International Journal of Social Robotics (2022) 14:297–312
https://doi.org/10.1007/s12369-021-00789-3

Are Tutor Robots for Everyone? The Influence of Attitudes, Anxiety, and Personality on Robot-Led Language Learning

Junko Kanero[1,2] · Cansu Oranç[2,3] · Sümeyye Koşkulu[2,4] · G. Tarcan Kumkale[5] · Tilbe Göksun[2] · Aylin C. Küntay[2]

Accepted: 17 April 2021 / Published online: 6 May 2021

Abstract

Do some individuals benefit more from social robots than others? Using a second language (L2) vocabulary lesson as an example, this study examined how individual differences in attitudes toward robots, anxiety in learning L2, and personality traits may be related to the learning outcomes. One hundred and two native Turkish-speaking adults were taught eight English words in a one-on-one lesson either with the NAO robot ($N=51$) or with a human tutor ($N=51$). The results in both production and receptive language tests indicated that, following the same protocol, the two tutors are fairly comparable in teaching L2 vocabulary. Negative attitudes toward robots and anxiety in L2 learning impeded participants from learning vocabulary in the robot tutor condition whereas the personality trait of extroversion negatively predicted vocabulary learning in the human tutor condition. This study is among the first to demonstrate how individual differences can affect learning outcomes in robot-led sessions and how general attitudes toward a type of device may affect the ways humans learn using the device.

Keywords Human–robot interaction (HRI) · Second language learning · Attitudes · Anxiety · Personality

1 Introduction

Are robots for everyone? With rapid advancements in robotics technology, it is only natural to consider different ways to integrate robots into our lives. Robots are already being implemented in a variety of environments including homes, factories, hospitals, and schools. In some environments, evaluating the benefits of robots may be relatively simple and easy. For example, if a robot is involved in the automation process of a manufacturing line, its benefits can be assessed based on whether the speed, quality, and ease of production increased after the installation of the robot. In the case of *social robots*, however, the assessment is not as straightforward. It can be challenging to determine and to fine-tune policies and practices regarding their use. The present study takes second language (L2) education as an example to evaluate the benefits of social robots. Learning an L2 is deemed particularly challenging because the age of acquisition is often late and exposure is sporadic [1]. Therefore, L2 learning provides the perfect opportunity to evaluate the potential of social robots as a learning aid and to explore how individual differences among learners can influence the effectiveness of lessons led by a social robot. Our main question here is whether there are individual differences in how L2 learners benefit from social robot tutors, and if so, what the characteristics of these successful learners are.

A social robot is "an autonomous or semi-autonomous robot that interacts and communicates with humans by following the behavioral norms expected by the people with whom the robot is intended to interact" ([2], p. 592). In educational settings, social robots are theorized to make unique contributions to learning especially when it is critical for the learner to interact with another agent. Interactive and responsive contexts facilitate language learning not only in human–human interaction but also when the learner

✉ Junko Kanero
jkanero@sabanciuniv.edu

1 Faculty of Arts and Social Sciences, Sabancı University, Orta Mahalle, Üniversite Caddesi No:27, Tuzla, 34956 Istanbul, Turkey

2 Department of Psychology, Koç University, Istanbul, Turkey

3 Max Planck Institute for Human Development, MPRG iSearch, Berlin, Germany

4 Department of Education and Pedagogy, Utrecht University, Utrecht, Netherlands

5 Department of Psychology, Özyeğin University, Istanbul, Turkey

uses digital devices [3]. As a social agent with a physical body, a social robot can play the role of a human tutor using vocal, gestural, and facial expressions [4–7]. Robots are also adaptive: robots can flexibly use their sensors to detect the motivational and educational needs of learners and change their behaviors accordingly [6, 7]. For example, the robot used in the current study NAO is equipped with not only a camera and a microphone but also touch sensors. It can be difficult for classroom teachers to meet the needs of each individual student in the class (e.g., [8]). In such situations, robot tutors may serve as a supplementary tool to provide additional one-on-one or small group lessons for struggling learners although, as we discuss in this paper, their benefits may vary across learners [9, 10]. These traits distinguish social robots from more commonly spread digital tools, such as smartphones and tablets, and call for the evaluation of their unique contributions in educational settings.

1.1 Social Robots in Language Education

Though a relatively new device, the use of robots in education has been explored extensively in both research and practical settings (e.g., schools, hospitals). For example, a meta-analysis study found 101 published papers (reporting 309 study results) about the use of robots in education prior to the time of their survey in May 2017 [7]. Importantly, most studies focused on whether robots boost the motivation of the learner and have not properly evaluated *learning outcomes* [6]. It is certainly reassuring that robots can motivate and engage learners. Yet, we also need to investigate the efficacy of robots as a learning aid since there is no need to introduce the relatively expensive technology to educational settings if it does not facilitate learning. The current study aims to directly assess the impact of social robots in language education and examines whether people can learn from a robot tutor, and if so, under what circumstances robots are actually beneficial to learners. In exploring the effectiveness of social robots as a tutor, we aim to identify individual differences among learners that affect the ways they learn from a robot tutor.

1.2 Individual Differences in Human–Robot Interaction

Although the importance of individual differences has been discussed for centuries, only in the past few decades, the exploration of individual differences became truly apparent across a broad range of psychological topics (see [11]). Whereas traditional psychological research focuses mainly on how humans think and behave on average, many new studies emphasize the need for examining individuals because humans employ vastly different approaches to the same cognitive task, including language learning (see [12] for a review).

Evaluating the relation between technology in general and individual differences in users, studies have focused almost exclusively on what leads the user to use technology. The most influential framework may be the Technology Acceptance Model (TAM) and its derivatives [13, 14]. These models suggest that perceived usefulness and perceived ease-of-use of the technology affect the users' attitudes toward the technology, which in turn predict whether or not the user actually uses the technology. Testing personality in a similar context, a study found that the use of technology was correlated positively with conscientiousness and negatively with neuroticism and extroversion [15].

Although not much is known specifically about the effects of individual differences in learning with robots, some studies have explored how attitudes and personality are related to the ways in which a person interacts with a robot. For example, the patterns of speech and eye gaze were observed in 56 adults while they built an object with the humanoid robot iCub [16]. The study found that individuals with negative attitudes toward robots tended to look less at the robot's face and more at the robot's hands. Furthermore, extroversion was found to be related to how much one talked with the robot. Interestingly, participants who were high on introversion interacted more with an introverted robot than an extroverted robot whereas participants who were high on extroversion interacted more with an extroverted robot than an introverted robot [17]. Other studies also reported that extroverts tend to be comfortable with robots physically approaching [18] and felt psychologically closer to robots [19]. Neuroticism also made a difference: individuals who are high in neuroticism did not feel psychologically close to the robot, and did not like the robot as much as those low in neuroticism [19]. In another study, when robots approached people, having high levels in the personality trait of neuroticism and negative attitudes toward robots increased personal space between the robot and the self [20] (see [21] for a review). Personality has also been suggested to affect our perception and acceptance of robots [22].

These studies suggest that individual differences in attitudes toward robots and certain personality traits are related to how humans behave when they interact with a robot. However, the results are far from consistent, and more importantly, no study has examined whether individuals with different attitudes toward robots and with different personality traits learn differentially from social robots. Observing differences in human behaviors has a scientific impact, but perhaps more important for human–robot interaction (HRI) is to move a step further and evaluate whether individuals with certain traits benefit, or fail to benefit, more from robot companions than others. Robot-led L2 learning is an excellent context to explore the issue because learning outcomes such as language test scores can be directly used to evaluate how effective a robot companion is.

No previous research focused specifically on the role of individual differences in robot-led L2 learning, but the idea has been suggested [23, 24]. In the examination of English word learning in fifth and sixth graders in Japan, children with some English proficiency or interest in English benefited more from extra learning opportunities provided by social robots than did their peers with lower proficiency or interest [23]. Some researchers also argue that robots may be particularly helpful for individuals with impaired social and communicative skills such as children with autism spectrum disorder (ASD). Social interactions with humans can be difficult or stressful for children with ASD because humans behave in very complex and unpredictable manners. Robots can be good communication partners for those children as they can provide simpler and less stressful environments [24]. Based on these theoretical arguments and the previous studies reporting that negative attitudes toward robots affect how a person interacts with a robot [16, 20, 25], we asked whether individuals who have more negative attitudes toward robots are less likely to learn words in the robot-led lesson compared to their peers with more positive attitudes.

