels access to the intelligence, intentions, and sensory impressions of another' (Biocca, 1997, section 7.2). It reflects paying attention to each other, understanding each other, and adapting behaviour and emotions towards each other. It is no surprise that such values are also crucial to successful vocabulary training programmes (Marulis & Neuman, 2010), and may thus apply to the robot-assisted vocabulary training in the current study. It may be worthwhile to design robots in such a way that they make learners feel as if it has a social presence, but the learner's perception of the robot and its social presence may be just as important. It is possible that a robot's benefits as a peer learner or tutor depend on the degree to which the learners anthropomorphize it. In other words, it is possible that a robot perceived as more human-like is more effective when learning an L2 than a robot that is perceived as a machine. This begs the question if and how anthropomorphism and learning are related to each other, which is the central research question of the current study.

Research that comes closest to answering this question is that of Chandra et al. (2018). This study did not directly focus on anthropomorphism, but the researchers did measure children's perception of a robot in terms of intelligence, likeability and friendliness, and whether this affected their learning in a learning-by-teaching paradigm. In this study, 25 seven- to nine-year-old children taught a NAO robot to write over the course of four sessions as a way to improve their own writing. There were two conditions: (a) the robot improved its handwriting for half of the children, and (b) the robot did not improve its writing for the other half of the children. Children in the first condition were able to perceive the robot's improvement by the last session, but this as such did not change how they perceived the robot's intelligence, likeability, and friendliness. However, children's own improvement in writing was positively correlated with the likeability of the robot. In the condition in which the robot did not improve, children's perceptions of the robot's intelligence, likeability, and friendliness did not change either, but in this condition, children's own learning was correlated with the perceived friendliness of the robot. These findings need to be interpreted with caution because of the small sample size and because they did not measure anthropomorphism, but they suggest that children's perception of the robot may indeed be related to their learning.

Our study expands on this previous work. It includes an L2 vocabulary training of multiple sessions, thus enabling us to study children's anthropomorphism of a robot and changes therein over a longer period of time. This increases ecological validity, as robot-based interventions aimed at teaching children a particular topic usually span a few weeks, causing novelty effects of the robot that wear off after multiple interactions (e.g., Kanda, Hirano, Eaton, & Ishiguro, 2004). We assess the degree to which children anthropomorphize the robot both before and after having interacted intensively with it, allowing to observe changes in anthropomorphism, and examine how children's anthropomorphism and changes therein relate to language-learning performance.

1.4 | This study

The current study was part of the L2TOR project, which evaluated the effectiveness of a multiple-session L2 learning intervention for young children using a social robot in a large-scale randomized control trial (Vogt et al., 2019). This long-term control study was preregistered on AsPredicted[1] and included four conditions: (a) an L2 vocabulary training with a tablet and a robot that performed iconic and deictic gestures to support word learning (gestures that visualize target words and pointing gestures), (b) an L2 vocabulary training with a tablet and a robot without iconic gestures (only pointing gestures), (c) an L2 vocabulary training with a tablet only (no robot involved) and (d) a control condition in which children only played dancing games with the robot.

Word knowledge was tested on three occasions, during a pretest, an immediate post-test and a delayed post-test (administered between 2 and 4 weeks after the training). The results of this preregistered study regarding children's word knowledge are reported in Vogt et al. (2019) and showed that, irrespective of condition, children knew significantly more words after the tutoring sessions than before. Moreover, children in the experimental conditions (robot with iconic gestures, robot without iconic gestures, and tablet-only) scored significantly higher than children in the control condition on word-knowledge tests during the immediate and delayed post-tests. There were no differences between the experimental conditions, such that children who had taken the tutoring sessions with the robot (with or without iconic gestures) did not know more words than children who had taken the sessions with the tablet only.

In the current study, we only included the experimental robot conditions (i.e., Conditions 1 and 2) to investigate the degree to which children anthropomorphized the robot and the way in which this relates to their word knowledge. In our analyses, we did not include the tablet-only and control conditions because, children in these conditions either did not interact with the robot (tablet condition) or were not taught any English words by the robot (control condition). We addressed the following research questions and hypotheses:

1. Are there individual differences in the degree to which children anthropomorphize the robot? We expect children to differ in the degree to which they anthropomorphize the robot, in line with previous research on individual differences in anthropomorphism (Epley et al., 2007, 2008).
2. How does the degree to which children anthropomorphize the robot change through multiple L2 tutoring sessions with the robot? Although the evidence is mixed (e.g., Bernstein & Crowley, 2008; Kory-Westlund et al., 2016; Michaelis & Mutlu, 2018), we expect that anthropomorphism will change over time in different ways, due to the multiple interactions children have with the robot. On the one hand, children may come to perceive the robot more as a friend after repeated interactions, thus perceive the robot as more human-like. On the other hand, it is also possible that children initially have high expectations of the robot's interactive qualities, which the robot, however, cannot meet. In that case, their perception will change over time towards considering the robot as less human-like.
3. How are children's anthropomorphistic perceptions of the robot and their knowledge of L2 words related? We expect word knowledge and attributing human-like cognitive, emotional and biological qualities to the robot to be positively related to each other. Specifically, we anticipate that children who anthropomorphize the robot more will treat the robot as a peer that has social presence, and, as such, benefit more from its presence in terms of increased motivation and engagement, that, in turn, will foster word learning. It should be noted that while this design does not enable us to study causal relations between anthropomorphism and word knowledge, we study whether the two are related and therefore provide evidence pertinent to the possible role anthropomorphism can play in the effectiveness of robot-based educational interventions.

2 | METHOD

2.1 | Participants

This study reports on a part of the sample described in Vogt et al. (2019), that is, the children in the two experimental robot conditions. Data were used from 104 monolingual Dutch children (50 girls) with an average age of 5 years and 8 months (*SD* = 5 months) who followed the vocabulary training in one of the two robot-assisted conditions (with or without iconic gestures). These children were recruited from the kindergartens of nine primary schools in the Netherlands. Within schools, children were randomly assigned to one of the conditions, while ensuring a similar gender distribution over the conditions. There were 53 children (23 girls) in the iconic-gesture condition (*M* age = 5 years and 8 months, *SD* = 4.8 months) and 51 children (26 girls) in the no-iconic-gesture condition (*M* age = 5 years and 8 months, *SD* = 4.6 months). Sixteen additional children were excluded because they: (a) knew more than half of the target words in the pre-test (*n* = 3), (b) did not complete the experiment due to technical issues (*n* = 2), (c) did not want to participate anymore (*n* = 8), or (d) did not complete the anthropomorphism questionnaire during the pre-test (*n* = 3). All children's parents signed an informed-consent form to allow their children to participate in this study. Children received a small gift at the end of the study to thank them for participation. The project in which the study was embedded, the L2TOR project, received ethical approval from Utrecht University's Ethics Committee under protocol number FETC16-039.

FIGURE 1 A child playing with the robot [Colour figure can be viewed at wileyonlinelibrary.com]

FIGURE 2 Example of one of the virtual environments that was used as a context for the language-learning interaction [Colour figure can be viewed at wileyonlinelibrary.com]

2.2 | L2 tutoring sessions

The aim of the L2 tutoring sessions was to teach each child 34 English words in the domains of mathematical and spatial language. Each child received seven tutoring sessions involving the robot and a tablet. During each of the sessions children were introduced to five or six new target words. The Softbank Robotics NAO robot was used, which was sitting in a 90° angle next to the child (see Figure 1).

A three-dimensional game was developed for the tablet, in which a particular scenario was displayed (e.g., animals in the zoo that had escaped their cages). This served as the context in which the L2 words were introduced (see Figure 2). For each word, the child and the robot had to perform different tasks on the tablet (e.g., selecting or dragging objects on the screen, repeating target words out loud or acting out target words). For instance, the robot would ask the child to drag three escaped animals back into their cage on the tablet. While dragging, the robot would count in English the number of animals in the cage.

