Social Robots for Language Learning: A Review

Rianne van den Berghe
Utrecht University

Josje Verhagen
Utrecht University
University of Amsterdam

Ora Oudgenoeg-Paz, Sanne van der Ven, and Paul Leseman
Utrecht University

In recent years, robots have increasingly been implemented as tutors in both first- and second-language education. The field of robot-assisted language learning (RALL) is developing rapidly. Studies have been published targeting different languages, age groups, and aspects of language and using different robots and methodologies. The present review presents an overview of the results obtained so far in RALL research and discusses the current possibilities and limitations of using social robots for first- and second-language learning. Thirty-three studies in which vocabulary, reading skills, speaking skills, grammar, and sign language were taught are discussed. Beside insights into learning gains attained in RALL situations, these studies raise more general issues regarding students' motivation and robots' social behavior in learning situations. This review concludes with directions for future research on the use of social robots in language education.

Technologies such as computers, tablets, and smartphones offer a wide array of possibilities for first- and second-language learning. These forms of technology, in particular interactive white boards, automatic speech recognition programs, instructive virtual games, tablets, and animated books, are increasingly being integrated into language education for both children and adults (Golonka, Bowles, Frank, Richardson, & Freynik, 2014; Takacs, Swart, & Bus,

2015; Young et al., 2012). These technologies allow for forms of language learning that are not always present in traditional classrooms, such as one-to-one and tailored instruction, access to native language input, direct feedback, and the possibility to practice with a virtual agent, which may be less intimidating than practicing with a peer or classmate (Golonka et al., 2014).

One of the newest forms of technology used in education—and the focus of the present review—are social robots. Social robots are robots that are specifically designed to interact and communicate with people, either semiautonomously or autonomously (i.e., with or without a person controlling the robot in real time), following behavioral norms that are typical for human interaction (Bartneck & Forlizzi, 2004). These robots are different from, for example, robotic arms in factories, which are often designed to perform a specific task and generally do not interact with people. They also differ from virtual agents or computer-based intelligent tutoring systems, as social robots always have a physical body of some sort and are therefore present in the real world, rather than being only virtually present via a screen. The field of robotics has developed rapidly over the past decade, leading to the availability of robots that can be employed for educational purposes. In recent experiments, robots have been used as tutors, for example, in teaching prime numbers (Kennedy, Baxter, Senft, & Belpaeme, 2015), puzzle-solving skills (Leyzberg, Spaulding, Toneva, & Scassellati, 2012), and, even more recently, language (e.g., Alemi, Meghdari, & Ghazisaedy, 2014; Kennedy, Baxter, Senft, & Belpaeme, 2016). The main aims of this review are to present the current state of knowledge about robot-assisted language learning (RALL), discuss advantages and disadvantages of RALL, and identify potential areas for future research on this topic.

Robots are presumed to have at least two advantages over most other forms of technology. First, they allow the learner to interact with the real-life physical environment, which is thought to be important for language development (Barsalou, 2008; Hockema & Smith, 2009; Iverson, 2010; Wellsby & Pexman, 2014). Specifically, both the manipulation of real-life objects (Kersten & Smith, 2002) and the use of whole-body movement and gestures (Mavilidi, Okely, Chandler, Cliff, & Paas, 2015; Rowe & Goldin-Meadow, 2009; Toumpaniari, Loyens, Mavilidi, & Paas, 2015) have been found to help children's vocabulary learning. Because of the possibility of acting on the physical environment, robots offer possibilities not provided by traditional computer-assisted lessons, such as manipulating objects and using gestures to support language teaching (e.g., Alemi et al., 2014).

The second advantage is that robots allow for more natural interaction than other forms of technology because of their appearance, which is often humanoid or in the shape of an animal. Many robots can use nonverbal cues such as eye gaze, pointing, and other types of gestures. While this also holds for animated characters on a screen, robots are generally perceived as more helpful, credible, informative, and enjoyable to interact with than animated characters (Kidd & Breazeal, 2004; Wainer, Feil-Seifer, Shell, & Matari, 2007). Furthermore, robots are more likely to be perceived as a typical teacher, peer, or friend rather than as a machine: Both children and adults have a tendency to anthropomorphize robots, that is, to ascribe human-like characteristics and behaviors to robots (Bartneck,

Kulić, Croft, & Zoghbi, 2009; Beran, Ramirez-Serrano, Kuzyk, Fior, & Nugent, 2011; Duffy, 2003). Therefore, robots can be programmed to take up a specific role, for example, the role of a teacher or friend, depending on whether the aim of the learning tasks is to instruct or correct students on a task or to have them practice newly learned information with peers.

Even though it is clear which advantages robots potentially have, there are a number of issues that need to be addressed in order for robots to be effective language tutors (see also Kanero et al., 2018, for a review on early language learning). The present review presents the current state of RALL research, with a special focus on affective aspects such as students' motivation and their responses to robots' social behavior. The overall goal of our review is to gain insight into the potential of robots as first- and second-language tutors and to identify areas for further research. Studies on preschool children, school-aged children, and adults will be reviewed. Throughout our review, studies will be described in relative detail to allow a thorough evaluation of the studies conducted and the possibilities robots offer for supporting language learners.

Our review is organized as follows. First, we describe our search criteria and the studies that were selected for review. Second, we present studies focusing on the effects of RALL on participants' language-learning outcomes. In these studies, word learning has been investigated more extensively than other aspects of language and will be discussed first, followed by a discussion of RALL studies on grammar learning, reading skills, speaking skills, and sign language. Third, we describe studies focusing on the role of affective aspects of RALL, addressing how robots may affect learners' motivation, the role of the robot's novelty, and the effect of robots' social behavior on learning. Finally, we discuss the meaning of these findings and offer directions for future research on the use of social robots for first- and second-language learning.

Method

In our review, we take a narrative approach. Specifically, we synthesize the relevant literature in order to provide a comprehensive overview of the work conducted so far (cf. Cronin, Ryan, & Coughlan, 2008). Given the limited number of RALL studies to date, we have adopted an inclusive approach in selecting studies. We did not apply rigorous criteria with respect to the quality of the studies, as due to the emerging nature of the field this could have led to a loss of information (Arksey & O'Malley, 2005).

Figure 1 shows the search, screening, and identification procedure. Studies were included if they (a) used an empirical design in which language was taught to children or adults (i.e., reviews and studies in which a specific robot or design of a study were proposed were excluded); (b) used a physically present robot (rather than a virtual robot), as we were interested in physical robots that have an embodied presence during the learning task; (c) assessed students' language-learning gains or affective aspects; (d) contained sufficient details to evaluate the design and outcomes (i.e., number of participants, number of target words, learning gains); (e) were published papers in journals or conference proceedings[1]; and (f) were written in English.

```
┌─────────────────────────────────────────────────────────────────────┐
│ Databases searched: PsycInfo, Web of Science, Google Scholar          │
│ Search terms used: robot-assisted language learning, robot vocabulary │
│ learning, robot language teaching, robot children second language,    │
│ robot assisted English learners                                       │
└─────────────────────────────────────────────────────────────────────┘
```

```
┌──────────────────────────────┐   ┌──────────────────────────────────┐
│ Identified articles (without │   │ Inclusion criteria:               │
│ duplicates)                  │   │  • Used an empirical study to teach│
│                              │   │    language                        │
│ Google Scholar: 750          │   │  • Used a physically present robot │
│ PsycInfo: 750                │   │  • Assessed students' learning or  │
│ Web of Science: 160          │   │    affective aspects               │
└──────────────────────────────┘   │  • Contained sufficient details to │
                                    │    evaluate design and outcomes    │
                                    │  • Published in journals or        │
                                    │    conference proceedings          │
                                    │  • Written in English              │
┌──────────────────────────────┐   └──────────────────────────────────┘
│ Title screening yielded:     │
│                              │   ┌──────────────────────────────────┐
│ 102                          │   │ Exclusion criteria:               │
└──────────────────────────────┘   │  • Did not focus on language       │
                                    │    learning                        │
┌──────────────────────────────┐   │  • Did not report on an empirical  │
│ Abstract screening yielded:  │   │    study                           │
│                              │   │  • Proposed a robot or the design  │
│ 56                           │   │    of a study                      │
└──────────────────────────────┘   │  • Reported a subsection of a      │
                                    │    study that was fully reported   │
                                    │    in a later paper                │
┌──────────────────────────────┐   │  • Made use of a virtual robot or  │
│ Full-text screening yielded: │   │    on-screen avatar                │
│                              │   └──────────────────────────────────┘
│ 29                           │
└──────────────────────────────┘
                    ┌──────────────────────────────┐
                    │ Cross-referencing yielded:   │
                    │                              │
                    │ 4                            │
                    └──────────────────────────────┘
┌──────────────────────────────┐
│ Studies meeting criteria:    │
│                              │
│ 33                           │
└──────────────────────────────┘
```
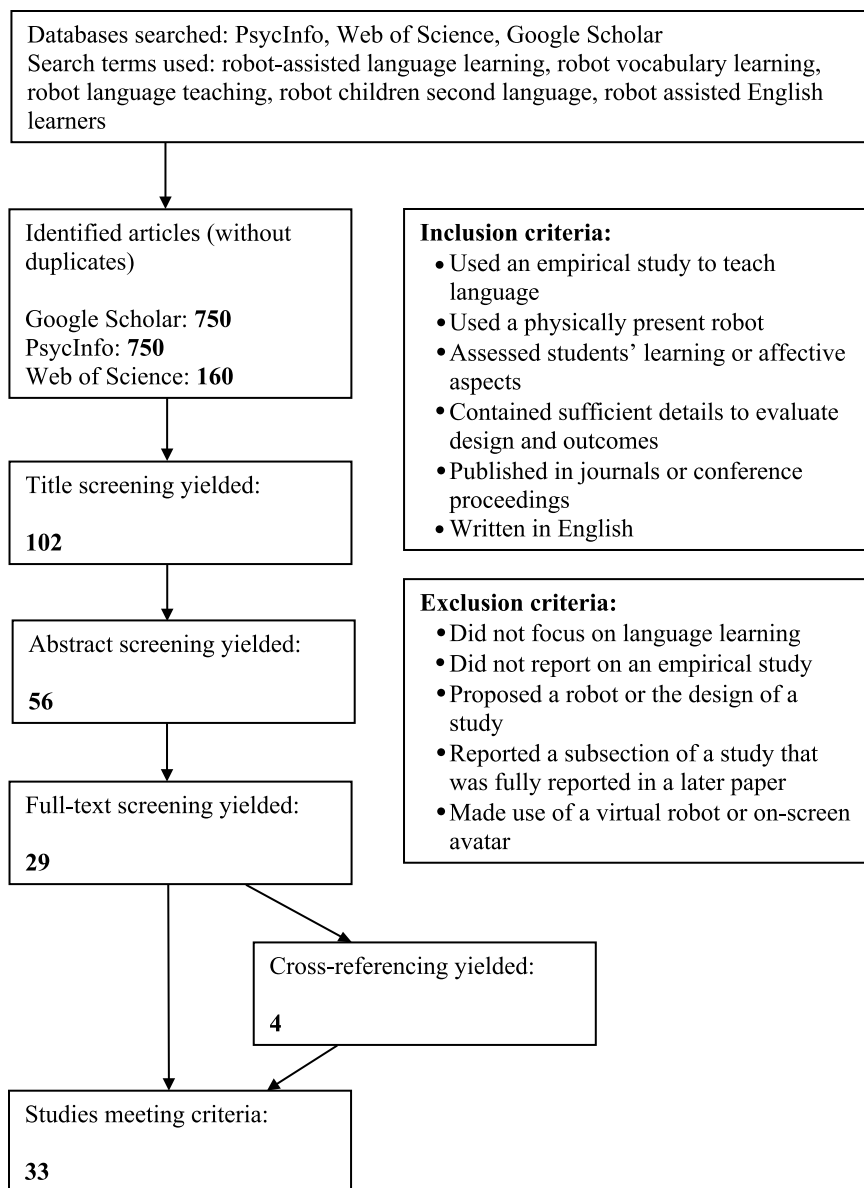
FIGURE 1. *Study selection process.*

Our literature search was conducted using PsycInfo, Web of Science, and Google Scholar. For Google Scholar and PsycInfo, the first 150 results were examined for each search term (cf. Falagas, Pitsouni, Malietzis, & Pappas, 2008;

Shultz, 2007). The following five search terms were used: "robot-assisted language learning," "robot vocabulary learning," "robot language teaching," "robot children second language," and "robot assisted English learners." A total of 750 papers in Google Scholar, 750 papers in PsycInfo, and 160 papers in Web of Science were examined based on their titles. A total of 102 studies were identified as potentially relevant, as their titles included (parts of) our search terms.