1.3 Individual Differences in L2 Learning

In L2 learning, recognizing individual differences may be particularly important. For example, some individuals prefer to be immersed in L2 and enjoy discovering ways to use words and phrases whereas others may prefer and learn better when instructions are given in L1 and they learn correspondences between L2 expressions and L1 translations of the expressions (e.g., [26, 27]). Meeting the preferences and needs of each individual learner has the potential to improve the outcome of L2 learning.

In their pioneering work, Horwitz, Horwitz, and Cope (1986) developed the Foreign Language Classroom Anxiety Scale (FLCAS) that assesses different types of anxiety that are related to language learning such as communication apprehension, test anxiety, and fear of negative evaluation [28]. Research using the scale generally finds that L2 anxiety and L2 achievement or performance negatively correlate with each other [29]. It is worth exploring whether a similar relation can be found when the language lesson is delivered by a robot. Using the FLCAS, a study with high school ESL students in Iran reported that students who took a group English lesson with a robot teaching assistant had less anxiety toward L2 learning than their peers who went through the same lesson content without a robot [30]. However, the robot and no-robot conditions differed in multiple ways (e.g., the robot made mistakes), and neither individual differences in L2 anxiety prior to the lesson nor learning outcomes were assessed. In contrast, the current study used the FLCAS to measure the predispositions of learners and tested whether anxiety about learning L2 impedes learning from a robot.

Personality traits have also been studied in relation to L2 learning. In particular, the personality trait of extroversion/introversion has been receiving the most attention [31, 32]. Despite the large size of literature, the evidence is far from conclusive [32, 33]. Whereas some studies identified a positive correlation between extroversion and learning outcomes [31], other studies found no relation [34]. Others, however, reported that introverted individuals were better L2 learners than extroverted individuals [35, 36]. The nature of outcome measures has often been identified as a possible cause of the discrepancy—perhaps extroverted learners are better at learning basic interpersonal communication skills whereas introverted learners are better at academic language abilities [33, 37]. More recently, a study with 115 high school students in Poland also concluded that extroversion might be beneficial only when oral communication skills are measured as the outcome [31]. In this study, no relation was found between extroversion/introversion and the English course grade in one school year, but extroversion predicted a higher grade in the following school year when the teacher focused more on oral skills. Several possibilities including the nature of the task (written language vs. spoken language) [31, 33, 37, 38], situation (e.g., introverts perform better when they study in a familiar environment whereas extroverts perform better in an unfamiliar study environment) [39], and interaction between personality and learning styles [40], have been proposed. Further research in general is needed regarding the association between personality traits and L2 learning, and the current study contributes to the discussion by evaluating whether learners' personality traits such as extroversion predict learning outcomes of robot-led L2 lessons.

1.4 Present Study

To examine whether and how individual differences are related to adults' L2 learning from social robot tutors, the present study tested English vocabulary learning by native Turkish speakers. To assess the unique nature of robot-led L2 lessons, it is critical to specifically examine individuals' attitudes toward robots. By assessing both negative attitudes toward robots and general personality traits, we are able to understand whether the observed relations between individual differences and learning outcomes are likely to be specific to robot-led L2 lessons as opposed to L2 lessons in general. For example, extroverted individuals may benefit from language lessons, whether with another person or a robot, because they enjoy communicating with another agent. In addition to the general personality traits, we examined whether anxiety specific to L2 learning affects learning outcomes. The personality trait of neuroticism is known to be associated with anxiety in general [41, 42], but assessing personality may not capture the anxiety learners feel specifically toward learning language. Finally, to identify the unique characteristics of robot-led L2 lessons and to understand whether the robot L2 tutor can be considered effective and beneficial, we also tested a human-led lesson in which a human tutor taught the same set of L2 words following the same teaching protocol.

In summary, this study examined whether (1) the robot and human tutors differ in their effectiveness in a vocabulary lesson when the two are following the same teaching protocol, (2) individuals who have more negative attitudes toward robots are less likely to learn words from the robot than those with more positive attitudes, (3) anxiety about learning L2 is negatively correlated with the number of words individuals learn in robot-led lessons, and (4) general personality traits such as extroversion predict the learning outcomes of robot-led L2 lessons.

2 Methods

2.1 Participants

The dataset consisted of 102 native Turkish-speaking young adults: 51 in the robot tutor condition (*age range* = 18–26 years; M_{age} = 19.99 years; *SD* = 1.84; 34 females), and 51 in the human tutor condition (*age range* = 18–24 years; M_{age} = 19.84 years; *SD* = 1.21; 34 females). All participants were undergraduate and graduate students at a university in Istanbul, Turkey. The academic language of the university is English, and all participants had intermediate to advanced English skills. Participants had no known vision or hearing impairments. They were given the option of receiving monetary compensation or course credits for their participation.[1] One participant in the robot tutor condition and two participants in the human tutor condition did not show up for the second visit and thus their scores for the two delayed language tests are missing. In addition, one participant in the robot tutor condition was not taught one of the eight vocabulary words due to a technical error, and thus the test data for that word were not used. One data point for the delayed production test for a participant in the human condition was also voided because the experimenter accidentally revealed the answer.

[1] We offered two forms of compensation to increase the diversity of our subject pool. A GLMM testing the form of compensation (money vs. course credit) as the sole fixed factor and word as a random intercept suggests that the form of compensation did not affect the scores of any of the outcome measures ($p = 0.828$ for the immediate production test; $p = 0.900$ for the immediate receptive test; $p = 0.927$ for the delayed production test; and $p = 0.275$ for the delayed receptive test).

2.2 Materials and Procedures

The overview of the tasks and measures used in the present study was as follows: In the first visit, participants completed a computerized questionnaire assessing attitudes toward robots (NARS), L2 anxiety, and personality traits. Then, based on a random assignment, participants received the one-on-one English lesson either from the robot or the human tutor. Immediately after the lesson, participants in both conditions completed measures of learning (i.e., production and receptive vocabulary tests). Finally, participants visited the lab again after a week to complete the delayed measures of learning, intended to assess retention of learning from both types of tutors.

2.2.1 Negative Attitudes Toward Robots

Negative Attitudes toward Robots Scale (NARS; [43]) was used to assess attitudes toward robots. The NARS consists of 14 questions divided into three subordinate scales: negative attitude toward interacting with robots (S1), negative attitude toward the social influence of robots (S2), and negative attitude toward emotions involved in the interaction with robots (S3). The Turkish version of the NARS was developed by the first and second authors based on both the original Japanese version and the English version [43] (see Appendix 1). Participants rated how well each of the statements represented their attitudes toward robots on a scale of 1–5 (1: I strongly disagree/ Kesinlikle katılmıyorum, 2: I disagree/Katılmıyorum, 3: Undecided/Kararsızım, 4: I agree/Katılıyorum, 5: I strongly agree/Kesinlikle katılıyorum). The Cronbach's alpha for the items used in the study was 0.83 (see 3.2. *Individual differences and learning outcomes* for more details).

2.2.2 L2 Anxiety

The Foreign Language Classroom Anxiety Scale (FLCAS; [28]) consists of 33 statements (e.g., I never feel quite sure of myself when I am speaking in my foreign language class/Yabancı dil derslerinde konuşurken kendimden asla emin olamıyorum.) to be rated on a scale of 1–5 (1: I fully disagree/Hiç katılmıyorum, 2: I disagree/Katılmıyorum, 3: I neither agree nor disagree/Ne katılıyorum ne de katılmıyorum, 4: I agree/Katılıyorum, 5: I fully agree/Tamamen katılıyorum). The Turkish version of the scale translated by Aydın and colleagues was used in the study [44]. The Cronbach's alpha was 0.93.

2.2.3 Personality Traits

Based on the five-factor model of personality, we assessed the following five traits using the Turkish version of a

Table 1 The target words and their definitions used in the study

Word	Definition
Dromedary	*Bu kelime tek hörgüçlü deve anlamına gelir* (This word means a one-humped camel)
Derrick	*Bu kelime petrol kuyusu üzerindeki kule anlamına gelir* (This word means a tower over an oil well)
Cairn	*Bu kelime taş yığını anlamına gelir* (This word means a mound of stones)
Angler	*Bu kelime olta ile balık tutan kimse anlamına gelir* (This word means a person who fishes with hook and line)
Caster	*Bu kelime bir şeye takılan küçük tekerlek anlamına gelir* (This word means a little wheel attached to something)
Cupola	*Bu kelime bir çatı üstüne inşa edilen küçük kubbe benzeri yapı anlamına gelir* (This word means a rounded vault-like structure built on top of a roof)
Barb	*Bu kelime çengel ya da kanca anlamına gelir* (This word means the tip of an arrow or fishhook)
Upholstery	*Bu kelime döşemelik kumaş anlamına gelir* (This word means fabric used to make a soft covering)

previously validated personality inventory—openness to experience, conscientiousness, extroversion, agreeableness, and neuroticism [45]. This inventory included 44 questions addressing each of the five traits—7 items for conscientiousness (e.g., I stick to my plans/Yaptığım planlara sadık kalırım); 10 items for neuroticism (e.g., I am depressed/Depresifimdir); 9 items for each of openness to experience (e.g., My interests are very diverse/İlgi alanlarım çok çeşitlidir), extroversion (e.g., I am talkative/Konuşkanımdır), and agreeableness (e.g., I am helpful/Yardımseverimdir). Participants rated how well each of the statements represented their personality on a scale of 1–5 (1: I strongly disagree/Kesinlikle katılmıyorum, 2: I disagree/Katılmıyorum, 3: I neither agree nor disagree/Ne katılıyorum, ne de katılmıyorum, 4: I agree/Katılıyorum, 5: I strongly agree/Kesinlikle katılıyorum). The Cronbach's alphas were 0.83, 0.81, 0.88, 0.66, and 0.84 for openness to experience, conscientiousness, extroversion, agreeableness, and neuroticism, respectively.