During these tasks, the robot acted as a slightly more knowledgeable peer who was also being taught English words, but could provide feedback on the child's actions when needed. For example, when a child dragged the wrong animal to a cage on the tablet, the robot could ask the child to drag the correct animal to the cage. See Table 1 for an example of the child and robot interaction.

The lessons were designed without relying on children's speech because speech recognition is currently still unreliable with children (Kennedy et al., 2017). For the few times children had to repeat a word, a Wizard of Oz was used where the researcher pressed a button on a control panel after the child had repeated after the robot. The rest of the interaction was carried out autonomously. The interaction was one-on-one in a separate room, but the experimenter stayed in the same room to intervene when necessary and to control the Wizard.

2.3 | Robot behaviour

During the sessions, the robot was in breathing-mode (moving with its arms) to appear more lively. As the robot motors can be quite loud when the robot moves, the breathing-mode also reduced the initial sound shock when the robot was going to make a gesture and moved up its arms. In both conditions, the robot used deictic gestures, such as pointing, to draw the child's attention to the tablet, and head movements to look at the child when the child was asked to perform a certain task on the tablet. The only addition to the iconic + deictic gesture condition on top of the deictic gesture condition was the robot's use of iconic gestures. Specifically, an iconic gesture was designed for each of the included target words, and the robot would perform this gesture whenever it produced that word in the L2.

Gestures were designed using key framing (Pot, Monceaux, Gelin, & Maisonnier, 2009), an animation technique where the designer defines a number of key positions of a character's limbs, and smooth transitions between these points are automatically generated.

TABLE 1 An example of an interaction between robot and child

Robot	Child's action
Tablet shows an environment with three cages, and three giraffes outside of the cages	
Let us put the <giraffe> in its cage!	Drags giraffe into cage
Well done!	
There are still <two giraffes> outside of the cage. There are <more giraffes> outside of the cage than inside the cage. Can you <add one> giraffe?	Drags giraffe into cage
Well done!	
We had to <add one giraffe> and now there are <two giraffes> in the cage. There are <more giraffes> inside the cage than outside the cage. Can you add <one giraffe>?	Drags giraffe into cage
Well done!	
Please touch the cage with the three animals, so we can hear what three is in English	Touches cage with the three giraffes
Tablet says < three>	
Repeat after me: <three>	
	Says three
Well done! < .... >	
Tablet display adds three trees to the tablet environment	
Cool! The last thing we need to do is to put food in the cage with the giraffes. This cage has the <most> animals so they need the <most> trees. Put the trees in the cage so the giraffes can eat from them. We have <three> giraffes, so we need <three> trees. Put the trees in the cage. Count them while dragging	Drags first tree
Let's do one more	Drags second tree
And the last one	Drags third tree
Well done!	
Great! Now each giraffe has their own tree because there are <three> trees and <three> giraffes. The cage is pretty full because <most> animals are in the giraffe cage with the <most> food. You did very well! Let's do something else!	

Note: The whole interaction was in Dutch, except for the words between brackets<>.

The design was based on human-performed gestures, which were recorded by means of a gesture-elicitation procedure where participants were asked to come up with a gesture depicting each of the target words. The resulting robot gestures were recreated based on the recorded examples, while taking into account the robot's physical limitations (such as its inability to move individual fingers) and the fact that the robot would be sitting down rather than standing, as the human performers were. Figure 3 shows examples of the robot gestures for the target words *running* and *behind*.[2]

2.4 | Materials and measurements

2.4.1 | Anthropomorphism questionnaire

This anthropomorphism questionnaire was constructed for the purposes of the present study and administered by an experimenter in a one-on-one session with the child. The questionnaire took about 10 min to complete. It consisted of 12 questions (for an overview, see Table 2 in Section 3) and assessed the degree to which children anthropomorphized the robot with regard to various types of properties: biological (e.g., feeling pain, need for food and ability to grow), cognitive (e.g., thinking, remembering), and emotional (e.g., being happy, being sad). Each question could be answered with 'yes'/'no'/'I don't know' and was followed by an open-ended query asking children why they gave this response. The items were based on Jipson and Gelman (2007), who investigated to what extent children make a distinction between living and non-living items. The questionnaire was adapted to fit the present study by adding several items to more thoroughly assess anthropomorphism (e.g., rather than measuring the robot's emotional abilities by only asking whether the robot could feel happy, an item was added on whether the robot could feel sad). The children's closed-ended answers were compared with the open-ended answers to find out whether the children understood the question. Two of the included questions (i.e., 'Can the robot break?' and 'Is the robot made by humans?') proved unreliable as children's answers to the open-ended query did not correspond to their answers on the close-ended questions. Therefore, we removed these items from our analysis. The children were awarded one point for each 'yes'-answer, which indicated that they attributed human-like properties to the robot, and their anthropomorphism score was the proportion of 'yes'-answers. We used proportions rather than total scores because there were missing values on some items for some children. This was the case for one child at the pre-test (four of the twelve questions were not administered) and for five children at the post-test (for each of whom one question was not administered). Thus, the maximum score was 1, with a score closer to 1 denoting a child's tendency to consider the robot as human-like. Cronbach's alpha indicated that the internal consistency of the questionnaire was satisfactory, α = .72 at the pre-test and α = .75 at the post-test.

2.4.2 | Comprehension test

The comprehension test was a picture-selection task. In this test, children were presented with a prerecorded target word and asked to choose which one out of three pictures or short video clips matched this word best ('Where do you see: [heavy]?'). Each target word was presented three times with different target and distractor stimuli in random order to decrease the chance of children guessing the correct answer. Only half of the 34 target words that were presented in the vocabulary training were included, as a test including all target words would have taken too long for these young children. The same test was used in both post-tests. The internal consistency of the

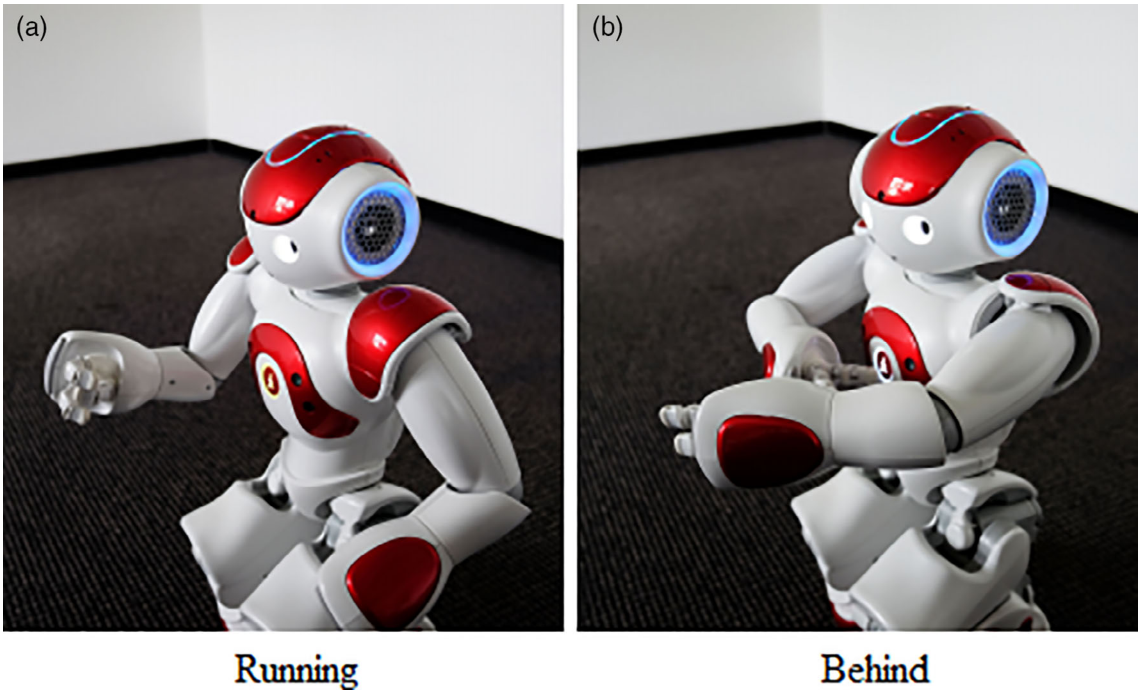

FIGURE 3 Examples of iconic gestures used in this study, photographed from a position where the child would sit. (a) *Running* is gestured by moving both arms back and forth as if the robot is running. (b) The word *behind* is gestured by moving the right hand up and down behind the left hand. Figures taken from Vogt et al. 2019 [Colour figure can be viewed at wileyonlinelibrary.com]

comprehension test was good, with Cronbach's alpha α = .84 at the first post-test and α = .87 at the second post-test.