After reading the abstracts of all 102 papers, 46 papers were excluded based on the criteria mentioned above. Specifically, we excluded papers that did not report on an empirical study ($N = 14$; e.g., Belpaeme et al., 2015), did not focus on language learning ($N = 13$; e.g., Arsénio, 2014), proposed a specific robot or a design of a study, rather than an empirical study assessing students' (affective aspects of) learning ($N = 14$; e.g., Funakoshi, Mizumoto, Nagata, & Nakano, 2011), or reported on a part of a study (e.g., preliminary results or a subset of the data), which was fully described in a later published paper that was included in the review ($N = 5$; e.g., Tanaka & Ghosh, 2011).

Subsequently, the full texts of the remaining 56 papers were read, and 27 further papers that did not meet the inclusion criteria were excluded. Reasons for exclusion included proposing a specific robot or a design of a study ($N = 10$; e.g., Nagata, Mizumoto, Funakoshi, & Nakano, 2010), not focusing on language learning ($N = 6$; e.g., Hood, Lemaignan, & Dillenbourg, 2015), reporting on a part of a study only ($N = 9$; e.g., Köse et al., 2015), and the use of a virtual robot rather than a physical one ($N = 2$; e.g., Moriguchi, Kanda, Ishiguro, Shimada, & Itakura, 2011).

Thus, 29 studies met the inclusion criteria. The references of these articles were checked and Google Scholar's "cited by" function was used for each of these articles to identify other potentially relevant studies. In so doing, four additional studies that met the inclusion criteria were found, yielding a total of 33 studies for our review.

Information on the design, characteristics, and main findings were extracted from all 33 studies. Studies were then assigned to one of two categories: language-learning outcomes or affective aspects of RALL. Studies on learning outcomes were grouped according to whether they addressed word learning or other language skills. Studies on affective aspects were grouped according to whether they focused on motivational aspects, the robot's novelty, or the robot's social behavior. For an overview of all the studies and their characteristics, see Table 1. For an overview of the types of robots used in these studies and their main characteristics, see the appendix.

Learning Gains in Robot-Assisted Language Learning Studies

Robot-Assisted Word Learning

Word Learning in Preschool and Young School-Aged Children

Out of all 33 RALL studies included in the review, 13 focused on word learning. Most of these included preschool children or children who just entered school. In three of these, children and were presented with words in a second language (L2) or in their first language (L1) over multiple sessions. Pretests indicated that the children did not yet know these words prior to the studies, and posttests indicated that the children learned only few words in each of the three studies.

TABLE 1

Methodological details and results of studies included in this review

Topic	Study	Participants	Method	Robot	Results
Word learning	Alemi, Meghdari, and Ghazisaedy (2014)	N = 46; age = 12; country: Iran	Aim: 45 L2 English words; duration: 5 weeks; 2 conditions: robot-assisted group and control group	NAO; autonomous; role: teaching assistant	Robot-assisted group learned more words and learned faster than control group.
	de Wit et al. (2018)	N = 61; age = 5; country: the Netherlands	Aim: 6 L2 English words; duration: 1 session; 4 conditions: with and without iconic gestures, with and without adaptivity	NAO; Autonomous; role: Peer tutor	Gestures benefited retention and engagement; adaptivity benefited engagement.
	Eimler, von der Pütten, and Schächtle (2010)	N = 18; age = 9–11; country: Germany	Aim: 20 L2 English words; duration: 1 session; 2 conditions: robot-assisted group and pencil-and-paper group	Nabaztag; autonomous; role: peer tutor	Robot-assisted group had a more positive learning experience; There were no significant differences in recall.
	Gordon et al. (2016)	N = 18; age = 3–5; country: United States	Aim: 8 L2 Spanish words; duration: 1 session; 2 conditions: personalized and nonpersonalized	Tega; autonomous; role: peer	Children felt more positive toward the personalized robot, although they did not learn more from that robot.
	Kanda, Hirano, Eaton, and Ishiguro (2004)	N = 228; age = 6 and 11; country: Japan	Aim: 6 L2 English words/phrases; duration: period of 2 weeks; 1 condition: robot-assisted learning	Robovie; autonomous; role: peer tutor	Most children stopped playing with the robot; only children who continued playing with the robot learned some words.
	Kory Westlund et al. (2015)	N = 19; Age = 4–6; country: United States	Aim: 6 L1 English words; duration: 1 session; 3 conditions: learning with adult, tablet, or robot	Dragonbot; teleoperated; role: tutor	Children learned as much from an adult or a tablet as the robot but preferred learning with the robot.
	Kory Westlund, Jeong, et al. (2017)	N = 45; age = 5; country: United States	Aim: 3 L1 English words; duration: 1 session; 2 conditions: flat or expressive storytelling	Tega; teleoperated; role: tutor	Children learned a similar amount of words in both conditions, but children were more engaged for the expressive robot than the flat robot.
	Kory Westlund, Dickens, et al. (2017)	N = 36; age = 2–5; Country: united States	Aim: 6 L1 English words; duration: 1 session; 4 conditions: learning with adult or robot (within-subject) and pictures close or far apart (between-subject)	Dragonbot; teleoperated; role: peer	Children could use nonverbal cues from a robot as well as those from a human. Distance did not affect results.

(continued)

TABLE 1 (continued)

Topic	Study	Participants	Method	Robot	Results
Word learning	Mazzoni and Benvenuti (2015)	$N = 10$; age = 4–6; country: Italy	Aim: 6 L2 English words; duration: 1 session; 2 conditions: with child peer or with robot peer	MecWilly; teleoperated; role: peer	Children learned as much with a robot peer as with a child peer.
	Meiirbekov, Balkibekov, Jalankuzov, and Sandygulova (2016)	$N = 22$; age = 9–10; country: Kazakhstan	Aim: 10 L2 English words; duration: 1 session; 2 conditions: ever-winning or ever-losing robot	NAO; autonomous; peer	Girls learned most from the ever-winning robot, while boys learned most from the ever-losing robot.
	Movellan, Eckhardt, Virnes, and Rodriguez (2009)	$N = 9$; age = 1.5; country: United States	Aim: 10 L1 words; duration: 10 sessions; 1 condition: robot-assisted (with 10 control words)	RUBI-4; autonomously; role: tutor	Children learned one word.
	Schodde, Bergmann, and Kopp (2017)	$N = 40$; age = 24; country: Germany	Aim: 10 Vimmi (artificial language) words; duration: 1 session; 2 conditions: adaptive or random	NAO; autonomous; role: tutor	Participants in the adaptive condition did not perform better than those in the random condition, although they improved more during the training.
	Tanaka and Matsuzoe (2012)	$N = 17$; age = 3–6; country: Japan	Aim: 4 L2 English words; duration: 1 session; 2 conditions: teaching the robot or not (within-subject)	NAO; autonomous; role: younger peer	Children learned words that they taught to the robot better than the words they did not teach.
	van den Berghe et al. (2018)	$N = 67$; age = 4–6; country: the Netherlands	Aim: 6 L2 English words; duration: 1 session; 3 conditions: learning alone, with child peer, or with robot peer	NAO; teleoperated; role: peer	Children learned as much with a robot peer as with a child peer, although they were outperformed by children learning alone on one task.
Reading	Gordon, Breazeal, and Engel (2015)	$N = 46$; age = 3–8; country: United States	Aim: ~11 L1 English words; duration: 1 session; 3 conditions: inquisitive tablet, inquisitive robot, or noninquisitive robot	Dragonbot; autonomous; role: younger peer	Children found the inquisitive robot more engaging but learned more from the noninquisitive robot.
	Hong, Huang, Hsu, and Shen (2016)	$N = 52$; age = 10–11; country: Taiwan	Aim: L2 English skills; duration: unclear; 2 conditions: robot-assisted and control group	Bioloid; autonomous; role: teaching assistant	The robot-assisted group outperformed the control group on reading and listening skills.
	Hsiao, Chang, Lin, and Hsu (2015)	$N = 46$; age = 2–3; country: Taiwan	Aim: emergent L1 literacy skills; duration: 2 months; 2 conditions: robot assisted and tablet-assisted group	iRobiQ; autonomous; role: peer tutor	Children in the robot condition improved more on their storytelling ability, word recognition, and story retelling skills than the tablet-assisted children.

(continued)

TABLE 1 (continued)

Topic	Study	Participants	Method	Robot	Results
	Hyun, Kim, Jang, and Park (2008)	N = 34; age = 4; country: Korea	Aim: L1 reading skills; duration: 4 weeks; 2 conditions: robot-assisted and tablet-assisted group	iRobiQ; autonomous; role: peer tutor	Children in the robot condition improved more on their story-making, story understanding, and word recognition skills than the tablet-assisted children.
Grammar	Herberg, Feller, Yengin, and Saerbeck (2015)	N = 23; age = 10; country: Singapore	Aim: L2 Latin or French rules; duration: 2 sessions; 2 conditions: watchful robot or looking-away robot (within-subject)	NAO; autonomous; role: tutor	Children performed worse on language tasks when the robot was watching them than when the robot looked away.
	Kennedy, Baxter, Senft, and Belpaeme (2016)	N = 67; age = 8; country: United Kingdom	Aim: L2 French rules; duration: 1 session; 3 conditions: high verbal availability, low verbal availability, and control	NAO; teleoperated; role: tutor	The robot's verbal availability did not affect learning gains.
Speaking skills	In and Han (2015)	N = 16; age = 10; country: Korea	Aim: L2 English intonation; duration: 1 session; 2 conditions: adult or child text-to-speech engine	ROBOSEM; unclear; role: tutor	Children did not learn to vary their intonation more with either text-to-speech engine.
	S. Lee et al. (2011)	N = 21; age = 10–12; country: Korea	Aim: speaking and listening skills; duration: 8 weeks; 1 condition: robot-assisted	Mero & Engkey; autonomous; role: tutor	Children improved their speaking skills but not their listening skills.
	Rosenthal-von der Pütten, Straßmann, and Krämer (2016)	N = 130; age = 26; country: Germany	Aim: L2 German skills; duration: 1 session; 6 conditions: prerecorded speech or robot speech, and physical robot, virtual robot, or no robot	NAO; unclear; role: tutor	Participants did not align lexically or syntactically to the virtual or physical robot.
	Wang, Young, and Jang (2013)	N = 63; age = 10–11; country: Taiwan	Aim: L2 English skills; duration: 1 semester; 2 conditions: robot-assisted and traditional class	Tangible learning companions; unclear; role: peer	Children working with the companions had higher motivation, confidence, and engagement.
Sign language	Uluer, Akalın, and Köse (2015)	N = 42; age = 21–32, 6–14, and 10–16; country: Turkey	Aim: 10 signs; duration: 1 session; 2 experiments: one with Robovie and one with the NAO	Robovie R3 and NAO; teleoperated; role: tutor	Experienced sign language learners learned similarly from both robots. Inexperienced learners learned more from the Robovie R3 robot.

(continued)

TABLE 1 (continued)

Topic	Study	Participants	Method	Robot	Results
Motivation	Alemi, Meghdari, and Ghazisaedy (2015)	$N = 46$; age $= 12$; country: Iran	Aim: 45 L2 English words; duration: 5 weeks; 2 conditions: robot-assisted group and control group	NAO; autonomous; role: teaching assistant	Students indicated lower anxiety and a more positive attitude toward learning English.
	Alemi, Meghdari, and Sadat Haeri (2017)	$N = 19$; age $= 3–6$; country: Iran	Aim: L2 English skills; duration: 10 sessions; 1 condition: robot-assisted group	NAO; autonomous; role: teaching assistant	Children enjoyed learning from the robot.
	Han, Jo, Jones, and Jo (2008)	$N = 90$; age $= 10–11$; country: Korea	Aim: L2 English skills; duration: 1 session; 3 conditions: robot-assisted, noncomputer-based media, and web-based instruction	IROBI; autonomous; role: tutor	Children showed higher sustained interest and concentration in the robot-assisted group than in the other groups.
	Shin and Shin (2015)	$N = 66$; age $= 14$; country: Korea	Aim: L2 English conversational skills; duration: 1 session; 2 conditions: robot-assisted and computer-assisted	NAO; unclear; role: tutor	Students in the robot-assisted class participated more and were more satisfied than students in the computer-assisted class.
Novelty	Rintjema, van den Berghe, Kessels, de Wit, and Vogt (2018)	$N = 15$; age $= 5$; country: the Netherlands	Aim: 17 L2 English words; duration 3 sessions; 1 condition: robot-assisted	NAO; teleoperated; role: peer tutor	Children learned most in later sessions, while engagement decreased.
	You, Shen, Chang, Liu, and Chen (2006)	$N = 100$; age $= 10$; country: Taiwan	Aim: L2 English skills; duration: 2 weeks; 1 condition: robot-assisted	RoboSapien; unclear; role: tutor	Attention to the robot declined in the second week
Social behavior	De Haas, Baxter, de Jong, Krahmer, and Vogt (2017)	$N = 59$; age $= 3$; country: the Netherlands	Aim: 4 L2 English words; duration: 1 session; 3 conditions: adult-like feedback, peer-like feedback, or no feedback	NAO; teleoperated; role: peer tutor	Feedback did not affect engagement, but children worked more independently in the peer-feedback condition.
	Saerbeck, Schut, Bartneck, and Janse (2010)	$N = 16$; age $= 10–11$; country: the Netherlands	Aim: Tokipona (artificial language) skills; duration: 1 session; 2 conditions: socially supportive or neutral robot	iCat; autonomous; role: tutor	Children learning from the socially supportive robot learned more and were more intrinsically and task-motivated.