2.2.4 Post-Lesson Vocabulary Tests

The production vocabulary test (hereafter the production test) and receptive vocabulary test (hereafter the receptive test) were administered immediately after the lesson (immediate post-lesson tests). To assess to what extent vocabulary was retained over time, participants completed the same measures again after a delay of one week (delayed post-lesson tests). The definitions of the target words used in the production test were the same as the definitions used in the lesson. In the receptive test, the pictures from the Peabody Picture Vocabulary Test, Fourth Edition (PPVT-4) [46], which correspond to the target words were used. The following section presents the details of the procedure. The delayed post-lesson tests were obtained in the lab seven days after the lesson. Due to schedule conflicts, however, three participants completed these tests after six days, while another participant completed the tests after eight days.

2.2.5 English Lesson with the Robot or Human Tutor

Participants were taught eight English nouns in the following order—dromedary, derrick, cairn, angler, caster, cupola, barb, and upholstery (see Table 1 for the definitions of the words). The words were selected from the last 40 items of the PPVT-4, as those items were supposed to be advanced even for native English speakers. The eight words were carefully selected so that (1) the Turkish equivalents of the words were not phonetically similar to them, and (2) pronouncing the words should not be too difficult for Turkish speakers.

The robot tutor was controlled through a Wizard-of-Oz interface [47]. We set one microphone behind the participant and four cameras at the corners of the ceiling, with which the "wizard" in another room monitored the participant. With regard to the voice of the robot, we used the female voice available on Amazon Polly ("Filiz" for Turkish and "Salli" for American English). All speech was prerecorded as WAV sound files.[2] The robot tutor provided no facial expressions, but moved its head and arms during the lesson to keep the participant engaged. While pronouncing the target English words and definitions, the robot stood still without

[2] We did not use the default text-to-speech (TTS) library in NAO because native Turkish speakers in the research team (co-authors and research assistants) found the Turkish speech to be unnatural and difficult to comprehend. The Amazon Polly "Filiz" was the most natural Turkish option we found, and "Salli" was chosen for English speech as it sounded most similar to "Filiz" among available options. We also modified the input text when the generated speech was unnatural or difficult to comprehend.

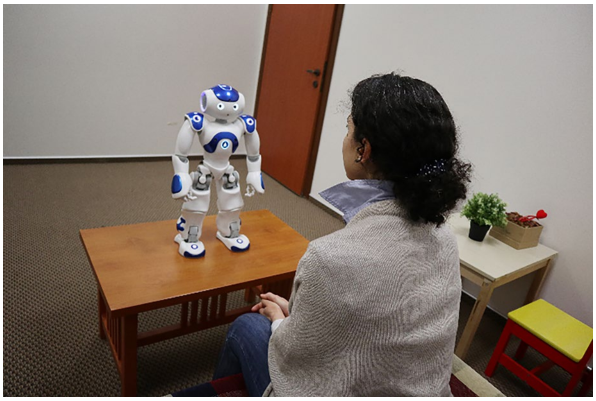

Fig. 1 The participant was instructed to go into a living room-like room by herself and to sit in front of the tutor

any movements because the motor sound of the robot could hinder the hearing. Most actions were chosen from the Animated Speech library of SoftBank Robotics (http://doc.aldebaran.com/2-1/naoqi/audio/alanimatedspeech_advanced.html), although some were created by the first author. There were unavoidable behavioral differences between the two tutors (e.g., the motor sound of the robot), but otherwise, the behavioral differences between the two tutors were kept minimal so that any differences in the tests and surveys can be attributed to differences in how the learner sees the robot and human tutors. In addition, a female experimenter served as the human tutor so that the difference in the voice is also kept minimal. In the robot tutor condition, the lesson began when the participant said "*Merhaba* (Hello)" and the NAO robot recognized the speech (Fig. 1). The robot or human tutor first briefly explained the structure of the lesson, and then introduced the words one by one. Each target word was taught in four steps:

1. The tutor introduced the target English word and asked the participant whether she already knew the word (Note that none of the participants knew any of the target words).
2. The tutor introduced the definition of the target word in Turkish (see Table 1).
3. The tutor asked the participant to utter the target word following the tutor three times.
4. The tutor again defined the word and asked the participant to repeat the definition.

After learning every two target words, the participant was given a mini quiz in which the tutor provided the definitions of the target words and asked the participant for the corresponding word. The lesson lasted for about 20 min.

At the end of the lesson, the robot or the human tutor asked the participant to return to the previous room and find the experimenter they met prior to the lesson. The experimenter administered the immediate production and receptive vocabulary tests. In the *production test*, the experimenter provided the definitions of the learned English words one by one in a randomized order, and the participant was asked to say the corresponding English word. In the *receptive test*, the participant heard the learned English word and was asked to choose a picture that matched the word from four options. As mentioned above, participants in both conditions completed the same post-lesson tests again after about a week.

3 Results

3.1 Robot Tutor versus Human Tutor

One of the goals of this study was to examine whether participants learned L2 vocabulary differently from social robots than they do from human tutors. To examine this question, we compared the two tutor conditions across all four learning outcome measures: immediate production test, immediate receptive test, delayed production test, and delayed receptive test. Figure 2 presents the number of correct responses for each of these tests across the two tutors. As shown in Fig. 2, the post-lesson test scores were generally similar across the two tutor conditions.

These data suggest that L2 vocabulary could be learned from a social robot just as well as it could be learned from a human tutor. To verify this conclusion, we conducted simple Generalized Linear Mixed Models (GLMMs) on each post-lesson test with Tutor Type (robot vs. human) as the sole fixed factor and Word as a random intercept.[3] The results showed that participants scored higher in the human tutor condition than in the robot tutor condition in the immediate receptive test, $B=-0.39$, $SE=0.16$, $Z=-2.41$, $p=0.016$; but no difference was found in the immediate production test, $B=-0.07$, $SE=0.16$, $Z=-0.45$, $p=0.650$; delayed production test, $B=0.28$, $SE=0.19$, $Z=1.52$, $p=0.128$; or the delayed receptive test, $B=-0.28$, $SE=0.15$, $Z=-1.82$; $p=0.069$.

[3] We used GLMMs in these analyses because they can be more powerful than parametric tests such as an ANOVA that assumes a normal distribution, as they allow us to analyze the responses of participants without averaging across trials [48]. As the outcome (the scores of the four post-lesson tests) was a binary variable (correct vs. incorrect), logit (log-odds) was used as the link function. The GLMMs constructed here also tested by-item random intercept to ensure that no effect is driven by specific test items. GLMMs were generated in R [49] using the *lme4.glmer* function [50]. In all models, we included the random effect of item (e.g., L2 words) as some L2 vocabulary words may be inherently more difficult to learn than others. All models were fit by maximum likelihood using adaptive Gauss-Hermite quadrature (nAGQ = 1).