2.4.3 | Additional measures

In addition to the anthropomorphism questionnaire and comprehension task, we administered several tasks assessing general cognitive skills. These tasks are beyond the scope of this study as they did not assess anthropomorphism (see Vogt et al., 2019). They were: (a) a Dutch receptive vocabulary test (Peabody Picture Vocabulary Test; Dunn & Dunn, 2005), (b) a selective-attention task (visual-search task; Mulder, Hoofs, Verhagen, van der Veen, & Leseman, 2014) and (c) a phonological-memory test (quasi-universal nonword-repetition test, Boerma et al., 2015).

Moreover, we administered two translation tests to measure children's knowledge of the English words, in which children listened to the target words in L2 and were asked for their Dutch translations, or vice versa. The English-to-Dutch translation test was used as a pre-test. Note that the main purpose of this translation test during the pre-test was to enable us to exclude any children who knew many words prior to the lesson series, although it also allowed us to compare pre- and post-test scores (see Vogt et al. 2019 for these analyses). We chose not to include a comprehension test as a pre-test, as children may learn from such tests, given that, unlike in the translation task where no answer is provided, a word is presented with pictures, one of which depicts the word's meaning. Moreover, in this article, we did not include the translation tests in the analyses, as there was low variability in children's scores. Thus, in the current paper, we only include the comprehension test as a measure of children's word knowledge.

2.5 | Procedure

Prior to the experiment, all children participated in a group introduction with the robot to familiarize the children with the robot, build trust, and explain the basic similarities and dissimilarities between the robot and humans (e.g., the robot speaks without moving its mouth, but looks at us while speaking in the same way as humans do; Belpaeme et al., 2018). These explanations were deemed necessary to make sure that children would know how to interact with the robot in the subsequent lessons. During the introduction, participants danced together with the robot, were allowed to shake the robot's hand, and played a brief gesture imitation game. The robot was not explicitly framed as either a human or a machine, by avoiding pronouns and by being called ‘Robin the robot’ (i.e., a combination of a gender-neutral human name and the label ‘robot’).

After the introduction, a pre-test was administered including the anthropomorphism questionnaire and several tests measuring general cognitive skills as well as children's knowledge of the English words. In the weeks thereafter, the children received seven one-on-one tutoring lessons with the robot. Each lesson took approximately 17 min to complete. One or 2 days after the last lesson, an immediate post-test was administered including the anthropomorphism questionnaire for the second time, the comprehension test, and other tasks measuring children's knowledge of the English words. Finally, a delayed post-test was administered in which the comprehension test and other English vocabulary tasks were repeated, between 2 and 4 weeks after the tutoring lessons.

2.6 | Analyses

In Section 3, each research question is addressed in a separate paragraph. First, we examined whether there were individual differences in the degree to which children anthropomorphized the robot before the tutoring sessions (RQ1). We used independent-samples *t*-tests to explore effects of gender and condition, and a linear-regression analysis for age. We used age as a continuous variable in our analyses, but reported means for a 'younger' and an 'older' age group in Table 3 in Section 3, calculated through a median split (at 68.2 months).

Second, we investigated how the degree to which children anthropomorphized the robot changed through multiple L2 tutoring sessions with the robot (RQ2). We used a paired-samples *t*-test to compare anthropomorphism scores before and after the tutoring sessions. We also explored effects of gender, condition, and age, using a mixed-design ANOVA with gender or condition as a between-subject variable and time as a within-subject variable, and a linear-regression analysis for age and change in anthropomorphism scores.

Third, we used Pearson's correlations to investigate how anthropomorphism and knowledge of L2 words were related (RQ3). We correlated children's scores on the anthropomorphism questionnaire before and after the tutoring sessions, change in scores on the anthropomorphism questionnaire, and scores on the comprehension test on each post-test (i.e., immediate and delayed).

TABLE 2 Proportions of children answering 'yes' on the questionnaire before and after the tutoring sessions

Do you think that Robin the robot…	Before	After
… can see things?	0.79 (82)	0.81 (84)
… can be sad?	0.66 (69)	0.41 (43)
… remember something?	0.64 (67)	0.69 (72)
… can feel it when you tickle Robin the robot?	0.45 (47)	0.33 (34)
… can think?	0.78 (81)	0.65 (68)
… has to eat?	0.27 (28)	0.17 (18)
… understands when you say something?	0.66 (69)	0.74 (77)
… can feel pain?	0.46 (48)	0.29 (30)
… can enjoy something?	0.92 (96)	0.92 (96)
… grows?	0.15 (16)	0.12 (12)
… can be happy?	0.94 (98)	0.87 (90)
… can recognize you?	0.49 (51)	0.89 (92)
Overall scores	0.60 (60)	0.57 (60)

Note: The total number of children can be found between brackets.

3 | RESULTS

3.1 | Anthropomorphism before tutoring sessions

We investigated our first research question: Are there individual differences in the degree that children anthropomorphize the robot? Table 2 displays the questions of the questionnaire and the proportions of children that answered the question with 'yes'.

As a group, children tended to attribute more human-like properties to the robot than machine-like properties as is reflected in the overall proportions being higher than 0.50 both before and after the tutoring sessions, but the scores varied strongly between the questions. Children highly agreed that the robot 'can enjoy something', 'can be happy', and 'can think'. They disagreed more often on various biological properties, such as 'Do you think Robin the robot feels it when you tickle Robin the robot?' and 'Do you think that Robin the robot can feel pain?'

We explored whether there were effects of gender, age, and condition. The mean anthropomorphism scores, separated for gender, age and condition, are displayed in Table 3. An independent-samples *t*-test showed no effect of gender, $t(102) = -0.30$, $p = .768$, $d = 0.06$, and a linear-regression analysis showed no effect of age, $F(1,102) = 2.24$, $p = 0.138$. With respect to condition, we explored whether children perceived the robot differently in the iconic-gesture condition compared to the condition without iconic gestures as measured before the robot interaction, using an independent-samples *t*-test. There were no differences between the two conditions in the degree to which children anthropomorphized the robot, $t(102) = -0.36$, $p = .722$, $d = 0.07$.