First, in a study on Japanese child learners of English (L2), an English-speaking Robovie robot was put into several classrooms of 6-year-olds and 11-year-olds over a period of 2 weeks (Kanda, Hirano, Eaton, & Ishiguro, 2004). Children were free in choosing how much to interact with the robot and could interact with the robot alone or with class mates. Children engaged in various activities with the robot, such as hugging, singing, and playing rock-paper-scissors. The robot used various English sentences, and the authors tested children's knowledge of six different target words and phrases that were commonly used in the interactions between the robot and the children, for example, "Hello" and "Let's play together." The study showed that learning gains were small. On average, the children knew only one or two of the six words or phrases examined in the posttest (Kanda et al., 2004).

These outcomes are similar to those obtained in a second RALL study on preschoolers' L2 word learning, by Gordon et al. (2016). In this study, a robot that personalized its motivational strategies depending on the child's affective state was used. Specifically, 3- to 5-year-old English-speaking children played several games on a tablet together with a Tega robot over the course of seven sessions in which they were taught a total of eight L2 (Spanish) words. On average, children learned only one or two out of eight words targeted in this study, as indicated by their scores on a posttest. We will discuss this study's results for personalized motivational strategies further in a later section on the effects of robots' social behaviors.

In the last study on preschoolers' word learning to be reviewed, English-speaking children were taught words in their L1 over multiple sessions (Movellan, Eckhardt, Virnes, & Rodriguez, 2009). Specifically, 2-year-old children interacted with a RUBI-4 robot for 10 sessions over a period of 12 days, in which the robot taught the children 10 words through digital and physical games. As in the other two studies, children showed limited learning in this study, as they learned only 1 or 2 out of 10 words.

To summarize, limited learning was found in each of the three studies. In all three studies, picture selection tasks were used to assess children's learning gains. In this type of task, children hear a target word and are asked to choose the picture corresponding to this target word out of several pictures (usually three or four). This task measures receptive vocabulary knowledge, that is, understanding of the meaning of a word. This is in contrast to productive knowledge, or the ability to produce a word with its correct meaning. Crucially, as it is a multiple-choice task, children can also obtain the right answer by guessing, and this chance level should be taken into account when interpreting results. If we do so in interpreting the results of the above studies, it appears that although the children in Gordon et al. (2016) improved as compared to their pretest performance, they did not score above chance level for seven out of eight words. Children did score above chance level in Movellan et al. (2009).

Importantly, in the second study, by Kanda et al. (2004), reviewed above, children were free to determine whether and for how long they interacted with the robot, and children's learning was related to the time they had spent interacting with the robot. Interaction time declined for most children already within the first week, and only children who maintained interest during the second week learned some words and phrases. In the studies by Gordon et al. (2016) and Movellan

et al. (2009), children's interaction time with the robot was not recorded, making it unclear how much time children actually spent playing with the robot and how active they were during the sessions. One possible explanation of the limited learning gains in these studies, therefore, is that children did not stay engaged enough over multiple RALL sessions to learn a substantial number of words.

Three other RALL studies found more positive results for robots teaching L1 or L2 words to preschoolers. Each of these studies used a different approach and, crucially, consisted of only one session. Specifically, children were taught (a) L1 words through shared book-reading with a robot (Kory Westlund, Jeong, et al., 2017), (b) L2 words by teaching a robot words rather than vice versa (Tanaka & Matsuzoe, 2012), or (c) L2 words by playing "I spy with my little eye" with a robot (de Wit et al., 2018). In all three studies, children learned a substantial number of new words.

In the first study, preschoolers were read one of two versions of a story by a Tega robot (Kory Westlund, Jeong, et al., 2017). The results indicated that on a receptive vocabulary (picture selection) task, children responded correctly to two out of three L1 target words on average. Moreover, to measure productive knowledge, a story retell task was used, where children had to retell the story to the experimenter. Results indicated that children used the target words from the story they had heard more often than those from the story they had not heard.[2]

In the second study that showed considerable word-learning gains in preschoolers, 17 Japanese-speaking preschoolers were taught four L2 (English) verbs by an experimenter who used objects to illustrate the meaning of the verbs (e.g., a cup for the verb "drink"; Tanaka & Matsuzoe, 2012). Then, the child taught the robot two of these words, randomly chosen, by making it act out the relevant verb. The results indicated that children learned the words that they taught the robot better than the words that they did not teach the robot, as evidenced in a picture selection task. Children demonstrated more knowledge of the verbs that they acted out than those that they did not act out not only in a direct posttest but also in a posttest 3 to 5 weeks after the training.

The last study showing clear word-learning gains in preschool children, conducted by de Wit et al. (2018), tested the effectiveness of a robot's use of gestures in teaching L2. In this study, 5-year-old children played the game "I spy with my little eye" with a NAO robot that used an iconic gesture to illustrate the meaning of each target word (e.g., it scratched its head and armpit for the word "monkey") with half of the children but did not produce such a gesture with the other half of the children. The children's task was to choose a picture of the animal corresponding to the English target word out of several pictures. Immediate posttest results indicated that children learned, on average, almost three out of six words. There was no immediate effect of the robot's iconic gesture use. However, iconic gestures did benefit retention of the target words: Children who had been presented with iconic gestures in the learning task showed better recall of the words in a delayed posttest a week later than children who had not been presented with iconic gestures.

Overall, these three studies suggest that RALL may benefit children's word learning (de Wit et al., 2018; Kory Westlund, Jeong, et al., 2017; Tanaka & Matsuzoe, 2012). Crucially, all three studies consisted of one session only,

suggesting that effects may differ between single- and multiple-session studies. We will return to this issue later in the section on the novelty effect. Some caution is needed, however, in interpreting the results of these studies. An important limitation of the study by Kory Westlund, Jeong, et al. (2017) is that children's potential prior knowledge of the words was not assessed. The finding that in this study children recognized not only target words but also control words indicates that they had prior knowledge of these words, as these words were not taught explicitly. This leaves open the possibility that they also had prior knowledge of the target words. A possible limitation of the study by Tanaka and Matsuzoe (2012) is that a human teacher was present in addition to the robot to teach children the L2 words. Since no condition was included in which only a robot was present, we do not know whether children are able to learn from teaching a robot by themselves, or whether the learning in this study was mostly due to being taught by a human adult. Finally, since children are known to learn from teaching someone else (Rohrbeck, Ginsburg-Block, Fantuzzo, & Miller, 2003), it is not clear whether children's word learning was due to the activity of teaching itself (i.e., the additional opportunity to practice the target words) or due to teaching a robot specifically. Despite their methodological limitations, the results of these three studies show the potential of using shared book-reading, learning by teaching a robot, and performing language games together with a robot for teaching young children new words in L1 or L2.[3]

An important question is how effective robots are in teaching words in comparison to human teachers. Even though robots are typically not developed with the aim of replacing human teachers, comparisons between robot and human teachers or peers are useful to investigate areas in which robots can complement humans. This question was addressed directly in a study comparing learning gains in an L1 (English) vocabulary-training task provided to preschoolers by a human teacher, a tablet, or a Dragonbot robot (Kory Westlund et al., 2015). Children saw pictures of animals on a tablet and were provided with L1 labels by the human teacher, tablet, or robot. The children in this study learned as much from the tablet or the robot as they learned from the human, that is, four out of six words. Similarly, in a more recent study by the same authors, preschoolers could use nonverbal cues (bodily orientation and eye gaze) of either a human teacher or a robot equally well when mapping unfamiliar L1 (English) words onto pictures (Kory Westlund, Dickens, et al., 2017). Two more studies have investigated how a robot peer compares to a human peer in language-learning experiments. Mazzoni and Benvenuti (2015) found that preschoolers learned as much (i.e., two to three out of six L2 words on average) when working either with a human peer or with a MecWilly robot. Similarly, van den Berghe et al. (2018) found that preschoolers generally learned as many L2 words when learning with a child peer or with a robot peer. However, children learning without a peer altogether showed the highest performance. Note that in this last study, children were provided with L2 words by a human experimenter and played games on a tablet with a child peer, with a robot peer, or without a peer. The presence of the experimenter may have attenuated the possible benefits of a (robot) peer.

This review of studies indicates that (robot) peers do not necessarily lead to higher learning gains than learning without such peers. Rather, the findings of the

studies described above suggest that children may be able to learn equally well when being taught by a robot or by a human teacher, or when being assisted by a robot or child peer.

Word Learning in School-Aged Children and Adults

As discussed above, RALL studies on word learning in young children show a mixed pattern of results, with some studies reporting small learning gains (Gordon et al., 2016; Kanda et al., 2004; Movellan et al., 2009) and others reporting more substantial learning gains (de Wit et al., 2018; Kory Westlund, Jeong, et al., 2017; Tanaka & Matsuzoe, 2012). Studies with older age groups—older school-aged children and adults—demonstrate a more consistent picture, showing clear word learning across studies. However, very different approaches have been taken across studies, with respect to both the role of the robot (i.e., acting like an assistant vs. a teacher) and whether it was controlled by a human or not, making it difficult to compare results directly.

In a study by Meiirbekov, Balkibekov, Jalankuzov, and Sandygulova (2016), the robot was used as a peer learner. Children's task in this study was to play a game together with a NAO robot in which they were provided with a letter and had to select images of words starting with that letter. After one lesson, children were on average able to produce over 3 out of the 10 L2 words that they had been taught.

In contrast, in another study (Alemi et al., 2014), the robot was used as a teaching assistant. Here, a NAO robot assisted in teaching young adolescents L2 (English) words by interacting with the students, making gestures depicting the target words, showing pictures, and telling stories. Students were taught a total of 45 words over the course of 10 sessions. The classes incorporating the robot were compared to an English class that did not have a robot assistant but engaged in the same type of activities. Results indicated that the students in the RALL classes learned faster, learned more, and retained more words than the students educated in the traditional class.[4]

Yet another study (Eimler, von der Pütten, & Schächtle, 2010) had 9- to 11-year-old German children play L2 English games with a Nabaztag robot for one session. The results indicated that children learned almost 14 out of 20 words on average. These are very high learning gains. Crucially, however, these learning gains did not significantly differ from those of children who had been taught these words through paper vocabulary lists. This suggests that children of this age may generally be skilled word learners and obtain high learning gains across different types of vocabulary interventions.

Finally, a study on adults learning words in an artificial language used the robot as a teacher (Schodde, Bergmann, & Kopp, 2017). The participants in this study were taught 10 words in the artificial language Vimmi via an "I spy with my little eye" game. In each trial, a NAO robot asked the participant to find the picture of the target word among distractor pictures. Participants' knowledge of the target words was assessed in an immediate posttest via two translation tasks: one from Vimmi to German and one from German to Vimmi. Participants produced, on average, 7 out of 10 words in the Vimmi-to-German translation task and 3.5 out of 10 words in the German-to-Vimmi translation task. These learning gains

are substantial, especially given that (a) translating words is more difficult than a receptive task (Mondria & Wiersma, 2004), (b) there was only one session, and (c) the learning task consisted of only three trials per target word.

Apart from the different roles assigned to the robot, another aspect that makes RALL studies on word learning difficult to compare is that in many of the studies described above, the robot was teleoperated by the experimenters (see Table 1 for information on whether studies used a teleoperated or autonomous robot). Teleoperation refers to a person controlling the robot, often without the participant's knowledge, in real time. Teleoperation is often the preferred or even the only option for certain tasks, as an autonomous robot (working without teleoperation) would require speech recognition and predefined scripts. Such scripts describe all the steps of an interaction, and the robot cannot divert from this script. Elaborate scripts are needed to have robots respond appropriately to the input, but even then the suitability of the responses cannot be guaranteed due to, among other reasons, the unpredictability of participants' behavior. Thus, previous studies that used an autonomous robot typically consisted of simple designs (e.g., "I spy with my little eye" games on a tablet) that allow for limited variability in the learners' responses (de Wit et al., 2018; Schodde et al., 2017). In more complex settings, experimenters can ensure through teleoperating that the robot answers appropriately, as they can simply type in contingent responses. Hence, given the current state of robot technology and the scientific literature, how effective robots are when operating autonomously remains an open question.