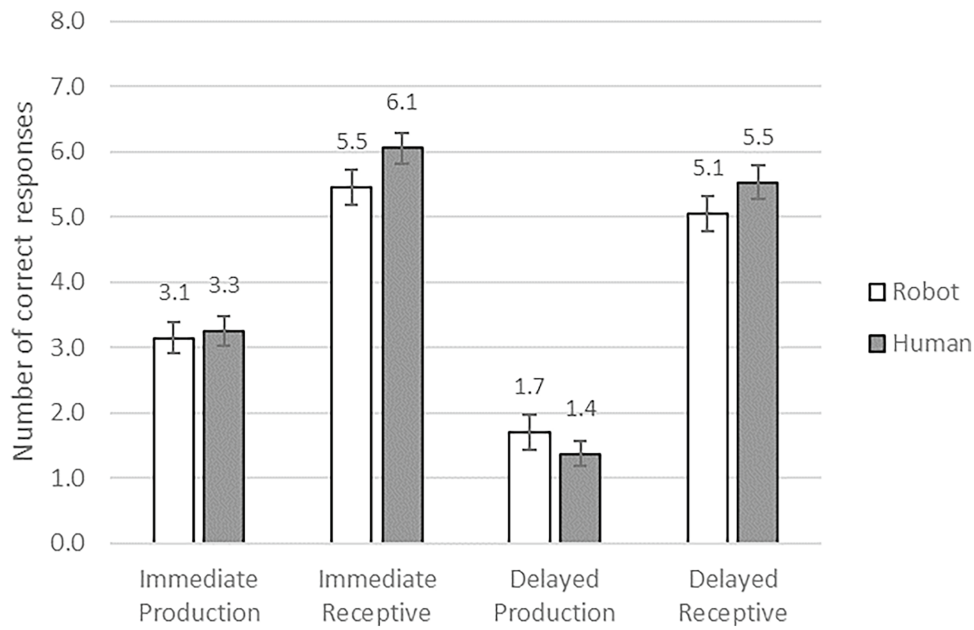

Fig. 2 Mean number of correct answers in the robot tutor and human tutor conditions in the four post-lesson tests. $N=102$ for the immediate production and receptive tests; $N=99$ for the delayed production and receptive tests. The highest possible score for each test was 8. The error bars indicate the standard errors

3.2 Individual Differences and Learning Outcomes

Next, we examined the question of whether some participants learned better or worse from robots depending on their attitudes toward robots, anxiety in L2 learning, and personality traits. The zero-order correlations among the individual difference measures indicated that the correlations between Subscale 3 (S3; negative attitude toward emotions involved in the interaction with robots) of the NARS and the other two subscales were exceptionally low ($r=0.21$ for S1 and $r=0.31$ for S2) in the current dataset. Therefore, the three S3 items were excluded from further analyses for internal consistency. As indicated by Cronbach's alphas in Table 2, each of these variables was measured reliably. Therefore, items measuring each construct were averaged to create relevant indices. For each variable index, values ranged between 1 and 5. Higher values for NARS indicated having more negative attitudes toward robots; similarly, higher values for L2 Anxiety indicated having greater anxiety. We conducted *t*-tests to confirm that participants in the two tutor conditions did not differ in their ratings of seven independent variables measured here (all p's > 0.24).

First, we performed a series of correlational analyses to understand whether these individual difference measures were related to performance in each post-lesson test and should be analyzed in regression models. Table 3 summarizes the correlations between the individual difference measures and the outcome test scores (i.e., number of words learned) within each tutor condition. Having greater anxiety and more negative attitudes toward robots were generally associated with poorer learning outcomes in the robot tutor condition. L2 anxiety scores, in particular, were negatively correlated with all four post-lesson test scores. Personality traits, on the other hand, did not seem to relate to performance significantly in the robot tutor condition. In the human tutor condition, only extroversion was associated with post-lesson tests, suggesting that extroverted participants scored lower on the post-lesson tests than less extroverted participants. No significant correlation between extroversion and learning outcomes was found in the robot condition. These correlations suggest that all three factors (NARS, L2 Anxiety, and Personality) deserve further analyses, and thus we conducted the regression analyses reported in the following sections.

Table 2 Descriptive statistics for the individual difference measures

		Robot tutor		Human tutor	
	α	*Mean*	*SD*	*Mean*	*SD*
NARS (11)	0.83	2.46	0.63	2.44	0.65
L2 Anxiety (33)	0.93	2.50	0.61	2.63	0.68
Personality (44)					
Openness (9)	0.83	4.10	0.55	4.07	0.50
Conscientiousness (7)	0.81	3.57	0.69	3.59	0.72
Extroversion (9)	0.88	3.58	0.65	3.75	0.79
Agreeableness (9)	0.66	3.71	0.51	3.77	0.52
Neuroticism (10)	0.84	3.18	0.69	3.11	0.75

$N=102$. The number in parenthesis indicates the number of items in the scale

Table 3 Correlations between individual difference measures and the outcome test scores (i.e., number of words learned) within each tutor condition

	Robot tutor				Human tutor			
	Immediate production	Immediate receptive	Delayed production	Delayed receptive	Immediate production	Immediate receptive	Delayed production	Delayed receptive
NARS	−0.22	−0.27*	−0.32*	−0.20	0.15	0.02	0.02	−0.03
L2 Anxiety	−0.30*	−0.39*	−0.23	−0.41*	−0.15	0.07	−0.10	0.10
Personality								
O[a]	0.11	−0.09	−0.14	−0.09	−0.05	0.01	−0.06	0.12
C[b]	0.19	0.11	0.11	0.14	0.02	−0.01	0.11	0.18
E[c]	0.09	−0.10	−0.06	−0.02	−0.27*	−0.27*	−0.24*	−0.27*
A[d]	0.01	−0.10	−0.06	−0.13	−0.12	0.00	0.03	−0.01
N[e]	−0.05	0.16	0.01	−0.02	0.01	0.13	−0.10	0.21

[a]O = Openness to experience; [b]C = Conscientiousness; [c]E = Extroversion; [d]A = Agreeableness; [e]N = Neuroticism; For the immediate tests, $N = 51$ in both conditions; For the delayed tests, $N = 50$ in the robot tutor condition and $N = 49$ in the human tutor condition. * $p < 0.05$

Table 4 GLMMs with NARS as the sole predictor for the four post-lesson scores

	Robot tutor				Human tutor			
	B	*SE*	*Z*	*p*	*B*	*SE*	*Z*	*p*
Immediate production	−0.40	0.19	−2.13	0.033	0.25	0.18	1.36	0.173
Immediate receptive	−0.50	0.18	−2.79	0.005	0.03	0.18	0.17	0.862
Delayed production	−0.82	0.22	−3.79	< 0.001	0.04	0.21	0.17	0.868
Delayed receptive	−0.33	0.17	−1.94	0.052	−0.05	0.17	−0.32	0.749

For the immediate tests, $N = 51$ in both conditions; For the delayed tests, $N = 50$ in the robot tutor condition and $N = 49$ in the human tutor condition

3.3 Negative Attitudes Toward Robots

We tested whether individuals who had negative attitudes toward robots learn less from the robot tutor than those with more positive attitudes. To test this proposition, we built four separate GLMMs for each of the four learning outcomes and examined whether negative attitudes toward robots predicted the number of words participants learned in the robot-led L2 lesson. As shown in Table 4, having relatively more negative attitudes toward robots made a difference across all for post-lesson tests. Thus, negative attitudes toward robots can impede learning from a robot tutor.

3.4 L2 Anxiety

We examined the role of L2 learning anxiety similarly by building a model for each post-lesson test for the robot tutor and human tutor conditions. As shown in Table 5, L2 Anxiety made a difference across all four tests in the robot tutor condition. However, to our surprise, L2 Anxiety was not a significant predictor for any of the four test scores in the human tutor condition (all ps > 0.15). Prior to the analysis, we speculated that the L2 anxiety would negatively predict the test scores in both robot tutor and human tutor conditions. However, the models suggest that L2 anxiety disrupts learning when the tutor is a robot, but not when the tutor is another person.

3.5 Personality Traits

To evaluate the relevance of the five personality traits, we followed the same steps and built models separately for the robot and human tutor conditions. We examined whether the personality trait of extroversion is positively related to the post-lesson scores. Unlike other factors, however, the personality traits were not directly relevant to L2 learning from robots, and thus weaker relationships were expected here. Indeed, the results of GLLMs exploring the relationship between each trait and learning outcomes for the robot tutor condition revealed a significant relationship involving only neuroticism in one of the models, for the immediate production test (Table 6). Participants who were high in neuroticism scored higher in the immediate production test when the tutor was a robot. In the human tutor condition, extroversion had a significant relationship to all four learning outcomes. Specifically, learners who were high in extroversion learned less in the lesson when the tutor was another

Table 5 GLMMs with L2 Anxiety as the sole predictor for the four post-lesson scores

	Robot tutor condition				Human tutor condition			
	B	*SE*	*Z*	*p*	*B*	*SE*	*Z*	*p*
Immediate production	−0.55	0.19	−2.88	0.004	−0.25	0.18	−1.43	0.153
Immediate receptive	−0.75	0.19	−3.98	< 0.001	0.13	0.18	0.72	0.472
Delayed production	−0.59	0.22	−2.73	0.006	−0.19	0.21	−0.87	0.383
Delayed receptive	−0.72	0.19	−3.88	< 0.001	0.16	0.17	0.96	0.339

For the immediate tests, $N = 51$ in both conditions; For the delayed tests, $N = 50$ in the robot tutor condition and $N = 49$ in the human tutor condition

person. The regression analysis also found a positive relation between openness to experience and the delayed production test though only in the human tutor condition.