3.2 | Change in anthropomorphism after tutoring sessions

Then, we investigated our second research question: How does the degree to which children's anthropomorphize the robot change through multiple L2 tutoring sessions with the robot? There was a

TABLE 3 Children's mean anthropomorphism scores (*SD*) before and after the tutoring sessions, separated for gender, age and condition

	Before	After
Gender		
Male	0.60 (0.20)	0.53 (0.22)
Female	0.61 (0.19)	0.62 (0.17)
Age		
Younger	0.62 (0.19)	0.56 (0.20)
Older	0.59 (0.20)	0.59 (0.20)
Condition		
No iconic gestures	0.60 (0.19)	0.59 (0.20)
Iconic gestures	0.61 (0.20)	0.57 (0.20)

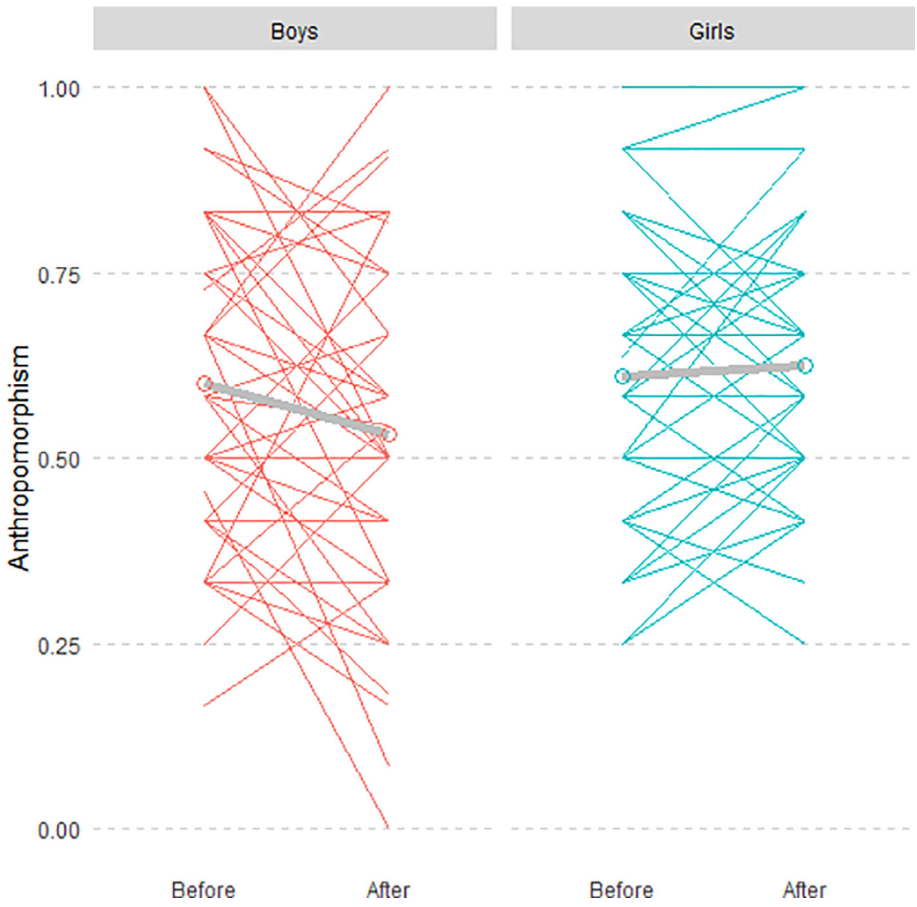

FIGURE 4 Anthropomorphism scores as a function of gender before and after the tutoring sessions [Colour figure can be viewed at wileyonlinelibrary.com]

positive and moderately strong correlation between scores before the tutoring sessions and after the tutoring sessions on the anthropomorphism questionnaire, *r*(104) = 0.505, *p* < .001, indicating moderate overall stability of anthropomorphism. However, there was also large variability among the children in whether and how the degree to which they anthropomorphized the robot changed before and after the tutoring sessions. Most children were consistent in the degree to which they anthropomorphized the robot (45 children), that is, their anthropomorphism scores during the two test moments were the same or differed by a maximum of one question. However, a relatively large number of children anthropomorphized the robot less after having interacted with it in the tutoring sessions (35 children). An increase in anthropomorphism also occurred, but was least common (24 children).

We compared children's answers on the anthropomorphism questionnaire after the tutoring sessions to their answers before the tutoring sessions. Table 2 shows that children changed their opinion drastically on a number of questions. Fewer children believed after the tutoring sessions that the robot could feel it when being tickled, that it could feel pain, or that it could be sad. More children believed after the tutoring sessions that the robot could understand what they said, and that the robot could recognize them. However, a paired samples *t*-test did not show significant differences between children's overall scores before and after the tutoring sessions on the anthropomorphism questionnaire, *t*(103) = 1.53, *p* = .130, *d* = 0.15.

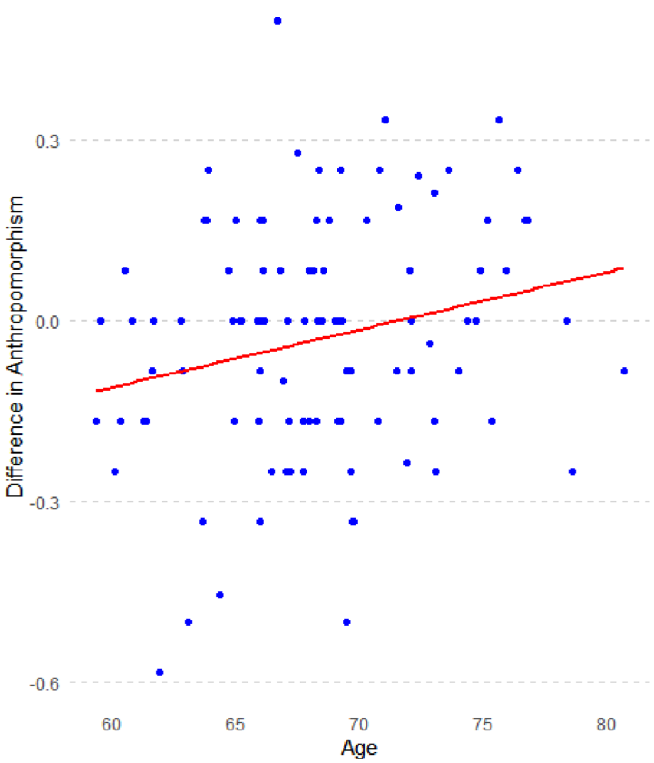

FIGURE 5 Age and the difference in anthropomorphism scores [Colour figure can be viewed at wileyonlinelibrary.com]

TABLE 4 Children's mean scores (*SD*) on the comprehension test

Condition	Immediate post-test	Delayed post-test
Iconic gestures	29.47 (5.85)	30.43 (6.22)
No iconic gestures	29.39 (6.08)	29.75 (6.44)

Note: The maximum score on the comprehension test was 54.

TABLE 5 Correlations between the anthropomorphism scores and the L2 comprehension scores

	Comprehension	
Anthropomorphism	Immediate post-test	Delayed post-test
Before tutoring sessions	−0.208*	−0.137
After tutoring sessions	−0.167	0.074
Change	0.036	0.212*

Note: *p < .05.

We then explored whether there were effects of gender, age, and condition (see Table 3 for the means). A mixed-design ANOVA with gender as a between-subject variable and test moment (before and after the tutoring sessions) as a within-subject variable showed an interaction between gender and test moment, $F(1,102) = 4.35$, $p = .039$, $\eta_p^2 = 0.04$. Boys assigned more human-like qualities to the robot before interacting with the robot than after the tutoring sessions, $t(54) = 2.28$, $p = .027$, $d = 0.32$, while there was no difference in girls' anthropomorphism scores between the two test moments, $t(48) = -0.54$, $p = .592$, $d = 0.07$. The interaction is displayed in Figure 4. Moreover, an interaction effect with age was found. A linear-regression analysis was used to predict the difference score in anthropomorphism from age. Age significantly predicted the change in anthropomorphism over time, $F(1,102) = 5.56$, $p = .020$, with an R^2 of .05. Children's predicted changed anthropomorphism score is equal to −0.68 + 0.01 *(age in months). Figure 5 shows that a younger age was associated with a larger decrease in anthropomorphism over the tutoring sessions. Participants' change in anthropomorphism increased 0.01 for each month of age.

To explore whether children perceived the robot differently in the iconic-gesture condition compared to the condition without iconic gestures over time, a mixed-design ANOVA with condition as a between-subject variable and test moment as a within-subject variable revealed that condition did not interact with time, $F(1,102) = 0.64$, $p = .434$, $\eta_p^2 = 0.01$. Thus, the use of iconic gestures apparently was not associated with a different change of children's anthropomorphizing of the robot.