Summarizing RALL Studies on Word Learning Across Age Groups

The L1 and L2 word-learning studies discussed above found mixed results regarding the robot's effectiveness for word learning. Specifically, several studies found only small (Movellan et al., 2009) or no learning gains (Gordon et al., 2016), or learning gains that held only for a subgroup of the children (Kanda et al., 2004). Other small-scale studies with preschool children showed positive effects of the use of robots in word learning and suggest that aspects such as learning by teaching and gestures might improve learning gains (de Wit et al., 2018; Tanaka & Matsuzoe, 2012). However, many studies were based on small samples and/or lacked control conditions and therefore provide only tentative evidence. Studies on school-aged children (Alemi et al., 2014; Eimler et al., 2010; Meiirbekov et al., 2016) and adults (Schodde et al., 2017) suggest that RALL benefits word learning more with these groups than with preschool children. However, direct comparisons between adults and children are needed to support this conclusion. Furthermore, it is important to note that most of the studies showing high word-learning gains employed the robot as a teaching assistant or peer learner rather than as an independent tutor (Alemi et al., 2014; Meiirbekov et al., 2016; Tanaka & Matsuzoe, 2012). Perhaps, in their current form, robots are not sufficiently technologically advanced (e.g., due to difficulties with speech recognition) to be effective tutors on their own. The current evidence bases suggest that teleoperation is still required for robots to be effective tutors and that technological advances and research on which robot behaviors are effective for learning are required to develop effective autonomous robot tutors.

Language Skills Beyond Word Learning

Language use comprises more skills than just vocabulary. These other skills, such as reading, speaking, grammatical skills, and sign language, have been studied less extensively in RALL research than word learning; only 11 of the 33 selected studies addressed (one of) these skills.

Reading Skills

RALL studies on reading skills show that a robot may be beneficial in assisting the teaching of reading skills, either in the function of an assistant or as a tutor. Specifically, comparing an L1 robot-assisted digital book-reading program to the same program without a robot, Hyun, Kim, Jang, and Park (2008) found that preschoolers in the robot-assisted program improved more on story-making, story-understanding, and word-recognizing skills over a 4-week period than children who were not assisted by the robot. Similar results were obtained in another study on early L1 reading (Hsiao, Chang, Lin, & Hsu, 2015). In this study, 2-year-old children followed an early L1 reading program over a period of 2 months, supported either by an iRobiQ robot with a screen or by a tablet without a robot. The results indicated that both groups improved on early literacy tests measuring comprehension, storytelling ability, retelling of stories, and word recognition. However, the children who had interacted with the robot improved more on their storytelling ability, word recognition, and story-retelling skills than children who had worked with a tablet only.

While the results of these two studies are promising, a third study on L1 reading in young children did not find such positive results. In this study, a relatively large group of 46 preschoolers performed a learning task in which they had to find out, together with a Dragonbot robot, how to read words (Gordon, Breazeal, & Engel, 2015). On average, the children learned the written word form of only 1 out of 11 new words. As in some of the other word-learning studies reviewed above (Gordon et al., 2016; Kanda et al., 2004; Movellan et al., 2009), these small learning gains were taken as evidence of the robot's effectiveness by the authors.

The only RALL study on L2 reading that has been performed to date found a positive effect of the presence of a robot teaching assistant on children's L2 (English) reading skills (Hong, Huang, Hsu, & Shen, 2016). In this study, either a human or robot teacher taught 10- to 11-year-old children reading, speaking, and listening skills by reading stories aloud, encouraging children to read sentences out loud, engaging in act-out games, and engaging in question-answer conversations. Children in the robot-assisted classroom outperformed children in the traditional classroom on a standardized English reading test. Children in the robot-assisted classroom were highly motivated by the robot, which may have benefited their learning as compared to children in the traditional classroom. Overall, these findings suggest that there is potential for robots supporting reading skills.

Grammar

Two RALL studies addressed L2 grammar learning, and both demonstrated positive effects of the robot on children's learning. First, Kennedy et al. (2016)

found that a NAO robot positively affected English-speaking children's learning of the French articles "le" and "la." The robot tutor taught 8- to 9-year-old children three rules regarding French articles. The children improved their knowledge of French articles and retained this knowledge in a posttest a week later. In the second RALL study on L2 grammar learning, Herberg, Feller, Yengin, and Saerbeck (2015) investigated children's learning of Latin and French rules, such as those governing plural and article use, in two separate sessions with a NAO robot. The robot either looked at them or looked away during tasks in which the children had to practice the newly acquired information. The study showed that children learned the rules from the robot. Unexpectedly, however, children performed worse if the robot had looked at them, although the effect was found for difficult items in Latin only. A possible explanation of this finding, proposed by the authors, is that instead of representing a comforting social presence during the task and putting the child at ease (which was the intended outcome), the robot increased pressure and, as such, made the children perform worse. These results indicate not only that the specific learning materials and their difficulty may affect experiment outcomes but also that the robot's behavior may affect learning in unexpected ways.

Speaking Skills

Studies addressing L2 speaking skills found mixed results. One study used a ROBOSEM robot to teach Korean-speaking children to use English intonation patterns (In & Han, 2015). Native English speakers vary their intonation more than native speakers of Korean, and less varied intonation shows Korean L2 English learners' nonnativeness. In the study by In and Han (2015), children did not learn to vary their English intonation upon interacting with the robot as compared to their pretest performance. The experimenters concluded that the robot's speech system (as opposed to human speech) is not effective enough to evoke changes in intonation. However, another study, also conducted in Korea and aimed at improving L2 English speaking and listening skills, did find improvement in other speaking skills (S. Lee et al., 2011). Specifically, this study examined children while they were playing with two robots, the Mero robot and the Engkey robot, with the purpose of improving their L2 (English) speaking and listening skills. The study showed that children's L2 listening skills did not improve upon interacting with the robots but that L2 speaking skills (measured through pronunciation, vocabulary, grammar, and communicative ability) did improve. Interestingly, the children in this study improved on all four aspects of speaking skills.

Even though both studies involved the same L1 and L2, they show opposing results, as the participants in S. Lee et al. (2011) improved their L2 pronunciation upon interacting with the robot, whereas the children in the study by In and Han (2015) did not. Contradictory results were also found in two studies that compared robot-assisted classrooms to traditional classrooms in teaching L2 English speaking skills to Taiwanese children: In one study, children improved their speaking skills more in the robot-assisted classroom than children in traditional classrooms (Wang, Young, & Jang, 2013), while in another study, children in a robot-assisted classroom outperformed students in a traditional classroom on L2

listening but not on L2 speaking (Hong et al., 2016). This contrast in results may be due to the different scope of the L2 training: The training of Wang et al. (2013) was aimed only at teaching speaking skills, while the training of Hong et al. (2016) was also aimed at teaching listening, reading, and writing.

Conflicting results across studies targeting the same skill (i.e., L2 speaking skills) in very similar participant groups may be due to the various ways in which speaking skills were evaluated. Speaking skills can be assessed in different ways, for example, by measuring intonation, speech rate, pronunciation, vocabulary, and grammatical complexity. Given the very different operationalizations of (L2) speaking skills in earlier work, future work on RALL assessing these different aspects would be useful to identify which speaking skills benefit most from robot tutoring. In pursuing this line of research, an important recommendation is that studies target the same L1s and L2s to test the effectiveness of robots for teaching speaking skills, as most L2 speaking skills are heavily dependent on learners' L1.

Before we conclude this section on RALL research on L2 speaking, a final study is noteworthy, in which adults' L2 lexical and syntactic alignment behavior was assessed. Lexical and syntactic alignment refers to the degree to which speakers adapt their words and sentence structures to those of their conversational partner (Rosenthal-von der Pütten, Straßmann, & Krämer, 2016) and thus involves a very different type of learning (implicit vs. explicit) and skill than the type of speaking skills (e.g., pronunciation and intonation) discussed above. Rosenthal-Von der Pütten et al. (2016) compared the L2 (German) alignment behavior of adults with various L1s to a physical robot, a virtual robot, and a computer system with prerecorded speech without a (virtual) agent. Contrary to the authors' expectations, participants showed no alignment to the physical or virtual NAO robot or the computer system (i.e., they did not use similar words and sentence structures). Furthermore, there were no significant differences in the perceived human likeness of the robot's text-to-speech system (i.e., the system that converts text into spoken voice output) and the prerecorded human speech. This is a striking result, as text-to-speech systems are often of inferior quality to human speech. It may also explain the absence of alignment effects: Participants may not have felt the need to align to a computer with such an advanced speech system. Note that alignment may also not result in implicit learning if the speech system is perceived to be of inferior quality: Learners may not learn advanced vocabulary or grammar from inferior systems. Clearly, more research is needed on how a robot's text-to-speech system affects L2 language learning in general and the learner's pronunciation of L2 words in particular.

Sign Language

RALL studies on sign language are nearly absent, and only 1 out of the 33 in our review addressed this topic. In this study, robots were found to be able to teach sign language successfully to various types of learners (Uluer, Akalın, & Köse, 2015). Uluer et al. (2015) compared the effectiveness of two robots in teaching Turkish sign language to three groups of Turkish participants: hearing adults, hearing children, and hearing-impaired children. The first robot, a Robovie R3 robot, has hands with five independent fingers, allowing for the production of signs that are more accurate than those of most other robots. The second, a NAO

robot, has only three fingers that cannot be moved independently. The three participant groups played imitation and act-out games with the robots. The results indicated that all groups learned most of the signs from the robot. Even though there was no difference between the effects of the two robots for the experienced sign language learners, the Robovie R3 robot resulted in significantly higher learning gains than the NAO robot in the inexperienced learners (typically hearing groups, who, unlike the hearing-impaired children, were novices in sign language). Thus, considering the specific characteristics of the robot seems especially relevant in learning situations like these, which rely more on the robot's physical interaction possibilities.

Summary

Studies on language skills other than word learning are rare in RALL research. Also, they are typically diverse, in the sense that they have looked at different age-groups and used very different research designs. The available studies, albeit few in number, suggest that a robot can successfully assist in teaching reading skills (Gordon et al., 2015; Hong et al., 2016; Hsiao et al., 2015; Hyun et al., 2008), grammar learning (Herberg et al., 2015; Kennedy et al., 2016), and sign language (Uluer et al., 2015), either in L1 or in L2. The evidence with respect to speaking skills is more mixed (Hong et al., 2016; In & Han, 2015; S. Lee et al., 2011; Rosenthal-von der Pütten et al., 2016; Wang et al., 2013) and may differ depending on which types of speaking skills are assessed (e.g., pronunciation, intonation, lexical alignment; cf. In & Han, 2015; S. Lee et al., 2011; Wang et al., 2013, for pronunciation; Rosenthal-von der Pütten et al., 2016, for alignment).

Affective Aspects of Robot-Assisted Language Learning

Robots' Positive Effects on Motivation

Robots not only affect language-learning gains but may also affect students' learning strategies and motivation to learn. Given that motivation has been found to be positively related to learning achievements (Dörnyei, 1994; Pekrun, Goetz, Titz, & Perry, 2002), it is important to look at how the use of robots in language-learning studies affects students' motivation. Previous studies indicate that robots generally have a positive effect on students' motivation in RALL.

A number of studies comparing a robot-assisted classroom to a traditional classroom found higher student motivation in robot-assisted classrooms than in traditional classrooms. In the study by Alemi et al. (2014) on L2 word learning in school-aged students that was reviewed above, robot-assisted students indicated that they felt very positive about learning with a robot. As discussed earlier, learning outcomes in this study indicated that the robot-assisted students learned faster, learned more, and retained more than the students in the traditional class. In fact, the students in the robot-assisted class needed less than a third of the time required by the traditional class to work through the materials.

The effects of RALL on these students' learning-related emotions were reported in a follow-up article (Alemi, Meghdari, & Ghazisaedy, 2015). Using self-report questionnaires, the authors found that students were less anxious to make mistakes and less self-conscious about making mistakes in the presence of the robot

than in the presence of a human teacher. Similar positive effects were found in studies on speaking skills. In Wang et al. (2013), 10- to 11-year-old students who learned together with a robot companion also displayed higher confidence, motivation, and engagement than children in a traditional classroom. A positive effect on students' motivation was also found by S. Lee et al. (2011), who observed that a robot improved learners' self-reported satisfaction, interest, confidence, and motivation. Finally, the 9- to 11-year-old children in the study by Eimler et al. (2010) were found to have a more positive learning experience when learning L2 English words with assistance from a robot than when they were taught these words through paper vocabulary lists, even though there were no significant differences in word learning between the two conditions.