In summary, we found that (1) learners who were high in extroversion learned less in the lesson when the tutor was another person, (2) participants who were high in openness to experience scored higher in the delayed receptive test when the tutor was another person, (3) participants who were high in neuroticism scored higher in the immediate production test when the tutor was a robot.

3.6 Cluster Analysis

To explore the relation between attitudes toward robots and L2 anxiety, we also performed an exploratory two-step cluster analysis. Participants in the robot tutor condition ($N = 51$) were classified into clusters based on their scores for NARS and L2 Anxiety. We used the auto-clustering function of IBM SPSS 18 to select the best cluster solution. The auto-clustering algorithm selects the best solution based on the largest change information criterion measure on the Schwarz Bayesian information criterion (BIC) and the highest ratio of distance measures. The algorithm revealed the two-cluster described in Table 7. The Silhouette measure of cohesion and separation considered the cluster quality as "good."

As shown in Table 7, Cluster 1 consisted of participants with relatively more positive attitudes toward robots. Further, they were less anxious about learning L2 than participants in Cluster 2. Consequently, we expected to observe that participants in Cluster 1 would learn from the robot better than participants in Cluster 2.

The data in Table 8 demonstrate that, when individuals do not have highly negative attitudes toward robots and have little or no anxiety in learning L2, they can learn L2 vocabulary from robot tutors just as well as they do from human tutors; perhaps even better from a robot tutor depending on how the learning outcome is measured. The effect size indices (Cohen's *d*) indicate that the score in the delayed production test was higher in Cluster 1 of the robot tutor condition than in the human tutor condition. The test scores were comparable between the two tutor conditions in the three other tests. When negative attitudes toward robots and anxiety for learning L2 were high, the robot tutor was not as effective as the human tutor.

4 Discussion

As the presence of social robots in our lives is becoming more and more prominent, it is critical to evaluate their efficacy in different settings. The present study used L2 language learning to investigate whether robots can be effective in educational settings. This study also examined how learners' individual differences in attitudes toward robots, anxiety about L2 learning, and personality traits affect their learning outcomes in robot-led L2 lessons. Our study design was more rigorous than most previous studies as it tested a large number of participants, used the number of learned words as a clear and objective measure of the robot's efficacy, and used the human tutor condition as a comparative control to the social robot. Through a stringent evaluation using four different tests of learning outcomes, we found that the robot tutor was as good as the human tutor in teaching L2 vocabulary, and individuals with negative attitudes toward robots learned fewer words in the robot-led lesson than those with more positive attitudes. Partially contrary to our expectations, individuals with higher L2 language anxiety learned fewer words only in the robot tutor condition whereas the personality trait of extroversion was negatively correlated with the learning outcomes only in the human tutor condition. Overall, our results indicate that participants learned L2 vocabulary from the robot tutor at almost equal levels as from the human tutor though attitudes about robots and L2 anxiety of learners played a role in the extent of learning.

4.1 Robot Tutor or Human Tutor?

On average, participants in the robot tutor and human tutor conditions received similar scores in the four post-lesson tests. However, the scores of the immediate receptive test were slightly higher in the human tutor condition than in the robot tutor condition. Receptive tests of vocabulary are sometimes considered a more sensitive measure of

Table 6 GLMMs with the five personality traits as predictors for the four post-lesson scores

	Robot tutor condition				Human tutor condition			
	B	*SE*	*Z*	*p*	*B*	*SE*	*Z*	*p*
Immediate production								
Openness to experience	0.25	0.24	1.05	0.296	0.18	0.27	0.67	0.505
Conscientiousness	0.01	0.17	0.08	0.938	−0.16	0.19	−0.87	0.385
Extroversion	0.14	0.21	0.69	0.493	−0.40	0.16	−2.47	0.014
Agreeableness	−0.07	0.24	−0.29	0.769	0.00	0.23	0.01	0.991
Neuroticism	0.38	0.17	2.15	0.032	−0.04	0.16	−0.28	0.782
Immediate receptive								
Openness to experience	−0.14	0.23	−0.63	0.527	0.28	0.28	0.98	0.328
Conscientiousness	−0.16	0.16	−0.97	0.334	−0.01	0.19	−0.06	0.951
Extroversion	−0.03	0.19	−0.13	0.894	−0.49	0.17	−2.81	0.005
Agreeableness	0.39	0.22	1.75	0.080	0.24	0.24	0.99	0.323
Neuroticism	0.17	0.16	1.03	0.303	−0.10	0.16	−0.58	0.562
Delayed production								
Openness to experience	−0.36	0.25	−1.44	0.150	−0.09	0.34	−0.26	0.796
Conscientiousness	−0.09	0.19	−0.50	0.617	0.25	0.23	1.11	0.269
Extroversion	0.07	0.22	0.32	0.753	−0.41	0.19	−2.19	0.029
Agreeableness	0.13	0.26	0.51	0.613	−0.40	0.29	−1.40	0.163
Neuroticism	0.22	0.19	1.15	0.251	0.21	0.19	1.10	0.273
Delayed receptive								
Openness to experience	−0.13	0.22	−0.60	0.550	0.59	0.28	2.14	0.032
Conscientiousness	−0.19	0.16	−1.15	0.250	−0.10	0.18	−0.59	0.557
Extroversion	0.07	0.19	0.35	0.729	−0.44	0.16	−2.72	0.007
Agreeableness	0.02	0.22	0.09	0.927	0.39	0.23	1.68	0.094
Neuroticism	0.21	0.16	1.30	0.194	0.17	0.16	1.10	0.270

For the immediate tests, $N=51$ in both conditions; for the delayed tests, $N=50$ in the robot tutor condition and $N=49$ in the human tutor condition

Table 7 Clustering participants in the robot tutor condition

Inputs	Cluster 1 (N=24) *M (SD)*	Cluster 2 (N=27) *M (SD)*	*F(1, 49)*	*p*	*Cohen's d*
NARS	2.19 (0.53)	2.71 (0.61)	10.42	0.003	0.91 (0.32/1.47)
L2 Anxiety	1.99 (0.34)	2.95 (0.41)	79.61	0.001	2.53 (1.76/3.23)

Cohen's *d* is a standardized effect size index corresponding to the difference between the two means divided by the pooled estimate of standard deviations; values in parentheses reflect 95% confidence intervals for *d*-values

language knowledge than production tests (e.g., [51]). For example, in a similar study with Dutch 5- to 6-year-olds, children also had a very low score (17.65% accuracy) in the immediate production test (in this case, children translated the Dutch words to the learned English words) [52]. It is difficult to compare adults and children, but those children performed much better in the receptive test (54.57% accuracy), and thus the pattern of difference between the receptive and the production tests was very similar to the current study. We may state that the human tutor was more successful than the robot tutor. Importantly, when participants in the robot tutor condition were divided into two groups through the cluster analysis, participants with relatively positive attitudes toward robots and low anxiety about learning L2 scored equally high or higher than participants in the human tutor condition. Therefore, we conclude that robot tutors can be as effective as human tutors when the attitudes and anxiety of learners do not impede their learning. As mentioned in the introduction, most previous HRI studies in L2 learning focus on motivation [6, 7]. The current study, on the other hand, directly evaluated the efficacy of robot tutors and the benefits of

Table 8 Post-lesson test scores across the clusters in the robot tutor condition compared to the human tutor condition

	Robot tutor		Human tutor	Standardized effect size for the group differences	
	Cluster 1 Mean (SD)	Cluster 2 Mean (SD)	Mean (SD)	Cluster 1 vs. Cluster 2	Cluster 1 vs. Human tutor
Immediate production	3.58 (1.61)	2.74 (1.68)	3.25 (1.58)	$d=0.50$ (−0.05/1.07)	$d=0.21$ (−0.28/0.69)
Immediate receptive	6.08 (1.79)	4.89 (1.99)	6.06 (1.65)	$d=0.62$ (0.06/1.19)	$d=0.01$ (−0.47/0.50)
Delayed production	2.33 (1.86)	1.07 (1.80)	1.35 (1.35)	$d=0.68$ (0.12/1.25)	$d=0.63$ (0.14/1.14)
Delayed receptive	5.75 (1.78)	4.22 (1.87)	5.53 (1.79)	$d=0.82$ (0.26/1.41)	$d=0.12$ (−0.36/0.61)

$N=51$ for the human tutor condition; $N=24$ for Cluster 1 and 27 for Cluster 2. The highest possible score for each test was 8. Cohen's d is a standardized effect size index corresponding to the difference between the two means divided by the pooled estimate of standard deviations; values in parentheses reflect 95% confidence intervals for d-values

learners by assessing the learning outcomes while also comparing them to human tutors.