3.3 | Anthropomorphism and word knowledge

Last, we investigated our third research question: How are children's anthropomorphistic perceptions of the robot and their knowledge of L2 words related? As already mentioned, we only included the children's comprehension-test scores to look at the relation with anthropomorphism. Table 4 displays children's scores on the comprehension test during both post-tests.

Pearson's correlations showed that anthropomorphism before the tutoring sessions was weakly related to the comprehension scores on the immediate post-test, $r(104) = -.208$, $p = .034$ (see Table 5). The relation was negative, suggesting that children who anthropomorphized the robot more prior to the lesson series knew fewer words during the immediate post-test than children who anthropomorphized the robot less. Anthropomorphism after the tutoring sessions was not related to comprehension scores on either post-test, both *ps* > .090.

Children's change in anthropomorphism was weakly but significantly and positively related to the comprehension scores on the delayed post-test, $r(104) = .212$, $p = .031$. Thus, the larger the change towards anthropomorphism of the robot over time, the higher the performance on the delayed post-test and vice versa.

4 | DISCUSSION

In the present study, we investigated (a) the degree to which 5-year-old children anthropomorphized a social robot, (b) whether the degree of their anthropomorphism changed after intensive experience with the robot acting as a peer tutor in an L2 word learning intervention and (c) whether anthropomorphism and the change therein were related to children's word knowledge.

4.1 | Anthropomorphism before tutoring sessions

We investigated the way children perceived the robot after a group-wise introduction session and prior to the tutoring sessions. Overall, children slightly more often agreed than disagreed with statements attributing human-like properties to the robot, but there were large differences between children in the degree to which they anthropomorphized the robot, in line with research on individual differences in the tendency to anthropomorphize objects (Epley et al., 2007, 2008). Moreover, children agreed more often with statements that attributed cognitive and, to some extent, also positive emotional states to the robot than biological properties and negative emotional states, in line with previous work that also found that young children are likely to ascribe cognitive mental states to robots (Beran et al., 2011).

As this was not the scope of the current study, we did not present and analyse children's answers to the open-ended questions, which asked them to motivate why they perceived the robot as more or less human-like. However, we noticed that there were large differences between the children, similar to their overall anthropomorphism scores, in the way they explained why they perceived the robot in the way they did. For example, some children thought that the robot would be sad if children did not want to play with it, while other children thought the robot would be sad if it was in pain. Some children thought that the robot could not be sad because it had no feelings, while other children thought the robot could not be sad because it could not handle water and, thus, could not cry.

Contrary to our expectation that gestures would make the robot more human-like, children did not anthropomorphize the robot more when it used iconic gestures than when it only used deictic gestures. This might be due to our design of the experiment, as the robot used the same repetitive gesture each time it used a target word. This repetitive behaviour could have reduced the positive effect of the gestures in respect to human-likeness of the robot. As humans do not use the same gesture each time they use a target word, variation in gesture use might increase the human-likeness of the robot again, which would be interesting to explore further. It is possible that in this study the iconic gestures did not convey the concepts as clearly as they were intended and as a consequence of that, the gestures did not impact anthropomorphism. The robot used iconic gestures for each target word and some words were more difficult to act out, such as 'more', for which an iconic gesture is not that iconic. This is supported by the lack of differences in learning outcomes (Vogt et al., 2019), in contrast to a different study (de Wit et al., 2018), where the iconic gestures clearly portrayed the meaning of the word and in which differences in word knowledge were found.

4.2 | Change in anthropomorphism after tutoring sessions

We investigated whether the degree to which children anthropomorphized the robot had changed after the L2 tutoring sessions. There were no significant differences in overall anthropomorphism, and, similar to the pre-test, children on average slightly more agreed than disagreed with attributing human-like properties to the robot at the post-test. However, with regard to specific properties some major changes were observed. Fewer children answered 'yes' to questions attributing biological properties and negative emotions to the robot at the post-test as compared to the pre-test. This concerned, for example, questions asking whether the robot 'could feel it when being tickled' or 'could feel pain'. This is in line with the study of Sciutti et al. (2014), who found that the robot's sensory and motor properties became more salient to children after they had interacted with a robot. At the post-test, more children answered 'yes' to questions addressing cognitive abilities, such as whether the robot can remember something, understand them when they say something, and is able to recognize them.

These changes together indicate an interesting shift in the way in which the robot is seen by children after intensive experience, namely as a basically mechanical being but with positive mental states, whereas initially children showed more confusion regarding the biological aspects and were less strongly convinced of the cognitive capabilities of the robot. We believe that this shift is due to the way in which the lessons were designed. At the start of each lesson, the robot greeted the children personally while mentioning their names, referred to the previous lessons and tracked the children's faces to suggest that the robot looked at them. The open-ended answers confirmed that possibility as their explanations for 'Do you think Robin the robot can recognize you?' changed from 'Robin the robot has not met me yet' to 'Robin the robot said my name every time we played'. It is likely that children were less inclined to believe that the robot could recognize them at the pre-test, simply because they had not yet played intensively with the robot in a one-on-one setting yet at that time. The same shift was found in the explanations of children for the question whether the robot can remember something: children started with many different explanations before the interaction like 'No, Robin the robot has small ears so cannot remember much', 'Yes, Robin the robot looks like a human so can also remember things' and changed their explanation after the interaction to 'Yes because Robin the robot remembered where we played before'.

Regarding negative emotional states, fewer children believed at the post-test that the robot 'could be sad', which can also be explained by the design of the lessons. Even though the robot expressed happiness (by changing the colours of its eyes) and also when it was *not* specifically happy (by not changing the colours of its eyes), it never expressed negative emotions, such as sadness or anger. Again, this was supported by the children's open answers where they mentioned the robot's coloured eyes during the post-test questionnaire.

Most children anthropomorphized the robot either to the same degree or to a lesser degree during the post-test as compared to the pre-test. Fewer children increased their anthropomorphism of the robot. Explorative analyses showed that age and gender had an influence on the change in anthropomorphism: Boys and younger children had a larger decrease in anthropomorphism than girls and older children.

It is possible that decreases in anthropomorphism were due to children initially having high expectations of the robot's interactive (human-like) qualities, which the robot could not meet (Dautenhahn & Werry, 2004). This effect could have affected younger children more as older children seem to anthropomorphize robots less in general (van Straten et al., 2019). Moreover, gender influenced the change in anthropomorphism. It is possible that girls were more forgiving of the robot's flaws than boys were and that girls therefore did not change their perception as much as boys (Tung, 2011).

The robot was largely autonomous during the tutoring sessions, but did not engage in personalized conversations with the children. The robot kept to the preprogrammed script and did not answer children's questions. For children with high expectations regarding the human-likeness of the robot, this could have led them to decrease their attribution of human-like properties to the robot. Conversely, children who had a less human-like perception of the robot prior to the tutoring sessions may have had low expectations of the robot's interactive (human-like) qualities. Since the robot displayed at least some human-like behaviours, such as mentioning the child by name (suggesting that it recognized the child) or indicating that it liked the sessions, this could have increased children's beliefs about the robot as human-like over repeated interactions. Thus, the observed changes in anthropomorphism may not only have been dependent on the robot's behaviours (in line with Tung, 2016), but also on whether this behaviour matched children's prior expectations.

A final possibility is that the observed change in anthropomorphism merely reflects the phenomenon of regression to the mean, with initially higher scores decreasing and initially lower scores increasing at post-test due to random measurement error. While we cannot fully rule out this explanation, it should be noted that more children decreased rather than increased in anthropomorphism, and the analysis at the item level revealed a complex but interpretable pattern of changes that pointed to a shift in how children perceived the robot within a similar overall anthropomorphism score at the pre- and post-test.