Other studies have compared the motivational aspect of the robot to other types of technology. The preschool children in Hsiao et al.'s (2015) reading experiment participated much more actively when assisted by a robot: They engaged more in reading, singing, and replying to questions than when working without robot. An observational study found that preschoolers in an L2 (English) learning class showed less anxiety, higher motivation, and higher engagement after interacting with a robot multiple times (Alemi, Meghdari, & Sadat Haeri, 2017). Furthermore, in a study comparing the at-home use of the IROBI robot for L2 (English) language learning to noncomputer-based media and web-based instruction, 10- to 12-year-old children working with a robot showed longer sustained interest and concentration than the other groups (Han, Jo, Jones, & Jo, 2008). Similarly, 14-year-old students were found to participate more and to be more satisfied when working with a NAO robot in an L2 (English) conversation class than when working with a computer (Shin & Shin, 2015). The students' motivation did not differ across conditions. These results must be interpreted with caution, however, as the students working with the robot engaged in an additional group conversation with the robot and thus had more exposure to the technology. Last, in Kory Westlund et al.'s (2015) study, preschoolers' learning with a robot was compared to learning with a tablet and a human teacher. The authors found that almost all the children preferred learning with the robot to learning with the human teacher or the tablet. Note, however, that this preference did not lead to higher learning outcomes. In summary, robots seem to have a more positive effect on students' motivation than other types of technology, such as tablets or web-based programs.

The positive effects of robots on learning-related emotions have been found not only in RALL studies but also in studies looking at other types of robot-assisted learning, such as programming and drawing and interpreting graphs (Chin, Hong, & Chen, 2014; E. Lee & Lee, 2008; E. Z. F. Liu, Lin, & Chang, 2010; Mitnik, Recabarren, Nussbaum, & Soto, 2009; Nourbakhsh et al., 2005). Interestingly, the picture that emerges from the literature on affective aspects of robot-assisted learning is much clearer than that on language-learning gains: The assistance and/or presence of social robots has a positive effect on students' engagement, attitude and motivation, and this holds across domains (language vs. other domains) and across age groups. This suggests that the potential of robots lies mainly in their ability to motivate students.

Interestingly, such positive effects on motivational aspects are generally not found for other types of technology, such as interactive white boards, blogs, and

virtual worlds, for which only weak evidence of positive effects is found (Barrett & Liu, 2016; Golonka et al., 2014). It should, however, be studied further, as people are likely to differ in the degree to which they feel intrinsically motivated to make use of technology for language learning (Stockwell, 2008). Future research could investigate the degree to which students are intrinsically motivated to work with robots and whether and how the positive effects of robots on motivation could benefit students' language learning. One caveat is noteworthy here: It is not completely clear to what extent the boost in motivation is due to the motivational actions of the robot itself or to the novelty of the robot. On the basis of the current state of knowledge, the possibility remains that the robot initially boosts motivation but that this effect fades out over time as people become accustomed to working with robots. This possibility will be discussed further in the next section.

The Novelty Effect in RALL research

Robots often spark a lot of enthusiasm in their users. This excitement can result in a so-called novelty effect on learning: Learners enjoy the new technology so much that their initial interest leads to higher learning outcomes, which would not have been attained if learners had been more familiar with the robot (cf. S. Liu, Liao, & Pratt, 2009). Once learners become used to the technology, their interest and boost in learning outcomes fade away. This effect might be particularly influential in experiments involving one session or a small number of sessions. In fact, it may, at least in part, explain why one-session word-learning studies found higher learning outcomes than word-learning studies consisting of multiple sessions.

Many authors do not report on how novelty may have affected their results or on how they controlled for a novelty effect. Some one-session experiments have addressed the issue of the novelty effect by having the children play with the robot for a few days before the actual experiment (e.g., Han et al., 2008). It is not clear, however, whether this procedure attenuated the novelty effect in this study, as the experiment itself consisted of only one session.

Studies reporting on students' interest in robots over time found mixed results. In the previously discussed study on L2 word learning by Kanda et al. (2004), the amount of time in which children wanted to interact with the robot quickly decreased within 2 weeks, and this decrease in interaction time with the robot in turn attenuated the learning effect. Specifically, in this study, learning gains were found only for the children who continued playing with the robot, a subset of about a quarter of the 200 participants in the study. Moreover, the continued interaction was not due to sustained interest. Rather, most children indicated that they continued playing with the robot out of pity (Kanda et al., 2004).

A similar decline in interest in working with the robot is reported in a study in which a RoboSapien robot assisted a teacher in English classes, engaging in several activities such as storytelling, answering questions, cheerleading, gesture games, and pronunciation exercises (You, Shen, Chang, Liu, & Chen, 2006). Overall, the children enjoyed the robot, although the attention they paid to the robot declined in the second week. After two lessons, children had already gotten used to the robot and became less interested in working with it. Language-learning

gains were not assessed, so it is not clear whether the decline in children's motivation affected learning.

Similarly, a decline in task engagement was found over the course of three sessions in a study on preschoolers learning L2 English words with a NAO robot (Rintjema, van den Berghe, Kessels, de Wit, & Vogt, 2018). This decline in task engagement did not affect learning gains, as children learned more in later sessions than in the first session. These results need to be interpreted with caution, as the specific target words taught in the lessons and the type of the lessons were not counterbalanced. Therefore, it is not clear whether changes in engagement and learning were due to a (dissipating) novelty effect of the robot or to differences in the content of the lessons. However, the studies discussed above do show that further development of both technology and content is needed to sustain children's interest and to make children enjoy interactions over time in order for robots to become effective learning companions or tutors.

In contrast to the studies summarized thus far that showed a decline in participants' interest in the robot, two previous studies found that participants sustained interest in working with a robot over a longer period. In the first, by Alemi, Meghdari, Mahboub Basiri, and Taheri (2015), a relatively large sample of students reported positive experiences after having worked with a robot for 10 sessions over 5 weeks, suggesting that they maintained their interest in the robot over multiple sessions. A possible explanation is that the robot functioned as an assistant to a human teacher, and that the teacher and robot together could sustain students' interest for a prolonged time. If a robot is solely responsible for maintaining an interaction, the behavioral and conversational demands on the robot's social interactional skills are higher than if a human teacher can act as a mediator.

The second study showing sustained interest found that the robot could maintain children's attention even if it interacted with the child independently, at least in very young children (Hsiao et al., 2015). The toddlers in this study interacted with a robot twice a week for a period of 8 weeks. Children did not lose interest in the robot and participated equally actively in the last 4 weeks as in the first 4 weeks. Note, however, that children in the control condition who worked with a tablet also sustained their interest over this period. This suggests that the e-book that was used as teaching material in both conditions and which contained many different activities was interesting enough to sustain interest over a long time period.

Raising and maintaining participants' interest is crucial to successful interactions, and recent work has addressed the issue of maintaining interest in RALL situations. Specifically, Han, Kang, Park, and Hong (2012) conducted several pilot RALL studies with an iRobiQ robot, and concluded that there are several strategies to encourage sustained interaction between a robot and children. These strategies are mostly focused on making the child seem important to the robot. This can be achieved by having voice- or face-tracking systems recognize and track the child, using pictures of the child on the screen, "remembering" the child's learning history, or working around quirks (e.g., framing quirks by telling the robot's "birth story"; Han et al., 2012). Therefore, a key recommendation for future RALL studies, according to these authors, is to teleoperate the robot in order to tailor the robot's speech and actions to the specific behavior and needs of

an individual child. Currently, artificial intelligence, visual recognition systems, and automatic speech recognition systems clearly do not yet allow for robots to interact autonomously with a child in such a way that the child will remain interested in the robot.

This recommendation is in keeping with the conclusion of a review of several (mostly non-RALL) robot studies in which robots interacted with children and adults over longer periods of time (Leite, Martinho, & Paiva, 2013). These studies found varying results, from a clear, short-lived novelty effect (Fernaeus, Håkansson, Jacobsson, & Ljungblad, 2010) to sustained interest over a period of 5 months (Tanaka, Cicourel, & Movellan, 2007). The authors propose several guidelines to encourage long-term interaction with social robots, involving the robot's appearance, behaviors, affect, memory, and adaptation. For example, one recommendation is that the robot should have both routine behaviors, such as greetings, as well as new and personalized behaviors over time (e.g., adding new games or suggesting games that match participants' interests). It is likely that the effectiveness of behaviors aimed to increase or sustain learners' interest differs per target group (e.g., depending on age, gender, or subject), and a robot's behaviors should be focused on its audience.

In short, the novelty effect is an important issue to be taken into account in robot studies. At least some results on learning-related emotions and learning gains obtained in previous robot studies are likely to stem from the initial excitement when learners work with a robot for the first time. Some ways in which long-term interaction could be fostered involve working around technical limitations (e.g., teleoperating the robot) or increasing (diversity in) the robot's social behavior. The next section will outline in more detail the outcomes, and concomitant complexities, of earlier work on robots' social behavior, and in particular on their supportiveness and motivational behavior.

The Complexity of Robots' Social Behavior

As noted in the Introduction, one of the advantages of robots is their appearance, and therefore their potential benefits in establishing more natural interactions. Robots can be programmed to behave socially via both nonverbal behaviors (e.g., gaze, body posture) and verbal behaviors (e.g., giving praise, saying someone's name). This section reviews evidence relating to robot behavior's ability to increase students' motivation and learning outcomes.

Several studies have examined how a robot's social behavior may positively affect learning gains. In one of these, the effect of social support on children's ability to learn an artificial language was investigated. Specifically, Saerbeck, Schut, Bartneck, and Janse (2010) studied how 10- to 11-year-old children interacted with the iCat to learn an artificial language. The experimenters manipulated the degree to which the robot was socially supportive, such that in one condition the robot engaged in a social dialogue, while in the other condition the robot focused solely on the desired transfer of knowledge. Children interacted with the robot for equally long periods across the two conditions, but children working with the socially supportive robot learned more and were more intrinsically motivated, as they reported having had more fun than children working with the neutral robot. This finding is similar to that of Gordon et al. (2016), who found that

children felt more positive toward a personalized robot. Crucially, in this study, the robot adapted its motivational utterances to the child's affective state (e.g., excited, thinking, or frustrated). Note, however, that this did not lead to higher learning gains, in contrast to the results of a non-RALL study by Leyzberg, Spaulding, and Scassellati (2014), in which a personalized robot tutor resulted in higher learning outcomes than a nonpersonalized tutor in a puzzle-solving task. Such mixed findings indicate that personalization is an important avenue for future research on exactly how robots can be used as effective tutors. Note that the two studies used very different age groups (preschoolers and adults), and personalization may affect age groups differently.

Another type of robot social behavior that may benefit learning is the robot's expressiveness (Kory Westlund, Jeong, et al., 2017). Using storytelling to teach preschoolers L1 (English) words, these researchers compared an expressive robot to a "flat" robot. The expressive robot spoke in a more expressive way, with changes in its intonation. The flat robot did not vary the intonation of its utterances. Both voices were recorded by a female adult rather than created via text-to-speech systems, as a computer-generated voice cannot reach the same variation in intonation as a human voice. The expressiveness of the robot did not affect how many target words children recognized or how they perceived the robot. However, children in the expressive condition used more target words in their retellings, told longer stories in a delayed posttest 4 to 6 weeks later, and were more likely to imitate the robot's phrasing in their story retellings. Crucially, concentration, engagement, and surprise (but not attention during the story) were significantly higher for children in the expressive condition than in the flat condition. Thus, although children did not learn more words receptively when interacting with an expressive robot, the expressiveness of the robot had a positive effect on the way in which children were involved in the task and on their production and retention of the target words.

Other aspects of social behavior, however, do not seem to have such positive effects on language learning. Specifically, previous work on the effects of verbal availability, feedback, and adaptivity has shown mixed results. The term *verbal availability* refers to a robot's sensitive response towards a student, for example, by using the student's name, giving praise, or asking for the student's opinion (Kennedy et al., 2016). In the study by Kennedy et al. (2016) that was also discussed above, a NAO robot taught 8- and 9-year-old children French articles, showing either high or low verbal availability. High verbal availability did not result in greater learning gains. Interestingly, however, another study by the same authors reporting on a math-learning experiment found that the NAO's nonverbal availability (i.e., the use of nonverbal cues such as gaze and posture to attend to the student) did affect children's learning gains positively (Kennedy et al., 2015), indicating that the effects of verbal and nonverbal availability may defer depending on the specifics of the tasks used.

Similarly, the effect of a robot's feedback during RALL tasks is unclear. To date, only one study has directly compared the effects of several types of feedback in a RALL task. In this task, 3-year-old children were taught L2 English count words by a NAO robot (de Haas, Baxter, de Jong, Krahmer, & Vogt, 2017). There were three conditions in which children were given (a) explicit positive and

implicit negative feedback (i.e., adult-like feedback), (b) explicit negative feedback (i.e., peer-like feedback), or (c) no feedback. The authors did not assess children's vocabulary learning gains but studied how feedback affected children's engagement, as measured via eye gaze and the amount of help children needed from the experimenter. The study showed that children looked more often at the experimenter in the no-feedback condition than in both feedback conditions, and that they needed more help from the experimenter in the explicit positive- and no-feedback conditions than in the explicit negative-feedback condition. There were no differences in the duration of the gaze toward the stimulus materials and the robot across the three conditions. This study indicates that the way in which children engage in a learning task is affected by the feedback the robot provides, but more research is needed to assess how a robot's feedback affects learning gains and motivation, and ideally to compare how children respond to feedback from robot and human tutors.