In this study, although there were some unavoidable differences between the two tutors (e.g., eye blinking of the human tutor, motor sound of the robot tutor), the robot tutor and the human tutor followed the exact same protocol, and their behavioral difference was minimal. As such, their behaviors were not adjusted much to the participant's verbal and non-verbal behaviors. This strict control is a strength of this study as it enables us to assess whether learners learn differently not because of variation in teaching styles or strategies, but purely because their tutor was a robot or a human. A downside of this study design is that we are unable to speak for the difference between an experienced human teacher and the latest technology. It is possible that if the human tutor was allowed to improvise her teaching methods, the lesson became more engaging and the participant learned vocabulary better (although the opposite outcome is technically possible too). The robot tutor could also be programmed to act differently because, as discussed in the introduction, a unique strength of social robots is their ability to adjust behaviors based on the information gathered from sensors.

As robotic technology is advancing very rapidly, the difference between social robots and human tutors may be expected to diminish. Discussing the topic, however, we must also reemphasize that robot tutors should not be considered as a substitute for human teachers, but as a supplemental tool that can provide additional support to learners. As our robot tutor demonstrated adequate abilities to teach L2 vocabulary comparable to that of the human tutor, this study supports the integration of social robots in educational settings where additional support for learners is needed. Nonetheless, as further discussed below, we found that the efficacy of robots also depends on the learner.

4.2 Negative Attitudes Toward Robots and Robot Tutoring

As expected, individuals who had more negative attitudes toward robots learned fewer words compared to their counterparts who had less negative attitudes. According to the aforementioned HRI studies using the NARS, negative attitudes toward robots predict shorter time looking at the robot's face [16], and longer time to until talking to a robot [43]. In another study examining the way participants played a computer game with or without a robot partner, the more negative the player's attitudes toward robots were, the less her behavior was affected by the presence of the robot [53]. Taken together, a possible explanation to the link between negative attitudes and learning outcomes identified in the present study would be that individuals with negative attitudes are less motivated to communicate and interact with the robot tutor and thus learned less in the robot-led lesson. To our knowledge, HRI studies thus far only examined how individuals' attitudes predict changes in their behaviors [16, 20, 25, 53], and the current study makes a novel and unique contribution to the field by demonstrating the direct relationship between the learner's attitudes and learning outcomes.

The NARS covers a wide range of topics concerning both participants' expectations about their personal interaction with robots and the social influence of robots in general. Regardless, all items assess participants' general attitudes toward a robot as a concept, and participants filled out the survey before meeting the NAO robot, and thus they had no expectation specific to the robot tutor of the study. It is noteworthy that participants' general impressions of a type of device (i.e., robot), which were formed prior to meeting the specific interaction partner (i.e., the NAO robot tutor from the lesson), were enough to predict the learning outcomes, though the difference in generality between the independent variables and dependent variables must have led to the relatively weak correlations. Further research may be conducted to examine whether the use of a less general measure (e.g., impression of the NAO robot) results in a stronger correlation.

4.3 L2 Anxiety and Robot Tutoring

An unexpected pattern was found with regard to L2 anxiety. Anxiety regarding learning L2 had a negative influence when an individual learned L2 vocabulary from a robot tutor but not when the lesson was given by a human tutor. Why did L2 anxiety affect the learning outcomes more in the robot tutor condition than in the human tutor condition? One possibility is that the novel and unfamiliar situation of being alone in a room with a robot and learning vocabulary from the robot might have heightened the anxiety [54]. Further, the situation might have been even more difficult for participants in the robot tutor condition who were high in both negative attitudes toward robots and L2 anxiety. The mechanism of the influence as well as ways to mitigate the anxiety should be explored in future research.

4.4 Personality Traits and Robot Tutoring

In contrast to L2 anxiety, personality traits showed more influence on participants who learned L2 vocabulary from the human tutor—extroverted individuals learned fewer words from the human tutor than their less extroverted peers. At first glance, our results may seem peculiar because one possibility we entertained was that extroverted individuals would learn well from the language lesson as they generally like to interact with others. However, as discussed in the introduction, previous findings on the link between extroversion and language learning are mixed with some studies suggesting that introverted individuals are better L2 learners than extroverted individuals [32, 33]. It should be emphasized that the current study tested vocabulary learning and did not test oral communication skills, which were thought to be related to extroversion [31, 33, 37]. But why did we see the relation only in the human tutor condition? One possibility is that, given their preference for social interactions, it is understandable if they felt less motivated with the human tutor who was strictly following the protocol instead of freely interacting with them. Perhaps when the tutor was a robot, the influence was reduced because the robot tutor's behaviors were not less social than the participants expected. In line with this possibility, a study with 7- and 8-year-olds found that a robot that is too social may even impede learning [55].

We also found neuroticism to be a significant predictor in the immediate production test. Among participants in the robot tutor condition, individuals who were high in neuroticism scored higher than their counterparts with relatively low neuroticism. The results seem inconsistent with previous studies that found neuroticism to be associated with general anxiety [41, 42], and to negatively affect factors such as psychological closeness to robots [19]. Though the results are very interesting, it is difficult to draw a conclusion regarding neuroticism based solely on one of the four tests, and experiments focusing specifically on these traits may be needed.

4.5 Limitations of the Study

The present study makes a unique contribution to research on social robots as the first to systematically evaluate the effects of individual differences on learning outcomes. Nevertheless, some shortcomings of the study also need to be addressed. First, participants in this study only learned the definitions and pronunciations of eight words. By administering two different tests, production and receptive tests, we made sure that participants truly understood and learned the meanings of the words. Nonetheless, vocabulary learning goes much further than what has been tested in the present research [56]. For example, language learners would not be able to use the learned words unless they learn how those words can be used in actual sentences. Second, with a single study, we cannot rule out the possibility that features that are specific to our robot tutor, such as its friendly appearance [57, 58] and female voice [59], have affected the learning outcomes. Finally, it should also be noted that, although our study demonstrated the link between attitudes toward robots and the learning outcomes, it does not speak for the exact mechanism underlying the effect nor the ways to improve the learning experience of individuals who have negative attitudes toward robots.

4.6 Future Research

One promising direction for future research is exploring other aspects of L2 learning, such as grammar and speaking. For instance, based on the research on the relation between extroversion/introversion and language learning [31, 37], we may expect that extroverted learners improve their speaking abilities whereas introverted individuals show an advantage in learning grammar. Language learning goes far beyond the memorization of vocabulary, and much more research is needed to gain a full picture of how robot tutors can improve L2 learning.

Future studies must also assess the mechanism by which attitudes toward robots affect vocabulary learning, though a few speculations can be already made based on the data. For instance, individuals who are worried about interacting with robots may be less engaged and pay less attention to the lesson content. It is also possible that they pay more attention to the appearance and behaviors of the robot tutor than the lesson content. Another construct that may underlie the relationship between negative attitudes toward robots and learning outcomes is trust. An individual's beliefs about an informant can influence their trust to learn from that informant [60]. As briefly discussed in the introduction, different models have been proposed to explain how the users' attitudes toward technology are generally formed, e.g., [12, 13]. Based on such general theoretical frameworks for technology, future HRI research should aim to identify the exact mechanism by which attitudes toward robots affect the learning outcomes of robot-led lessons and address whether the unique characteristics of social robots demand a different framework.

5 Conclusion

The current study is the first to document the relation between individual differences and the learning outcomes of robot-led learning. By empirically demonstrating that negative attitudes toward robots and L2 anxiety can impede learning, we highlight the importance of diverging from a one-size-fits-all model and recognizing the diversity among learners. In addition to the theoretical contribution, this study exemplifies a novel way to empirically explore the use of social robots in education and motivates a new set of research questions. Further, our findings provide valuable insights for not only HRI research but also research on language learning in general and contribute to the development of effective robot-led curriculums for individual learners. Educators and policymakers must carefully consider ways to pre-assess and alleviate L2 anxiety and negative attitudes toward robots before they introduce these tools to students.