4.3 | Anthropomorphism and word knowledge

Finally, we investigated whether anthropomorphism and word knowledge were related. We found two weak but significant correlations. Children's anthropomorphism of the robot at pre-test was negatively related to their comprehension scores at the immediate post-test, though not at the delayed post-test. In contrast, a change in perception towards more anthropomorphism was positively related to word knowledge at the delayed post-test, though not at the immediate post-test.

Possibly, both correlations point again to the role of children's expectations about the robot as a human-like being. If children had low expectations of the robot and the robot exceeded these expectations, they may have become more engaged, which is beneficial for learning. In contrast, children with high expectations which the robot could not meet, may have become disappointed while working with the robot over several tutoring sessions.

There are two important caveats concerning this link between anthropomorphism and word knowledge. First, the correlations, though statistically significant, were rather weak. Moreover, we did not include child characteristics such as age and cognitive ability that could possibly underlie the observed correlations. It is possible that the correlations are spurious and caused by a shared third factor. Second, the present design did not allow for testing the causal direction of the observed correlations. Thus, it is not clear whether children learn more from the robot because they come to perceive it more as human-like, or that they come to perceive the robot as more human-like because they have successful language-learning interactions with it.

4.4 | Limitations, strengths and future research

The current study has several limitations. First, we did not use a standardized questionnaire for anthropomorphism because of our young target group. Standardized tests, such as the Godspeed questionnaire (Bartneck et al., 2009), often use Likert scales or semantic differentials, which are too difficult for young children. In contrast, other measures that are specifically designed for young children and are therefore more appropriate to use (Kory-Westlund & Breazeal, 2019), do not capture the type of human-like properties children attribute to robots. We based our questionnaire on previous work (Jipson & Gelman, 2007) and the questionnaire was found to be reliable, showing also moderate stability between pre-test and post-test. The proposed questionnaire can therefore be seen as a first step towards a validated questionnaire to measure children's anthropomorphism of robots.

Furthermore, we do not know how the introduction of the robot before the pre-test affected the degree to which children anthropomorphized the robot. To ensure that children could establish a common ground with the robot and to decrease any anxiety, the introduction contained several statements about the properties of the robot that related to, amongst others the robot being a peer, speaking as a human, and looking as a human. It is possible that these statements may have biased children's perception towards anthropomorphism at the pre-test. However, administering the anthropomorphism questionnaire prior to the introduction would have had other disadvantages. For instance, it would not have been clear whether children's perceptions were based on actual interactions with similar robots, with different robots, or were based on cartoons, movies or television programs, or just on imagination. The large variation in scores indicates that children still formed their own opinions about the robot, but we do not know whether these opinions were biased towards anthropomorphizing. Note that despite this possible bias, the *changes* in anthropomorphism we observed, in particular at the item level, can be considered genuine and likely to relate to the intensive experience children had with the robot during the lessons.

Moreover, we could only conduct correlational analyses to examine how anthropomorphism and word knowledge were related. In addition, we could not rule out that other child-related factors underlie the relations that were observed between children's anthropomorphism and word knowledge. Future research with field experiments is needed to test whether framing the robot as a machine or as similar to a human affects children's learning differently. A high level of anthropomorphism in itself may not be required for successful tutoring sessions, as no positive main effects of anthropomorphism were found in our study. On the other hand, managing children's expectations of robots especially at first, may be important, as lower initial levels, indicating more reserved expectations of the robot, relate to more word knowledge than when expectations are (too) high.

Furthermore, it is difficult to translate these results to other fields in which technology is used to support learning, such as VR, AR, XR, or serious games. These types of technology often use virtual avatars, which users may anthropomorphize and which may thus be subject to similar relations between anthropomorphism and learning outcomes as in our study. It is possible that, since no differences could be found between the two different robot conditions, interacting with a robot over a longer period of time is more important for children's anthropomorphism than specific behaviours the robot displays, such as gestures. Such behaviours of the robot can still be important for anthropomorphism, but mainly in short interactions (Tung, 2016) and after multiple exposures, the interaction itself becomes more important (e.g., the conversations or type of activity that the child and robot engage in). This would give an indication that our results can also be translated to other fields. However, as we only measured children's perception with a robot, we will need to investigate more thoroughly to determine whether this is the case.

Lastly, there are studies suggesting that presenting robots as human-like to children is undesirable (Broadbent, 2017), as a robot expressing simulated feelings as real feelings is deceptive. Moreover, children may form relationships with robots that may come at the cost of relationships with people. It is important for developers to make sure that children realize that robots are different from human beings. Repeated exposure may more easily reveal a robot's flaws and thus lead to decreases in anthropomorphism, but a subset of the children in our study were found to increase in anthropomorphism, despite our robot's flaws. This is in line with a study finding higher anthropomorphism after repeated exposure (Michaelis & Mutlu, 2018). Thus, even after engaging with a 'flawed' robot, children may continue to anthropomorphize a robot. Therefore, researchers may want to consider whether presenting the robot as a social entity and suggesting it has cognitive, emotional or social abilities is required for their study. After all, even though transparency about the robot's lack of psychological abilities leads to lower anthropomorphism, children may feel as close to the robot as when children's expectations about the robot's psychological abilities are managed (van Straten et al., 2020).

Our study also has several strengths. It is one of the first studies to investigate anthropomorphism and changes therein after children had multiple interactions with a robot, and to relate it to children's word knowledge. Furthermore, we included a large sample of young children. Lastly, the different robot properties presented in the questionnaire allowed for a more thorough and differentiated understanding of the ways in which children perceive robots.

5 | CONCLUSION

The study presented in this article explored the degree to which children anthropomorphize a social robot, whether this had changed after seven tutoring sessions, and whether anthropomorphism correlated with children's word knowledge after these sessions. We found that children generally anthropomorphized the robot, although there were large differences between children in the degree to which they did. Our results showed that children's overall tendency to anthropomorphize had not significantly changed after the tutoring sessions, but the analysis at the item level revealed a complex pattern of changes indicating a shift within this overall tendency towards seeing the robot as more mechanical while at the same time attributing more cognitive capabilities to the robot. As an exploration, we found a weak but significant correlation between children's increased anthropomorphism and their word knowledge. Children who came to perceive the robot more as a human knew more words after the tutoring sessions. Although the causal direction of this relation is not yet clear, the results underscore the importance of taking children's anthropomorphism into consideration when designing robot-assisted tutoring sessions.

ACKNOWLEDGEMENTS

This study was carried out within the L2TOR project, funded by the European Union's Horizon 2020 research and innovation program under grant agreement no. 688014. We are grateful to all members of the L2TOR team who have helped us conduct this study. We would like to thank the children, their parents, and the schools for their participation. Furthermore, we would like to thank Laurette Gerts, Annabella Hermans, Esmee Kramer, Madee Kruijt, Marije Merckens, David Mogendorff, Sam Muntjewerf, Reinjet Oostdijk, Laura Pijpers, Chani Savelberg, Robin Sonders, Sirkka van Straalen, Sabine Verdult, Esmee Verheem, Pieter Wolfert, Hugo Zijlstra, and Michelle Zomers for their help in collecting data.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

PEER REVIEW

The peer review history for this article is available at https://publons.com/publon/10.1111/jcal.12497.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author upon reasonable request.

ORCID

Mirjam de Haas https://orcid.org/0000-0002-1001-8103

ENDNOTES

[1] https://aspredicted.org/6k93k.pdf

[2] Video recordings of all gestures are available at https://tiu.nu/l2tor-gestures

REFERENCES

Alemi, M., Meghdari, A., Basiri, N. M., & Taheri, A. (2015). The effect of applying humanoid robots as teacher assistants to help Iranian autistic pupils learn English as a foreign language. In *International conference on social robotics* (pp. 1–10). Cham: Springer.

Barsalou, L. W. (2008). Grounded cognition. *Annual Review of Psychology*, *59*, 617–645.