A final behavior that may affect learning is adaptivity. This is an area worth exploring, since robots can, at least in theory, be programmed to provide adaptive tutoring. Only two studies to date have studied the effects of robot adaptivity on language learning. In the first study, by Schodde et al. (2017), an adaptive robot system was compared to a random system in teaching German adults words from an artificial language called Vimmi. The adaptive robot system selected which item to teach (depending on the items that the participant showed difficulty with) and the difficulty of the item (i.e., the number of distractors). The adaptive robot did not result in higher scores on two translation tasks (from Vimmi to German and from German to Vimmi) than the random robot, but participants in the adaptive condition improved more within the "I spy with my little eye" game (i.e., they found the right target more often) than participants in the random control condition. Schodde et al. note that the fact that the participants' greater improvement did not result in higher learning gains could be due to the difficulty of the translation tasks as compared to the leaning task. If the participants' knowledge had been measured receptively, a benefit of adaptivity might have been found.

However, in the second study examining the effects of a robot's adaptivity on language learning (de Wit et al., 2018), no positive effects of adaptation were found on word knowledge tasks either. In this study, which is also discussed above, de Wit et al. (2018) investigated the effect of adaptivity on Dutch preschoolers' learning of L2 English animal names. For half of the children, the "I spy with my little eye" game was adapted to the child's needs (e.g., fewer distractor pictures for difficult target words), and for the other half of the children, the difficulty was not adapted. While children in the adaptivity condition remained engaged during the game, in contrast to the children in the no-adaptivity condition, adaptivity did not result in higher learning gains. As these studies do not convincingly show that adaptivity results in higher learning gains, more research is needed to study the effect of adaptive systems and to confirm the importance of adaptivity in RALL.

Thus, across studies, there are contradictory results with respect to the effects of the robot's personality and social behavior on learners' motivation and learning outcomes. Although one could adopt a "no harm in trying" policy with regard to incorporating personalized or social behavior in child-robot interactions, other

experiments indicate that caution is needed, as social behavior does not always lead to higher learning gains. In the study by Gordon et al. (2015), for example, a robot that was not curious led to higher learning gains than a curious robot that showed excitement about the learning task and interest in the progression of the story. A possible explanation for this finding is that the curious robot was indeed more engaging but distracted participants from the learning task and therefore resulted in smaller learning gains compared to a less engaging robot.

These results are reminiscent of the study by Herberg et al. (2015) described earlier, which showed that children performed worse when a robot looked at them during tasks than when it looked away. Finally, support for the idea that social behavior may result in lower learning gains comes from an experiment in which prime numbers were taught by a NAO robot to 7- and 8-year-old children (Kennedy et al., 2015). In this study, an antisocial robot (which actively avoided gaze) resulted in greater learning gains than a social robot. In-depth analyses revealed that children spent more time looking at the social robot than the antisocial robot, thus looking less at the educational content provided by the tablet. These findings suggest that when a robot is too social, it can distract the child and actually make the child learn less than when the robot is less social, at least when the educational content is provided by an external medium such as a tablet and not by the robot itself.

A finding that adds to the complexity of the effects of a robot's social behavior on task interest and learning is that gender can play a role in the beneficial or adverse effects of the robot's social behavior. As discussed above, the children in Meiirbekov et al.'s (2016) study learned to produce over three new L2 words when working with a robot. The experimenters compared learning gains for children working together with a robot that would always either win or lose the game to assess whether the robot making mistakes would make the child feel more at ease during the learning process. Interestingly, the child's gender determined which robot version led to the highest outcomes: Girls learned twice as many words as boys from the ever-winning robot, whereas boys learned twice as many words as girls from the ever-losing robot. The experimenters did not offer any possible explanations for this interaction effect of robot condition and gender, but it may suggest that there are differences between girls and boys in empathy or competitiveness, that is, in how they perceive the different robot versions (e.g., they may feel sorry for the robot when it loses or focus on their own wins) and in how they engage with the robot when it always either wins or loses.

To summarize, the existing evidence with respect to the robot's social behavior is mixed. A robot showing social behavior such as producing the child's name can increase children's engagement in learning tasks. At the same time, the social behaviors of the robot can distract children from learning and, as a consequence, result in poorer learning outcomes. Moreover, there may be interaction effects with child characteristics such as gender, and results may differ depending on learners' sociocultural backgrounds. The studies listed above have been conducted in countries all over the world (e.g., the United States, the United Kingdom, Singapore), and the different contexts may affect how children respond to the robot's behavior. Moreover, the studies reviewed in this section involved a single session only, and it is not clear whether effects of robots' social behavior differ

when learners interact with robots over multiple sessions. Thus, it is still difficult to disentangle the effects of the robot's social behavior shown from the novelty effect discussed above. Future research should try to optimize the social behavior of the robot for different learning tasks (e.g., grammar learning vs. speaking skills) and different groups of learners (e.g., preschool vs. school-aged children, girls vs. boys) and incorporate adaptivity and feedback.

Discussion

The goal of this review was to provide an overview of the current evidence on RALL and to identify potential topics for future research regarding the use of robots for language teaching. Thirty-three studies addressing word learning, reading skills, grammar learning, speaking skills, and sign language have been discussed, focusing on two important aspects: (a) the robot's effect on children's L1 and L2 language-learning gains and learning motivation and (b) the way robots should behave to maximize learning outcomes. Below, these aspects will be discussed separately, followed by a discussion of possible avenues for future research.

Mixed results were found with respect to L1 and L2 language learning outcomes. Most studies focused on word learning and did not clearly show whether robots are effective for word learning. More research is needed to determine the most effective role for the robot (e.g., teaching assistant or peer learner), the age-groups for which robots are most beneficial (e.g., preschool children, school-aged children, or adults), and the optimal number of sessions for word learning (one or multiple). The few studies examining reading skills, grammar learning, and sign language showed quite positive results, while the evidence with respect to speaking skills is more mixed. Note that the studies made different comparisons: Studies on grammar learning and sign language compared different robot behaviors or platforms to assess the most effective robot (behavior), while the studies on reading and speaking skills compared the effectiveness of a robot to other types of technology or traditional classrooms. Moreover, the conflicting results between skills may result from differences in demands on the robot's interactional qualities (e.g., being able to have contingent conversations), which are likely higher in lessons on speaking skills than in lessons on reading or grammar. Lessons on reading and grammar can be mediated through a tablet or other devices that display words or rules (thus combining the robot with other types of technology), while robots cannot fall back on such devices and need more skills (e.g., speech recognition, natural language generation) when practicing speaking skills with learners.

In contrast to the studies on language-learning outcomes, which showed mixed results, the studies addressing participants' learning-related emotions generally found positive effects and showed that both children and adults often enjoy working with the robot. Given that learning motivation and learning gains are often related (Dörnyei, 1994; Pekrun et al., 2002), the robot's potential to motivate learners could be a valuable property. However, higher motivation was not always linked to higher learning gains in the RALL studies reviewed, and motivation could, at least in part, be due to the initial novelty effect of robots, which soon disappears. Although addressed in some experiments (e.g., Han et al., 2008), it is not clear how novelty has affected the results of previous studies. A strong recommendation is to carefully consider how to introduce the robot to participants to

minimize novelty effects and to see whether the effects found are robust to prolonged exposure or wear off over time.

Conflicting results were especially striking with respect to the social behavior of the robot. Although some studies found positive effects of personalized and/or social behavior on learning gains and enjoyment, other studies found social behavior to negatively affect learning outcomes and behavior. Thus, it is clear that there is a thin line between the robot being social enough to sustain children's interest and being too social, leading to children being distracted or even intimidated by the robot. Furthermore, adaptivity and feedback have remained understudied and should have a more central role in future studies, given the importance of adapting to the learner's level and providing helpful feedback in L2 education and education in general (Li, 2010; Vygotsky, 1980).

One of the issues that makes it difficult to compare findings across studies is teleoperation. In the section on word learning, we discussed how teleoperation allows the experimenter to control the robot in real time, which may result in different interactions than when the robot is running autonomously. A teleoperated robot can respond in a contingent manner, while autonomous robots have to rely on predefined scripts and can respond contingently only when the participant behaves as expected. One of the values of RALL research is its contribution to developing autonomous robots that can be placed in classrooms. This does not mean, however, that robots should not be teleoperated during experiments, as teleoperating robots allows researchers to study more advanced interactions than those that can be achieved using autonomous robots. Such interactions are valuable to identify robot behaviors or properties that need to be developed further. However, we recommend that researchers clearly state whether they teleoperated their robot or used it autonomously to make the distinction between the two types of robots more clear and to facilitate comparisons between studies.

A further important issue follows from the newness of the field: Robots constitute a new form of technology, and too few studies have been conducted so far to conclude that robots are effective language tutors. Future studies will allow for firmer conclusions regarding robots' potential as language tutors. Furthermore, a subset of the previous studies is heavily underpowered and/or suffers from other methodological limitations (e.g., no control group), which warrants caution in interpreting and evaluating their results. These issues often come to the fore in studying the use of technology for language learning: New technologies are often met with great enthusiasm, but research on their effectiveness often does not meet empirical standards and/or does not necessarily provide conclusive evidence of the benefits of these technologies (Salaberry, 2001). However, even with their limitations, the studies reviewed in this article helped us to identify potential areas for future research.

In our Introduction, several advantages of robots over many other forms of technology were discussed. One advantage is that robots provide opportunities for the learner to interact with their real-life environment. They are physically present and make it possible to integrate physical exercises or objects into learning tasks. Thus far, motor activities with robots have rarely been incorporated in learning tasks due to their feasibility (e.g., walking is undesirable as robots are likely to fall over), with the exception of a few studies (e.g., Tanaka & Matsuzoe, 2012). This

is not surprising, because the more a robot acts and moves through space, the more likely it is that technical issues such as falling over or overheating will arise. However, it is possible that the use of objects and exercises could lead to higher learning gains (Kersten & Smith, 2002; Mavilidi et al., 2015; Toumpaniari et al., 2015). As robot technology is developing, motor activities are becoming more feasible to integrate, and this is therefore an area worth exploring.

Another advantage of robots over most other technologies is the interactional possibilities robots provide. Given the mixed evidence on the robot's social behavior, personalization of child-robot interaction is perhaps the most important line of research for the future. When humans teach a child new word forms and meanings, they carefully monitor the child's comprehension and, if necessary, adapt the tutoring strategy to the child's needs. Robots are not yet capable of monitoring the child's comprehension and adapting the tutoring strategy to individual children in such a careful manner. This makes it difficult to obtain "true" adaptivity, in which the robot adapts its lesson and behavior depending on the child's comprehension. This is partially due to speech recognition systems, which in their current form are often not suited to recognizing child speech (Kennedy et al., 2017), and to systems not being advanced enough to recognize children's emotions or comprehension. Furthermore, it is difficult to program robots to respond in a contingent manner, and even more so in interactions with young children, who are much more unpredictable in their verbal responses than older speakers. In other words, the research so far indicates that the important advantages of robots over most forms of technology still exist primarily in theory. Technical limitations prevent regular implementation of these possible advantages, and further technological developments are required to make full use of robots' potential and to put these theoretical advantages into practice.

A review of research on computer-assisted language learning (Garrett, 2009) mentioned how, in 1991, it was possible for one person to write "an overview . . . of the kinds of technological resources currently available to support language learning and of various approaches to making use of them" (Garrett, 1991, as quoted in Garrett, 2009, p. 719). In 2009, the same author noted that an update would fill an entire journal, requiring contributions from many different areas of expertise (Garrett, 2009). Perhaps in another two decades, we will say the same about RALL. We will go from one review aimed at capturing almost all extant RALL research to a great many possibilities we cannot even imagine at the moment. The use of robots may become such an everyday aspect of life that we will not even wonder about employing them. However, before we reach that point, we first have to find out how exactly robots should interact and behave socially to be effective language tutors.