Appendix 1

The Turkish version of the Negative Attitudes toward Robots Scale (NARS; [43]) used in the present study. The back-translation of the Turkish items are indicated in parentheses (the back-translation is different from the original English version of the NARS). The subscales are: S1 (negative attitudes toward interacting with robots), S2 (negative attitudes toward the social influence of robots), and S3 (negative attitude toward emotions involved in the interaction with robots). In this study, the whole survey was administered, but S3 was dropped from analyses as it did not highly correlate with S1 and S2.

Subscale	Item
S2	*Eğer robotların kendi duyguları olursa kaygılı hissederim* (I will feel anxious if robots have their own emotions.)
S2	*Robotların insanlara daha çok benzemesinin insanoğlu açısından olumsuz bir sonucu olacağını düşünüyorum* (I surmise that there will be negative consequences for humans when robots become more similar to humans.)
S3	*Robotlarla etkileşime girersem kendimi rahat hissederim* (I will feel comfortable if I interact with robots.)
S1	*Robotların kullanıldığı bir iş yerinde çalıştığımı hayal ettiğimde kaygılı hissederim* (I feel anxiety when I imagine that I may be employed or assigned to a workplace where robots are used.)
S3	*Eğer robotların kendi duyguları olursa kendimi onlara yakın hissederim* (I will feel close to robots if they have their own emotions.)
S3	*Robotların duygusal davrandıklarını gördüğümde kendimi daha rahat hissederim* (I feel more comfortable when I see robots behaving affectively.)
S1	*Robotlar hakkında bir şey duyduğumda bile kendimi çaresiz hissediyorum* (I feel helpless even by hearing something about robots.)
S1	*Başkalarının önünde robot kullanacak olursam kendimi utandırabilirim* (I am likely to be embarrassed when I use robots in public.)
S1	*"Yapay zekanın verdiği kararlar" veya "robotların verdiği kararlar" gibi ifadeler beni rahatsız ediyor* (The words "artificial intelligence" or "decision by robots" make me feel unpleasant.)
S1	*Sadece robotların önünde durmak bile bende gerginlik yaratır* (Even standing in front of robots will strain me.)
S2	*Robotlara aşırı bağlı olmak gelecekte olumsuzluğa sebep olabilir* (I surmise that becoming extremely dependent on robots will have negative consequences for humans in the future.)
S1	*Robotlarla etkileşime girersem kendimi tedirgin hissederim* (I will feel nervous if I interact with robots.)

Subscale	Item
S2	*Robotların çocukların zihnini olumsuz yönde etkileyeceklerinden korkuyorum* (I am afraid that robots may negatively influence children's minds.)
S2	*Gelecekteki toplumlara robotların hükmedeceği kanısındayım* (I surmise that robots may dominate future societies.)

Acknowledgments We would like to thank Dr. Tatsuya Nomura for permission to use the NARS and for his advice in translating the scale. We also thank Idil Franko, Orhun Uluşahin, Seref Can Esmer, and Mine Yürekli for their contribution as research assistants.

Funding This research was funded by the EC H2020 L2TOR Project (Grant 688014).

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

References

1. Ortega L (2009) Understanding second language acquisition. Routledge, Taylor & Francis Group, London
2. Bartneck C, Forlizzi J (2004) A design-centred framework for social human-robot interaction. In: RO-MAN 2004: 13th IEEE international workshop on robot and human interactive communication. Institute of Electrical and Electronics Engineers, Kurashiki, Japan, pp 591–594
3. Dore RA, Zosh JM, Hirsh-Pasek K, Golinkoff RM (2017) Chapter 4—Plugging into word learning: the role of electronic toys and digital media in language development. In: Blumberg FC, Brooks PJ (eds) Cognitive development in digital contexts. Academic Press, San Diego, pp 75–91
4. Han J-H, Jo M-H, Jones V, Jo J-H (2008) Comparative study on the educational use of home robots for children. J Inf Process Syst 4:159–168. https://doi.org/10.3745/JIPS.2008.4.4.159
5. Kennedy J, Baxter P, Belpaeme T (2015) Comparing robot embodiments in a guided discovery learning interaction with children. Int J Soc Robot 7:293–308
6. Kanero J, Geçkin V, Oranç C et al (2018) Social robots for early language learning: current evidence and future directions. Child Dev Perspect 12:146–151. https://doi.org/10.1111/cdep.12277
7. Belpaeme T, Kennedy J, Ramachandran A et al (2018) Social robots for education: a review. Sci Robot 3:eaat5954. https://doi.org/10.1126/scirobotics.aat5954
8. Bloom BS (1984) The 2 sigma problem: the search for methods of group instruction as effective as one-to-one tutoring. Educ Res 13:4–16. https://doi.org/10.3102/0013189X013006004
9. 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. ACM, pp 423–430
10. Schodde T, Bergmann K, Kopp S (2017) Adaptive robot language tutoring based on Bayesian knowledge tracing and predictive decision-making. In: Proceedings of the 2017 ACM/IEEE international conference on human-robot interaction. Association for Computing Machinery, New York, NY, pp 128–136
11. Chamorro-Premuzic T, von Stumm S, Furnham A (2015) The Wiley-Blackwell handbook of individual differences. Wiley-Blackwell, Malden, MA
12. Kidd E, Donnelly S, Christiansen MH (2018) Individual differences in language acquisition and processing. Trends Cogn Sci 22:154–169. https://doi.org/10.1016/j.tics.2017.11.006
13. Davis FD (1989) Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS Q 13:319–340. https://doi.org/10.2307/249008
14. Venkatesh V, Morris MG, Davis GB, Davis FD (2003) User acceptance of information technology: toward a unified view. MIS Q 27:425–478. https://doi.org/10.2307/30036540
15. Barnett T, Pearson AW, Pearson R, Kellermanns FW (2015) Five-factor model personality traits as predictors of perceived and actual usage of technology. Eur J Inf Syst 24:374–390. https://doi.org/10.1057/ejis.2014.10
16. Ivaldi S, Lefort S, Peters J et al (2017) Towards engagement models that consider individual factors in HRI: on the relation of extroversion and negative attitude towards robots to gaze and speech during a human–robot assembly task: experiments with the iCub humanoid. Int J Soc Robot 9:63–86. https://doi.org/10.1007/s12369-016-0357-8
17. Tapus A, Ţăpuş C, Matarić MJ (2008) User—robot personality matching and assistive robot behavior adaptation for post-stroke rehabilitation therapy. Intel Serv Robot 1:169. https://doi.org/10.1007/s11370-008-0017-4
18. Gockley R, Matarić MJ (2006) Encouraging physical therapy compliance with a hands-off mobile robot. In: Proceedings of the 1st ACM SIGCHI/SIGART conference on human–robot interaction. Association for Computing Machinery, Salt Lake City, UT, pp 150–155
19. Salem M, Lakatos G, Amirabdollahian F, Dautenhahn K (2015) Would you trust a (faulty) robot? Effects of error, task type and personality on human–robot cooperation and trust. In: 2015 10th ACM/IEEE international conference on human–robot interaction (HRI), pp 1–8
20. Takayama L, Pantofaru C (2009) Influences on proxemic behaviors in human-robot interaction. In: 2009 IEEE/RSJ international conference on intelligent robots and systems, pp 5495–5502
21. Robert L (2018) Personality in the human robot interaction literature: a review and brief critique. Social Science Research Network, Rochester, NY
22. der Pütten AR, Weiss A (2015) The uncanny valley phenomenon: does it affect all of us? Interact Stud 16:206–214. https://doi.org/10.1075/is.16.2.07ros
23. Kanda T, Hirano T, Eaton D, Ishiguro H (2004) Interactive robots as social partners and peer tutors for children: a field trial. Hum-Comput Interact 19:61–84
24. Robins B, Dautenhahn K, Boekhorst RT, Billard A (2005) Robotic assistants in therapy and education of children with autism: can a small humanoid robot help encourage social interaction skills? Univ Access Inf Soc 4:105–120. https://doi.org/10.1007/s10209-005-0116-3
25. Nomura T, Kanda T, Suzuki T, Kato K (2008) Prediction of human behavior in human–robot interaction using psychological scales for anxiety and negative attitudes toward robots. IEEE Trans Robot 24:442–451. https://doi.org/10.1109/TRO.2007.914004
26. Debreli E, Oyman N (2016) Students' preferences on the use of mother tongue in English as a foreign language classrooms: is it the time to re-examine English-only policies? Engl Lang Teach 9:148–162