Bartneck, C., Kulić, D., Croft, E., & Zoghbi, S. (2009). Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots. *International Journal of Social Robotics*, *1*(1), 71–81.

Belpaeme, T., Kennedy, J., Ramachandran, A., Scassellati, B., & Tanaka, F. (2018). Social robots for education: A review. *Science Robotics*, *3*(21), eaat5954.

Belpaeme, T., Vogt, P., Van den Berghe, R., Bergmann, K., Göksun, T., De Haas, M., … Papadopoulos, F. (2018). Guidelines for designing social robots as second language tutors. *International Journal of Social Robotics*, *10*(3), 325–341.

Beran, T. N., Ramirez-Serrano, A., Kuzyk, R., Fior, M., & Nugent, S. (2011). Understanding how children understand robots: Perceived animism in child–robot interaction. *International Journal of Human-Computer Studies*, *69*(7–8), 539–550.

Bernstein, D., & Crowley, K. (2008). Searching for signs of intelligent life: An investigation of young children's beliefs about robot intelligence. *The Journal of the Learning Sciences*, *17*(2), 225–247.

Biocca, F. (1997). The cyborg's dilemma: Progressive embodiment in virtual environments. *Journal of Computer-Mediated Communication*, *3*(2), JCMC324.

Boerma, T., Chiat, S., Leseman, P., Timmermeister, M., Wijnen, F., & Blom, E. (2015). A quasi-universal nonword repetition task as a diagnostic tool for bilingual children learning Dutch as a second language. *Journal of Speech, Language, and Hearing Research*, *58*(6), 1747–1760.

Breazeal, C., Kidd, C. D., Thomaz, A. L., Hoffman, G., & Berlin, M. (2005). Effects of nonverbal communication on efficiency and robustness in human-robot teamwork. In *2005 IEEE/RSJ International Conference on Intelligent Robots and Systems*, Canada: Edmonton, Alta. (pp. 708–713).

Broadbent, E. (2017). Interactions with robots: The truths we reveal about ourselves. *Annual Review of Psychology*, *68*, 627–652.

Caporael, L. R. (1986). Anthropomorphism and mechanomorphism: Two faces of the human machine. *Computers in Human Behavior*, *2*(3), 215–234.

Chandra, S., Paradeda, R., Yin, H., Dillenbourg, P., Prada, R., & Paiva, A. (2018, March). Do children perceive whether a robotic peer is learning or not?. In *Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interactions*, Chicago, IL, USA. (pp. 41–49).

Dautenhahn, K. (2004, September). Robots we like to live with?! – A developmental perspective on a personalized, life-long robot companion. In *RO-MAN 2004. 13th IEEE International Workshop on Robot and Human Interactive Communication*, Gyeongju, Korea (South). (pp. 17–22).

Dautenhahn, K., & Werry, I. (2004). Towards interactive robots in autism therapy: Background, motivation and challenges. *Pragmatics & Cognition*, *12*(1), 1–35.

de Nooijer, J. A., van Gog, T., Paas, F., & Zwaan, R. A. (2013). Effects of imitating gestures during encoding or during retrieval of novel verbs on children's test performance. *Acta Psychologica*, *144*(1), 173–179.

de Wit, J., Schodde, T., Willemsen, B., Bergmann, K., de Haas, M., Kopp, S., … Vogt. (2018, March). The effect of a robot's gestures and adaptive tutoring on children's acquisition of second language vocabularies. In *Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction*, Chicago, IL, USA. (pp. 50–58).

DiSalvo, C. F., Gemperle, F., Forlizzi, J., & Kiesler, S. (2002, June). All robots are not created equal: the design and perception of humanoid robot heads. In *Proceedings of the 4th Conference on Designing Interactive Systems: Processes, Practices, Methods, and Techniques*, London, England. (pp. 321–326).

Duffy, B. R. (2003). Anthropomorphism and the social robot. *Robotics and Autonomous Systems*, *42*(3–4), 177–190.

Dunn, L. M., Dunn, L. M., & Schlichting, L. (2005). Peabody Picture Vocabulary Test-III-NL, *Dutch version by Liesbeth Schlichting*. Amsterdam: Harcourt Assessment B.V.

Epley, N., Waytz, A., Akalis, S., & Cacioppo, J. T. (2008). When we need a human: Motivational determinants of anthropomorphism. *Social Cognition*, *26*(2), 143–155.

Epley, N., Waytz, A., & Cacioppo, J. T. (2007). On seeing human: A three-factor theory of anthropomorphism. *Psychological Review*, *114*(4), 864–886.

Eyssel, F., Kuchenbrandt, D., Hegel, F., & de Ruiter, L. (2012, September). Activating elicited agent knowledge: How robot and user features shape the perception of social robots. In *Proceedings of the 2012 RO-MAN*, Paris, France. (pp. 851–857).

Fink, J. (2012, October). Anthropomorphism and human likeness in the design of robots and human-robot interaction. In *International Conference on Social Robotics*, Chengdu, China. (pp. 199–208).

Hegel, F., Krach, S., Kircher, T., Wrede, B., & Sagerer, G. (2008, August). Understanding social robots: A user study on anthropomorphism. In *The 17th IEEE International Symposium on Robot and Human Interactive Communication*, Munich, Germany. (pp. 574–579).

Hockema, S. A., & Smith, L. B. (2009). Learning your language, outside-in and inside-out. *Linguistics*, *47*(2), 453–479.

Iverson, J. M. (2010). Developing language in a developing body: The relationship between motor development and language development. *Journal of Child Language*, *37*(2), 229–261.

Jipson, J. L., & Gelman, S. A. (2007). Robots and rodents: Children's inferences about living and nonliving kinds. *Child Development*, *78*(6), 1675–1688.

Kahn, P. H., Jr., Kanda, T., Ishiguro, H., Freier, N. G., Severson, R. L., Gill, B. T., … Shen, S. (2012). "Robovie, you'll have to go into the closet now": Children's social and moral relationships with a humanoid robot. *Developmental Psychology*, *48*(2), 303–314.

Kahn, P. H., Gary, H. E., & Shen, S. (2013). Children's social relationships with current and near-future robots. *Child Developmental Perspectives*, *7*(1), 32–37.

Kanda, T., Hirano, T., Eaton, D., & Ishiguro, H. (2004). Interactive robots as social partners and peer tutors for children: A field trial. *Human–Computer Interaction*, *19*(1–2), 61–84.

Kanero, J., Geçkin, V., Oranç, C., Mamus, E., Küntay, A. C., & Göksun, T. (2018). Social robots for early language learning: Current evidence and future directions. *Child Development Perspectives*, *12*(3), 146–151.

Kelly, S. D., McDevitt, T., & Esch, M. (2009). Brief training with co-speech gesture lends a hand to word learning in a foreign language. *Language and Cognitive Processes*, *24*(2), 313–334.

Kennedy, J., Lemaignan, S., Montassier, C., Lavalade, P., Irfan, B., Papadopoulos, F., … & Belpaeme, T. (2017, March). Child speech recognition in human-robot interaction: Evaluations and recommendations. In *Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interactions*, Vienna, Austria. (pp. 82–90).

Kidd, C. D. (2003). *Sociable robots: The role of presence and task in human-robot interaction* (Doctoral dissertation), Cambridge, Massachusetts, USA: Massachusetts Institute of Technology.

Kory-Westlund, J. M., & Breazeal, C. (2019, June). Assessing children's perceptions and acceptance of a social robot. In *Proceedings of the 18th ACM International Conference on Interaction Design and Children*, Boise ID, USA. (pp. 38–50).

Kory-Westlund, J. M., Martinez, M., Archie, M., Das, M., & Breazeal, C. (2016, August). Effects of framing a robot as a social agent or as a machine on children's social behavior. In *2016 25th IEEE International Symposium on Robot and Human Interactive Communication*, New York City, USA. (pp. 688–693).