Appendix

TABLE A1

Robots used in the various studies and their characteristics

Robot	Research or company based	Developed by	Target age-group	Purpose	Body	Face	Arms
Bioloid	Company	Robotis	All ages	Build your own robot	A.O. humanoid, spider, or dinosaur	Fixed	Yes, movable by robot
Dragonbot	Research	MIT Media Lab	Children	Long-term interactions with children	Shaped like a dragon	Animated on a screen	Yes, to be posed by others
EngKey	Company-made for research purposes	Korea Institute of Science and Technology	Children	Real-time interaction between students and their remote teachers	Semihumanoid	Avatar on a screen	Yes, movable by robot
iCat	Company-made for research purposes	Philips Electronics	All ages	Research on human-robot collaboration, joint attention, gaming, and ambient intelligence	Shaped like a cat	Movable eyes, eyebrows, lips	No
iRObi(Q)	Company	Yujin Robot	Children	Early childhood education	Semihumanoid	Fixed, with emotion expression through colors	Yes, movable by robot
MecWilly	Company	MecWilly Project	All ages	Education	Humanoid	Movable eyes, eyebrows, lips	Yes, movable by robot
Mero	Company-made for research purposes	Korea Institute of Science and Technology	All ages	Various types of interactions	No	Movable eyes, eyebrows, lips	No
Nabaztag	Company	Violet/Mindscape/ Softbank Robotics	All	Smart object	Shaped like a rabbit	Fixed, with movable ears	No
NAO	Company	Softbank Robotics	All ages	Various types of interactions	Humanoid	Fixed	Yes, movable by robot
Robo Sapien	Company	WowWeeBioloidRoo	All ages	Making programming robots easy	Humanoid	Fixed	Yes, movable by robot
ROBOSEM	Company	Yujin Robot	All ages	L2 English teaching	Humanoid	An avatar on a screen	Yes, movable by robot
Robovie (R3)	Company-made for research purposes	Vstone & ATR	All ages	Assistance in everyday tasks	Humanoid	Fixed	Yes, movable by robot
Tangible robot companions	Research	Wang, Young, and Jang (2013)	Children	L2 English teaching	Animals	Fixed	Yes, movable by robot

Notes

This research was supported by the European Union's Horizon 2020 research and innovation program under grant agreement no. 688014. We would like to thank Robyn Berghoff for her feedback on an earlier version of this manuscript and James Kennedy for sharing his thoughts on directions for future research in the field of RALL.

[1] Note that in the field of robotics, many results are presented at conferences rather than in peer-reviewed journals to allow for more rapid development of the field.

[2] These data are supported by another study by the same authors, which also showed that children can learn to use words productively through playing storytelling games with a robot (Kory Westlund & Breazeal, 2015). However, only preliminary data from this study have been published so far.

[3] A pilot study suggests that the learning-by-teaching paradigm could also be effective in improving children's writing skills by having them teach a robot how to write (Hood et al., 2015). However, this finding needs to be investigated further.

[4] A pilot study with four autistic children and a design similar to that in Alemi et al. (2014) suggests that RALL classes are also effective for autistic children (Alemi, Meghdari, Mahboub Basiri, & Taheri, 2015), although further research is required on this subject.

References

References marked with an asterisk indicate studies included in the review.

*Alemi, M., Meghdari, A., & Ghazisaedy, M. (2014). Employing humanoid robots for teaching English language in Iranian junior high-schools. *International Journal of Humanoid Robotics, 11*, 1450022-1–1450022-25. doi:10.1142/S0219843614500224

*Alemi, M., Meghdari, A., & Ghazisaedy, M. (2015). The impact of social robotics on L2 learners' anxiety and attitude in English vocabulary acquisition. *International Journal of Social Robotics*, *7*, 523–535. doi:10.1007/s12369-015-0286-y

Alemi, M., Meghdari, A., Mahboub Basiri, N., & Taheri, A. (2015). The effect of applying humanoid robots as teacher assistants to help Iranian autistic pupils learn English as a foreign language. In *Proceedings of the International Conference on Social Robotics 2015* (pp. 1–10). Basel, Switzerland: Springer International. doi:10.1007/978-3-319-25554-5

*Alemi, M., Meghdari, A., & Sadat Haeri, N. (2017). Young EFL learners' attitude towards RALL: An observational study focusing on motivation, anxiety, and interaction. In *Proceedings of the International Conference on Social Robotics 2017* (pp. 252–261). Basel, Switzerland: Springer International. doi:10.1007/978-3-319-70022-9

Arksey, H., & O'Malley, L. (2005). Scoping studies: Towards a methodological framework. *International Journal of Social Research Methodology*, *8*, 19–32.

Arsénio, A. M. (2014). Developmental language learning from human/humanoid robot social interactions: An embodied and situated approach. In *Robotics: Concepts, methodologies, tools, and applications* (pp. 1328–1353). New York, NY: IGI Global.

Barrett, N. E., & Liu, G.-Z. (2016). Global trends and research aims for English academic oral presentations: Changes, challenges, and opportunities for learning technology. *Review of Educational Research*, *86*, 1227–1271. doi:10.3102/0034654316628296

Barsalou, L. W. (2008). Grounded cognition. *Annual Review of Psychology*, *59*, 617–645. doi:10.1146/annurev.psych.59.103006.093639

Bartneck, C., & Forlizzi, J. (2004). A design-centred framework for social human-robot interaction. In *Proceedings of the Ro-Man 2004* (pp. 591–594). Los Alamitos, CA: IEEE. doi:10.1109/ROMAN.2004.1374827

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*, 71–81.

Belpaeme, T., Kennedy, J., Baxter, P., Vogt, P., Krahmer, E. E. J., Kopp, S., . . .Deblieck, T. (2015, October). *L2TOR: Second language tutoring using social robots*. In Paper presented at the 1st International Workshop on Educational Robotics at the International Conference on Social Robotics, Paris, France.

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*, 539–550.

Chin, K., Hong, Z.-W., & Chen, Y.-L. (2014). Impact of using an educational robot-based learning system on students' motivation in elementary education. *IEEE Transactions on Learning Technologies*, *7*, 335–345.

Cronin, P., Ryan, F., & Coughlan, M. (2008). Undertaking a literature review: A step-by-step approach. *British Journal of Nursing*, *17*, 38–43.

*de Haas, M., Baxter, P., de Jong, C., Krahmer, E., & Vogt, P. (2017). Exploring different types of feedback in preschooler and robot interaction. In *Proceedings of the Companion of the 2017 ACM/IEEE International Conference on Human-Robot Interaction* (pp. 127–128). New York, NY: ACM.

*de Wit, J., Schodde, T., Willemsen, B., Bergmann, K., de Haas, M., Kopp, S., . . . Vogt. (2018). 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* (pp. 50–58). New York, NY: ACM.

Dörnyei, Z. (1994). Motivation and motivating in the foreign language classroom. *Modern Language Journal*, *78*, 273–284.

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

*Eimler, S., von der Pütten, A., & Schächtle, U. (2010). Following the white rabbit—A robot rabbit as vocabulary trainer for beginners of English. In G. Leitner, M. Hitz, & A. Holzinger (Eds.), *Proceedings of the 6th Symposium of the Workgroup Human-Computer Interaction and Usability Engineering* (pp. 322–339). *Berlin, Germany: Springer*.

Falagas, M. E., Pitsouni, E. I., Malietzis, G. A., & Pappas, G. (2008). Comparison of PubMed, Scopus, Web of Science, and Google Scholar: Strengths and weaknesses. *The FASEB Journal*, *22*, 338–342. doi:10.1096/fj.07-9492LSF

Fernaeus, Y., Håkansson, M., Jacobsson, M., & Ljungblad, S. (2010). How do you play with a robotic toy animal? A long-term study of Pleo. In *Proceedings of the 9th international Conference on Interaction Design and Children* (pp. 39–48). New York, NY: ACM.

Funakoshi, K., Mizumoto, T., Nagata, R., & Nakano, M. (2011). The chanty bear: A new application for HRI research. In *Proceedings of the 6th International Conference on Human-Robot Interaction* (pp. 141–142). New York, NY: ACM.

Garrett, N. (2009). Computer-assisted language learning trends and issues revisited: Integration innovation. *The Modern Language Journal*, *93*, 719–740.

Golonka, E. M., Bowles, A. R., Frank, V. M., Richardson, D. L., & Freynik, S. F. (2014). Technologies for foreign language learning: A review of technology types and their effectiveness. *Computer Assisted Language Learning*, *27*, 70–105. doi:10 .1080/09588221.2012.700315

*Gordon, G., Breazeal, C., & Engel, S. (2015). Can children catch curiosity from a social robot? In *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction* (pp. 91–98). New York, NY: ACM.

*Gordon, G., Spaulding, S., Kory Westlund, J. M., Lee, J. J., Plummer, L., Martinez, M., . . . Breazeal, C. L. (2016). Affective personalization of a social robot tutor for children's second language skills. In *Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence* (pp. 3951–3957). New York, NY: ACM.

*Han, J., Jo, M.-H., Jones, V., & Jo, J.-H. (2008). Comparative study on the educational use of home robots for children. *Journal of Information Processing Systems*, *4*, 159–168. doi:10.3745/JIPS.2008.4.4.159

Han, J., Kang, B., Park, S., & Hong, S. (2012). How to sustain long-term interaction between children and ROBOSEM in English class. In *Proceedings of the seventh annual ACM/IEEE International Conference on Human-Robot Interaction* (pp. 421–422). New York, NY: ACM.

*Herberg, J. S., Feller, S., Yengin, I., & Saerbeck, M. (2015). Robot watchfulness hinders learning performance. In *Proceedings of the 24th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)* (pp. 153–160). Los Alamitos, CA: IEEE. doi:10.1109/ROMAN.2015.7333620

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

*Hong, Z.-W., Huang, Y.-M., Hsu, M., & Shen, W.-W. (2016). Authoring robot-assisted instructional materials for improving learning performance and motivation in EFL classrooms. *Educational Technology & Society*, *19*, 337–349.

Hood, D., Lemaignan, S., & Dillenbourg, P. (2015). When children teach a robot to write: An autonomous teachable humanoid which uses simulated handwriting. In *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction* (pp. 83–90). New York, NY: ACM.

*Hsiao, H.-S., Chang, C.-S., Lin, C.-Y., & Hsu, H.-L. (2015). "iRobiQ ": The influence of bidirectional interaction on kindergarteners' reading motivation, literacy, and behavior. *Interactive Learning Environments*, *23*, 269–292. doi:10.1080/10494820.2012.745435

*Hyun, E., Kim, S., Jang, S., & Park, S. (2008). Comparative study of effects of language instruction program using intelligence robot and multimedia on linguistic ability of young children. In *Proceedings of the 17th IEEE International Symposium on Robot and Human Interactive Communication* (pp. 187–192). Los Alamitos, CA: IEEE.

*In, J.-Y., & Han, J.-H. (2015). The acoustic-phonetic change of English learners in robot assisted learning. In *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction* (pp. 39–40). New York, NY: ACM. doi:10.1145/2701973.2702003

Iverson, J. M. (2010). Developing language in a developing body: The relationship between motor development and language development. *Journal of Child Language*, *37*, 229–261. doi:10.1017/S0305000909990432

*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*, 61–84. doi:10.1080/07370024.2004.9667340

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*, 146–151. doi:10.1111/cdep.12277

Kennedy, J., Baxter, P., Senft, E., & Belpaeme, T. (2015). Higher nonverbal immediacy leads to greater learning gains in child-robot tutoring interactions. In *Proceedings of the International Conference on Social Robotics*. Basel, Switzerland: Springer International. doi:10.1007/978-3-319-25554-5

*Kennedy, J., Baxter, P., Senft, E., & Belpaeme, T. (2016). Social robot tutoring for child second language learning. In *Proceedings of the Eleventh ACM/IEEE International Conference on Human Robot Interaction* (pp. 231–238). New York, NY: ACM.

Kennedy, J., Lemaignan, S., Montassier, C., Lavalade, P., Irfan, B., Papadopoulos, F., . . . Belpaeme, T. (2017). Child speech recognition in human-robot interaction: Evaluations and recommendations. In *Proceedings of the 2017 ACM/IEEE Human-Robot Interaction Conference* (pp. 82–90). New York, NY: ACM.

Kersten, A. W., & Smith, L. B. (2002). Attention to novel objects during verb learning. *Child Development*, *73*, 93–109.

Kidd, C. D., & Breazeal, C. (2004). Effect of a robot on user perceptions. In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems* (pp. 3559–3564). Los Alamitos, CA: IEEE.

Kory Westlund, J., & Breazeal, C. (2015). The interplay of robot language level with children's language learning during storytelling. In *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction* (pp. 65–66). New York, NY: ACM. doi:10.1145/2701973.2701989

*Kory Westlund, J. M., Dickens, L., Jeong, S., Harris, P. L., DeSteno, D., & Breazeal, C. L. (2017). Children use non-verbal cues to learn new words from robots as well as people. *International Journal of Child-Computer Interaction*, *13*, 1–9. doi:10.1016/j.ijcci.2017.04.001

*Kory Westlund, J., Dickens, L., Jeong, S., Harris, P., DeSteno, D., & Breazeal, C. (2015). A comparison of children learning new words from robots, tablets, & people. In *Conference Proceedings of New Friends: The 1st International Conference on Social Robots in Therapy and Education* (pp. 7–8).