27. Liu J (2008) L1 use in L2 vocabulary learning: facilitator or barrier. Int Educ Stud 1:65–69
28. Horwitz EK, Horwitz MB, Cope J (1986) Foreign language classroom anxiety. Mod Lang J 70:125. https://doi.org/10.2307/327317
29. Horwitz E (2001) Language anxiety and achievement. Annu Rev Appl Linguist 21:112–126. https://doi.org/10.1017/S0267190501000071
30. Alemi M, Meghdari A, Basiri NM, Taheri A (2015) The effect of applying humanoid robots as teacher assistants to help Iranian autistic pupils learn English as a foreign language. In: Tapus A, André E, Martin JC, Ferland F, Ammi M (eds) Social robotics. ICSR 2015. Lecture Notes in Computer Science, vol 9388. Springer, Cham
31. Ożańska-Ponikwia K (2017) Extraverts and introverts in the FL classroom setting. In: Second language learning and teaching, pp 93–105
32. Dornyei Z, Ryan S (2015) The psychology of the language learner revisited. Routledge, New York, NY
33. Dewaele J-M (2012) Personality in second language acquisition. In: The encyclopedia of applied linguistics. American Cancer Society
34. Naiman N (1978) The good language learner: a report. Ontario Institute for Studies in Education, Toronto
35. Ehrman M (2008) Personality and good language learners. In: Griffiths C (ed) Lessons from good language learners. Cambridge University Press, Cambridge, pp 61–72
36. Carrell PL, Prince MS, Astika GG (1996) Personality types and language learning in an EFL context. Lang Learn 46:75–99. https://doi.org/10.1111/j.1467-1770.1996.tb00641.x
37. Ellis R (2004) Individual differences in second language learning. In: The handbook of applied linguistics. Blackwell, Malden, MA, pp 525–551
38. Dewaele J-M, Furnham A (1999) Extraversion: the unloved variable in applied linguistic research. Lang Learn 49:509–544. https://doi.org/10.1111/0023-8333.00098
39. MacIntyre PD, Clement R, Noels KA (2007) Affective variables, attitude and personality in context. In: Ayoun D (ed) French applied linguistics. John Benjamins Publishing Company, Amsterdam, pp 270–298
40. Wakamoto N (2009) Extroversion/introversion in foreign language learning: interactions with learner strategy use. Peter Lang, Bern
41. Kotov R, Gamez W, Schmidt F, Watson D (2010) Linking "big" personality traits to anxiety, depressive, and substance use disorders: a meta-analysis. Psychol Bull 136:768–821. https://doi.org/10.1037/a0020327
42. Paulus DJ, Vanwoerden S, Norton PJ, Sharp C (2016) From neuroticism to anxiety: examining unique contributions of three transdiagnostic vulnerability factors. Personal Individ Differ 94:38–43. https://doi.org/10.1016/j.paid.2016.01.012
43. Nomura T, Kanda T, Suzuki T (2006) Experimental investigation into influence of negative attitudes toward robots on human–robot interaction. AI Soc 20:138–150. https://doi.org/10.1007/s00146-005-0012-7
44. Aydın S, Harputlu L, Güzel S et al (2016) A Turkish version of foreign language anxiety scale: reliability and validity. Procedia Soc Behav Sci 232:250–256. https://doi.org/10.1016/j.sbspro.2016.10.011
45. Demir B, Kumkale GT (2013) Individual differences in willingness to become an organ donor: a decision tree approach to reasoned action. Personality Individ Differ 55:63–69. https://doi.org/10.1016/j.paid.2013.02.002
46. Dunn LM, Dunn DM (2007) Peabody picture vocabulary test—fourth edition (PPVT). NCS Pearson, Minneapolis, MN
47. Dahlbäck N, Jönsson A, Ahrenberg L (1993) Wizard of Oz studies—why and how. Knowl-Based Syst 6:258–266. https://doi.org/10.1016/0950-7051(93)90017-N
48. Jaeger TF (2008) Categorical data analysis: away from ANOVAs (transformation or not) and towards logit mixed models. J Mem Lang 59:434–446. https://doi.org/10.1016/j.jml.2007.11.007
49. R Core Team (2020) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna
50. Bates D, Mächler M, Bolker B, Walker S (2015) Fitting linear mixed-effects models using lme4. J Stat Softw 67:1–48. https://doi.org/10.18637/jss.v067.i01
51. Shin D, Chon YV, Kim H (2011) Receptive and productive vocabulary sizes of high school learners: what next for the basic word list? Engl Teach 66:127–152
52. Vogt P, van den Berghe R, de Haas M et al (2019) Second language tutoring using social robots: A large-scale study. In: 2019 14th ACM/IEEE international conference on human–robot interaction (HRI), pp 497–505
53. Hinz NA, Ciardo F, Wykowska A (2019) Individual differences in attitude toward robots predict behavior in human–robot interaction. In: Salichs M et al (eds) Social robotics. ICSR 2019. Lecture Notes in Computer Science, vol 11876. Springer, Cham. https://doi.org/10.1007/978-3-030-35888-4_7
54. Daly J (1991) Understanding communication apprehension: an introduction for language educators. In: Horwitz EK, Young DJ (eds) Language anxiety: from theory and research to classroom implications. Prentice Hall, Englewood Cliffs, NJ, pp 3–14
55. Kennedy J, Baxter P, Belpaeme T (2015) The robot who tried too hard: social behaviour of a robot tutor can negatively affect child learning. In: Proceedings of the tenth annual ACM/IEEE international conference on human–robot interaction. Association for Computing Machinery, Portland, OR, pp 67–74
56. Clenton J, Booth P (2021) Vocabulary and the four skills: pedagogy, practice, and implications for teaching vocabulary. Routledge, Abingdon
57. Goetz J, Kiesler S, Powers A (2003) Matching robot appearance and behavior to tasks to improve human-robot cooperation. In: The 12th IEEE international workshop on robot and human interactive communication, 2003. Proceedings. ROMAN 2003, pp 55–60
58. Kanda T, Miyashita T, Osada T et al (2008) Analysis of humanoid appearances in human–robot interaction. IEEE Trans Rob 24:725–735. https://doi.org/10.1109/TRO.2008.921566
59. Schermerhorn P, Scheutz M, Crowell CR (2008) Robot social presence and gender: Do females view robots differently than males? In: 2008 3rd ACM/IEEE international conference on human–robot interaction (HRI), pp 263–270
60. Landrum AR, Eaves BS, Shafto P (2015) Learning to trust and trusting to learn: a theoretical framework. Trends Cogn Sci 19:109–111. https://doi.org/10.1016/j.tics.2014.12.007

Junko Kanero is an Assistant Professor of Psychology and the Principal Investigator of the Mind, Language, and Technology Lab at Sabancı University in Istanbul, Turkey. She received her Ph.D. in Developmental Psychology and Neuroscience from Temple University in Philadelphia, PA, USA, and worked as a Postdoctoral Researcher at Koç University in Istanbul. Her main research focuses are language development in infancy and childhood, the neural processing of language, the interaction between language and non-linguistic cognition, and the use of technology in language education.

Cansu Oranç is a postdoctoral researcher at the research group iSearch (Information Search, Ecological and Active learning Research with Children) at the Max Planck Institute for Human Development in Berlin, Germany. She received her Ph.D. degree in Psychology from Koç University in Istanbul, Turkey. Her research focuses on children′s learning with technology. She is specifically interested in children′s intuitions about and interactions with intelligent artificial agents, such as social robots and voice assistants, in formal and informal learning settings.

Sümeyye Koşkulu is a Ph.D. Candidate at the Department of Education and Pedagogy at Utrecht University in Utrecht, the Netherlands. She received her M.A. degree in the Department of Psychology from Koç University, Istanbul, Turkey. Her research focuses on the relation of parent/teacher-child interactions to cognitive development particularly language and executive function development of children.

G. Tarcan Kumkale is an Associate Professor of Psychology at Özyegin University. He received his Ph.D. in Social Psychology from the University of Florida in 2004. His areas of research expertise include the social psychology of attitudes and persuasion, social influence, health communication, behavior change interventions, and personality.

Tilbe Göksun is an Associate Professor of Psychology at the Department of Psychology at Koç University in Istanbul, Turkey. She received her Ph.D. in Developmental Psychology from Temple University in Philadelphia, PA, USA, and worked as a postdoctoral researcher in the Center for Cognitive Neuroscience at the University of Pennsylvania in Philadelphia, PA, USA. She leads the Language and Cognition Lab at Koç University. Her research focuses on early language and cognitive development, language-thought interactions, neuropsychology of language, and multimodal communication and thinking.

Aylin C. Küntay is a Professor of Psychology and the Dean of the College of Social Sciences and Humanities at Koç University. She received her Ph.D. in Developmental Psychology in 1997 from the University of California, Berkeley. Her work is on the relation of early communicative and language development to social interaction and cognitive development in bilingual and monolingual children.