Leite, I., Martinho, C., & Paiva, A. (2013). Social robots for long-term interaction: A survey. *International Journal of Social Robotics*, *5*(2), 291–308.

Leite, I., Pereira, A., & Lehman, J.F. (2017). Persistent memory in repeated child–robot conversations. In *Proceedings of the 16th International Conference on Interaction Design and Children*, New York, USA. (pp 238–247).

Lemaignan, S., Fink, J., & Dillenbourg, P. (2014, March). The dynamics of anthropomorphism in robotics. In *2014 9th ACM/IEEE International Conference on Human-Robot Interaction*, Bielefeld, Germany. (pp. 226–227).

Lemaignan, S., Fink, J., Dillenbourg, P., & Braboszcz, C. (2014). The cognitive correlates of anthropomorphism. In *2014 Human-Robot Interaction Conference, Workshop "HRI: A bridge between Robotics and Neuroscience"*, Bielefeld, Germany.

Lemaignan, S., Fink, J., Mondada, F., & Dillenbourg, P. (2015, October). You're doing it wrong! studying unexpected behaviors in child-robot interaction. In *International Conference on Social Robotics*, Paris, France. (pp. 390–400).

Macedonia, M., Müller, K., & Friederici, A. D. (2011). The impact of iconic gestures on foreign language word learning and its neural substrate. *Human Brain Mapping*, *32*(6), 982–998.

Marulis, L. M., & Neuman, S. B. (2010). The effects of vocabulary intervention on young children's word learning: A meta-analysis. *Review of Educational Research*, *80*(3), 300–335.

Michaelis, J. E., & Mutlu, B. (2018). Reading socially: Transforming the in-home reading experience with a learning-companion robot. *Science Robotics*, *3*(21), eaat5999.

Monaco, C., Mich, O., Ceol, T., & Potrich, A. (2018). Investigating mental representations about robots in preschool children. *arXiv preprint arXiv*:1806.03248, Trondheim, Norway.

Moon, A., Troniak, D. M., Gleeson, B., Pan, M. K., Zheng, M., Blumer, B. A., … & Croft, E.A. (2014, March). Meet me where I'm gazing: How shared attention gaze affects human-robot handover timing. In *Proceedings of the 2014 ACM/IEEE International Conference on Human-Robot Interactions*, Bielefeld, Germany. (pp. 334–341).

Mulder, H., Hoofs, H., Verhagen, J., van der Veen, I., & Leseman, P. P. (2014). Psychometric properties and convergent and predictive validity of an executive function test battery for two-year-olds. *Frontiers in Psychology*, *5*, 733.

Obaid, M., Barendregt, W., Alves-Oliveira, P., Paiva, A., & Fjeld, M. (2015). Designing robotic teaching assistants: Interaction design students' and children's views. In *International Conference on Social Robotics*, Paris, France. (pp. 502–511).

Oudgenoeg-Paz, O., Leseman, P. P., & Volman, M. C. J. M. (2015). Exploration as a mediator of the relation between the attainment of motor milestones and the development of spatial cognition and spatial language. *Developmental Psychology*, *51*(9), 1241–1253.

Pereira, A., Martinho, C., Leite, I., & Paiva, A. (2008, May). iCat, the chess player: The influence of embodiment in the enjoyment of a game. In *Proceedings of the 7th International Joint Conference on Autonomous Agents and Multiagent Systems*, Estoril Portugal. (Vol. 3, pp. 1253–1256).

Phillips, E., Zhao, X., Ullman, D., & Malle, B. F. (2018, March). What is human-like? Decomposing robots' human-like appearance using the Anthropomorphic roBOT (ABOT) database. In *Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interactions*, Chicago, IL, USA. (pp. 105–113).

Pot, E., Monceaux, J., Gelin, R., & Maisonnier, B. (2009, September). Choregraphe: A graphical tool for humanoid robot programming. In *RO-MAN 2009 - The 18th IEEE International Symposium on Robot and Human Interactive Communication*, Toyama, Japan. (pp. 46–51).

Riek, L. D., Rabinowitch, T. C., Chakrabarti, B., & Robinson, P. (2009, March). How anthropomorphism affects empathy toward robots. In *Proceedings of the 4th ACM/IEEE International Conference on Human Robot Interactions*, La Jolla, California, USA. (pp. 245–246).

Salem, M., Eyssel, F., Rohlfing, K., Kopp, S., & Joublin, F. (2013). To err is human (-like): Effects of robot gesture on perceived anthropomorphism and likability. *International Journal of Social Robotics*, *5*(3), 313–323.

Sciutti, A., Rea, F., & Sandini, G. (2014). When you are young, (robot's) looks matter. Developmental changes in the desired properties of a robot friend. In *2014 RO-MAN: The 23rd IEEE International Symposium on Robot and Human Interactive Communication*, Edinburgh, Scotland. (pp. 567–573).

Tellier, M. (2008). The effect of gestures on second language memorisation by young children. *Gesture*, *8*(2), 219–235.

Topping, K. J. (2005). Trends in peer learning. *Educational Psychology*, *25*(6), 631–645.

Tung, F. W. (2011, July). Influence of gender and age on the attitudes of children towards humanoid robots. In *International Conference on Human-Computer Interaction*, Orlando, Florida, USA. (pp. 637–646).

Tung, F. W. (2016). Child perception of humanoid robot appearance and behavior. *International Journal of Human-Computer Interaction*, *32*(6), 493–502.

van den Berghe, R., Verhagen, J., Oudgenoeg-Paz, O., van der Ven, S., & Leseman, P. (2019). Social robots for language learning: A review. *Review of Educational Research*, *89*(2), 259–295.

van Straten, C. L., Peter, J., & Kühne, R. (2019). Child-robot relationship formation: A narrative review of empirical research. *International Journal of Social Robotics*, *12*, 325–344.

van Straten, C. L., Peter, J., Kühne, R., & Barco, A. (2020). Transparency about a robot's lack of human psychological capacities: Effects on child-robot perception and relationship formation. *ACM Transactions on Human-Robot Interaction (THRI)*, *9*(2), 1–22.

Verhagen, J., van den Berghe, R., Oudgenoeg-Paz, O., Küntay, A., & Leseman, P. (2019). Children's reliance on the non-verbal cues of a robot versus a human. *PLoS One*, *14*(12), e0217833.

Vogt, P., van den Berghe, R., de Haas, M., Hoffman, L., Kanero, J., Mamus, E., … & Papadopoulos, F. (2019, March). Second language tutoring using social robots: a large-scale study. In *2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, Daegu, Korea: IEEE. (pp. 497–505).

Vygotsky, L. (1978). Interaction between learning and development. In M. Cole, V. John-Steiner, S. Scriber, & E. Souberman (Eds.), *Mind in society: The development of higher psychological processes* (pp. 79–97). Cambridge, MA: Harvard University Press.

Waytz, A., Morewedge, C. K., Epley, N., Monteleone, G., Gao, J. H., & Cacioppo, J. T. (2010). Making sense by making sentient: Effectance motivation increases anthropomorphism. *Journal of Personality and Social Psychology*, *99*(3), 410–435.

Wellsby, M., & Pexman, P. M. (2014). Developing embodied cognition: Insights from children's concepts and language processing. *Frontiers in Psychology*, *5*, 506.

Zaga, C., de Vries, R. A., Li, J., Truong, K. P., & Evers, V. (2017, May). A simple nod of the head: The effect of minimal robot movements on children's perception of a low-anthropomorphic robot. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, Denver, Colorado, USA. (pp. 336–341).

How to cite this article: van den Berghe R, de Haas M, Oudgenoeg-Paz O, et al. A toy or a friend? Children's anthropomorphic beliefs about robots and how these relate to second-language word learning. *J Comput Assist Learn*. 2021; 37:396–410. https://doi.org/10.1111/jcal.12497