*Kory Westlund, J. M., Jeong, S., Park, H. W., Ronfard, S., Adhikari, A., Harris, P. L., . . . Breazeal, C. L. (2017). Flat vs. expressive storytelling: Young children's learning and retention of a social robot's narrative. *Frontiers in Human Neuroscience*, *11*, 295. doi:10.3389/fnhum.2017.00295

Köse, H., Uluer, P., Akalın, N., Yorgancı, R., Özkul, A., & Ince, G. (2015). The effect of embodiment in sign language tutoring with assistive humanoid robots. *International Journal of Social Robotics*, *7*, 537–548. doi:10.1007/s12369-015-0311-1

Lee, E., & Lee, Y. (2008). A pilot study of intelligent robot aided education. In *Proceedings of the ICCE Conference on Advanced Learning Technologies, Open Contents, & Standards* (pp. 595–596). Retrieved from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.372.9307&rep=rep1&type=pdf

*Lee, S., Noh, H., Lee, J., Lee, K., Lee, G. G., Sagong, S., & Kim, M. (2011). On the effectiveness of robot-assisted language learning. *ReCALL*, *23*, 25–58. doi:10.1017/S0958344010000273

Leite, I., Martinho, C., & Paiva, A. (2013). Social robots for long-term interaction: A survey. *International Journal of Social Robotics*, *5*, 291–308. doi:10.1007/s12369-013-0178-y

Leyzberg, D., Spaulding, S., & Scassellati, B. (2014). Personalizing robot tutors to individuals' learning differences. In *Proceedings of the 2014 ACM/IEEE international*

conference on Human-Robot Interaction (pp. 423–430). New York, NY: ACM. doi:10.1145/2559636.2559671

Leyzberg, D., Spaulding, S., Toneva, M., & Scassellati, B. (2012). The physical presence of a robot tutor increases cognitive learning gains. In *Proceedings of the Annual Meeting of the Cognitive Science Society* (pp. 1882–1887). Retrieved from https://escholarship.org/uc/item/7ck0p200

Li, S. (2010). The effectiveness of corrective feedback in SLA: A meta-analysis. *Language Learning*, *60*, 309–365. doi:10.1111/j.1467-9922.2010.00561.x

Liu, E. Z. F., Lin, C. H., & Chang, C. S. (2010). Student satisfaction and self-efficacy in a cooperative robotics course. *Social Behavior and Personality*, *38*, 1135–1146.

Liu, S., Liao, H., & Pratt, J. A. (2009). Impact of media richness and flow on e-learning technology acceptance. *Computers & Education*, *52*, 599–607. doi:10.1016/j.compedu.2008.11.002

Mavilidi, M.-F., Okely, A. D., Chandler, P., Cliff, D. P., & Paas, F. (2015). Effects of integrated physical exercises and gestures on preschool children's foreign language vocabulary learning. *Educational Psychology Review*, *27*, 413–426. doi:10.1007/s10648-015-9337-z

*Mazzoni, E., & Benvenuti, M. (2015). A Robot-partner for preschool children learning English using socio-cognitive conflict. *Journal of Educational Technology & Society*, *18*, 474–485. doi:10.1145/2696454.2696457

*Meiirbekov, S., Balkibekov, K., Jalankuzov, Z., & Sandygulova, A. (2016). "You win, I lose": Towards adapting robot's teaching strategy. In *Proceedings of the Eleventh ACM/IEEE International Conference on Human Robot Interaction* (pp. 475–476). New York, NY: ACM.

Mitnik, R., Recabarren, M., Nussbaum, M., & Soto, A. (2009). Collaborative robotic instruction: A graph teaching experience. *Computers & Education*, *53*, 330–342. doi:10.1016/j.compedu.2009.02.010

Mondria, J.-A., & Wiersma, B. (2004). Receptive, productive, and receptive + productive L2 vocabulary learning: What difference does it make? In P. Bogaards & B. Laufer (Eds.), *Vocabulary in a second language* (pp. 79–100). Amsterdam, Netherlands: John Benjamins. doi:10.1017/S0272263106280057

Moriguchi, Y., Kanda, T., Ishiguro, H., Shimada, Y., & Itakura, S. (2011). Can young children learn words from a robot? *Interaction Studies*, *12*, 107–118.

*Movellan, J. R., Eckhardt, M., Virnes, M., & Rodriguez, A. (2009). Sociable robot improves toddler vocabulary skills. In *Proceedings of the 4th ACM/IEEE international conference on Human robot interaction* (pp. 307–308). New York, NY: ACM. doi:10.1145/1514095.1514189

Nagata, R., Mizumoto, T., Funakoshi, K., & Nakano, M. (2010). Toward a chanting robot for interactively teaching English to children. In *Proceedings of INTERSPEECH Satellite Workshop on Second Language Studies: Acquisition, Learning, Education and Technology* (pp. 2–13). Retrieved from https://pdfs.semanticscholar.org/c013/d7207a546e0db9f124534ff3eb8cb6033bd0.pdf?_ga=2.47043492.2058360171.1544163163-1804566684.1541741064

Nourbakhsh, I. R., Crowley, K., Bhave, A., Hamner, E., Hsiu, T., Perez-Bergquist, A., . . . Wilkinson, K. (2005). The robotic autonomy mobile robotics course: Robot design, curriculum. *Autonomous Robots*, *18*, 103–127.

Pekrun, R., Goetz, T., Titz, W., & Perry, R. P. (2002). Academic emotions in students' self-regulated learning and achievement: A program of qualitative and quantitative research. *Educational Psychologist*, *37*, 91–105.

*Rintjema, E., van den Berghe, R., Kessels, A., de Wit, J., & Vogt, P. (2018). A robot teaching young children a second language: The effect of multiple interactions on engagement and performance. In *Companion of the 2018 ACM/IEEE International Conference on Human-Robot Interaction* (pp. 219–220). New York, NY: ACM. doi:10.1145/3173386.3177059

Rohrbeck, C. A., Ginsburg-Block, M. D., Fantuzzo, J. W., & Miller, T. R. (2003). Peer-assisted learning interventions with elementary school students: A meta-analytic review. *Journal of Educational Psychology*, *95*, 240–257. doi:10.1037/0022-0663.95.2.240

*Rosenthal-von der Pütten, A. M., Straßmann, C., & Krämer, N. C. (2016). Robots or agents—Neither helps you more or less during second language acquisition. Experimental study on the effects of embodiment and type of speech output on evaluation and alignment. In *Proceedings of the International Conference on Intelligent Virtual Agents* (pp. 256–268). New York, NY: Springer. doi:10.1007/978-3-319-47665-0_23

Rowe, M. L., & Goldin-Meadow, S. (2009). Differences in early gesture explain SES disparities in child vocabulary size at school entry. *Science*, *323*, 951–954.

*Saerbeck, M., Schut, T., Bartneck, C., & Janse, M. D. (2010). Expressive robots in education: Varying the degree of social supportive behavior of a robotic tutor. In *Proceedings of the 28th International Conference on Human Factors in Computing Systems* (pp. 1613–1622). New York, NY: ACM. doi:10.1145/1753326.1753567

Salaberry, M. R. (2001). The use of technology for second language learning and teaching: A retrospective. *Modern Language Journal*, *85*, 39–56.

*Schodde, T., Bergmann, K., & Kopp, S. (2017). Adaptive robot language tutoring based on Bayesian knowledge tracing and predictive decision-making. In *Proceedings of the 12th ACM/IEEE International Conference on Human-Robot Interaction* (pp. 128–136). *New York, NY: ACM*.

*Shin, J., & Shin, D.-H. (2015). Robot as a facilitator in language conversation class. In *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction Extended Abstracts* (pp. 11–12). New York, NY: ACM. doi:10.1145/2701973.2702062

Shultz, M. (2007). Comparing test searches in PubMed and Google Scholar. *Journal of the Medical Library Association*, *95*, 442–453. doi:10.3163/1536-5050.95.4.442

Stockwell, G. (2008). Investigating learner preparedness for and usage patterns of mobile learning. *ReCALL*, *20*, 253–270.

Takacs, Z. K., Swart, E. K., & Bus, A. G. (2015). Benefits and pitfalls of multimedia and interactive features in technology-enhanced storybooks: A meta-analysis. *Review of Educational Research*, *85*, 698–739. doi:10.3102/0034654314566989

Tanaka, F., Cicourel, A., & Movellan, J. R. (2007). Socialization between toddlers and robots at an early childhood education center. In *Proceedings of the National Academy of Sciences*, *104*, 17954–17958.

Tanaka, F., & Ghosh, M. (2011). The implementation of care-receiving robot at an English learning school for children. In *Proceedings of the 6th International Conference on Human-Robot Interaction* (pp. 265–266). New York, NY: ACM. doi:10.1145/1957656.1957763

*Tanaka, F., & Matsuzoe, S. (2012). Children teach a care-receiving robot to promote their learning: Field experiments in a classroom for vocabulary learning. *Journal of Human-Robot Interaction*, *1*, 78–95. doi:10.5898/JHRI.1.1.Tanaka

Toumpaniari, K., Loyens, S., Mavilidi, M.-F., & Paas, F. (2015). Preschool children's foreign language vocabulary learning by embodying words through physical activity and gesturing. *Educational Psychology Review*, *27*, 445–456. doi:10.1007/s10648-015-9316-4

*Uluer, P., Akalın, N., & Köse, H. (2015). A new robotic platform for sign language tutoring: Humanoid robots as assistive game companions for teaching sign language. *International Journal of Social Robotics*, *7*, 571–585. doi:10.1007/s12369-015-0307-x

*van den Berghe, R., van der Ven, S., Verhagen, J., Oudgenoeg-Paz, O., Papadopoulos, F., & Leseman, P. (2018). Investigating the effects of a robot peer on L2 word learning. In *Companion of the 2018 ACM/IEEE International Conference on Human-Robot Interaction* (pp. 267–268). New York, NY: ACM. doi:10.1145/3173386.3176986

Vygotsky, L. S. (1980). 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.

Wainer, J., Feil-Seifer, D. J., Shell, D. A., & Matari, M. J. (2007). Embodiment and human-robot interaction: A task-based perspective. In *Proceedings of the International Conference on Human-Robot Interaction* (pp. 872–877). New York, NY: ACM.

*Wang, Y. H., Young, S. S.-C., & Jang, J.-S. R. (2013). Using tangible companions for enhancing learning English conversation. *Journal of Educational Technology & Society*, *16*, 296–309.

Wellsby, M., & Pexman, P. M. (2014). Developing embodied cognition: Insights from children's concepts and language processing. *Frontiers in Psychology*, *5*, 506. doi:10.3389/fpsyg.2014.00506

*You, Z.-J., Shen, C.-Y., Chang, C.-W., Liu, B.-J., & Chen, G.-D. (2006). A robot as a teaching assistant in an English class. In *Proceedings of the Sixth International Conference on Advanced Learning Technologies* (pp. 87–91). Los Alamitos, CA: IEEE. doi:10.1109/ICALT.2006.1652373

Young, M. F., Slota, S., Cutter, A. B., Jalette, G., Mullin, G., Lai, B., . . . Yukhymenko, M. (2012). Our princess is in another castle: A review of trends in serious gaming for education. *Review of Educational Research*, *82*, 61–89.

Authors

RIANNE VAN DEN BERGHE is a PhD candidate at the Department of Special Education: Cognitive and Motor Disabilities, Utrecht University, Heidelberglaan 1, Utrecht, 3584 CS, the Netherlands; email: *m.a.j.vandenberghe@uu.nl*. She received an MA in linguistics from Utrecht University, in which she focused on (second) language acquisition and discourse processing. In her PhD research, she investigates the effectiveness of social robots for children's second-language education.

JOSJE VERHAGEN is a postdoctoral researcher at the Department of Dutch Studies at the University of Amsterdam, the Netherlands, and at the Department of Special Education: Cognitive and Motor Disabilities, Utrecht University, Heidelberglaan 1, Utrecht, 3584 CS, the Netherlands; email: *j.verhagen@uu.nl*. She received her PhD in 2009 on adult second language acquisition from the Max Planck in Psycholinguistics (Nijmegen) and

VU University (Amsterdam). Her research interests are first and second language acquisition, as well as bilingualism in young children.

ORA OUDGENOEG-PAZ is a postdoctoral researcher at the Department of Special Education: Cognitive and Motor Disabilities, Utrecht University, Heidelberglaan 1, Utrecht, 3584 CS, the Netherlands; email: *o.oudgenoeg@uu.nl*. She received her PhD in pedagogics from Utrecht University. Her research focuses on early language and motor development and the link between the two. She studies how motor development, sensorimotor interactions, and early language exposure facilitate the development of (spatial) cognition and (spatial) language. Her work concerns both first and second language acquisition.

SANNE VAN DER VEN is an assistant professor at the Department of Special Education: Cognitive and Motor Disabilities, Utrecht University, Heidelberglaan 1, Utrecht, 3584 CS, the Netherlands; email: *s.vanderven@uu.nl*. She received her PhD in 2011 on the development of executive functions and mathematical abilities in primary school children. Her current research focuses on the effects of feedback and emotion on learning.

PAUL LESEMAN is a full professor of education at Utrecht University, Heidelberglaan 1, Utrecht, 3584 CS, the Netherlands; email: *p.p.m.leseman@uu.nl*. He received an MA in psychology at the University of Amsterdam and a PhD at the Erasmus University Rotterdam. His research concerns early development and education of children in multicultural contexts. He has published on language development, bilingual development, emergent literacy and numeracy, and executive functions.