

University of Groningen

## Teaching Turkish-Dutch kindergartners Dutch vocabulary with a social robot

Leeuwestein, Hanneke; Barking, Marie; Sodacı, Hande; Oudgenoeg-Paz, Ora; Verhagen, Josje; Vogt, Paul; Aarts, Rian; Spit, Sybren; de Haas, Mirjam; de Wit, Jan

*Published in:*  
Journal of Computer Assisted Learning

*DOI:*  
[10.1111/jcal.12510](https://doi.org/10.1111/jcal.12510)

**IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.**

*Document Version*  
Publisher's PDF, also known as Version of record

*Publication date:*  
2021

[Link to publication in University of Groningen/UMCG research database](#)

### *Citation for published version (APA):*

Leeuwestein, H., Barking, M., Sodacı, H., Oudgenoeg-Paz, O., Verhagen, J., Vogt, P., Aarts, R., Spit, S., de Haas, M., de Wit, J., & Leseman, P. (2021). Teaching Turkish-Dutch kindergartners Dutch vocabulary with a social robot: Does the robot's use of Turkish translations benefit children's Dutch vocabulary learning? *Journal of Computer Assisted Learning*, 37(3), 603-620. <https://doi.org/10.1111/jcal.12510>

### **Copyright**

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.










### **Take-down policy**

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

*Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.*

## ARTICLE

# Teaching Turkish-Dutch kindergartners Dutch vocabulary with a social robot: Does the robot's use of Turkish translations benefit children's Dutch vocabulary learning?

Hanneke Leeuwestein<sup>1,4</sup>  | Marie Barking<sup>2</sup> | Hande Sodaci<sup>2</sup>  |  
Ora Oudgenoeg-Paz<sup>1</sup>  | Josje Verhagen<sup>3</sup>  | Paul Vogt<sup>2</sup>  | Rian Aarts<sup>2</sup>  |  
Sybren Spit<sup>3</sup> | Mirjam de Haas<sup>2</sup>  | Jan de Wit<sup>2</sup>  | Paul Leseman<sup>1</sup> 

<sup>1</sup>Utrecht University, Utrecht, The Netherlands

<sup>2</sup>Tilburg University, Tilburg, The Netherlands

<sup>3</sup>University of Amsterdam, Amsterdam, The Netherlands

<sup>4</sup>University of Groningen, Groningen, The Netherlands

## Correspondence

Hanneke Leeuwestein, Faculty of Behavioural and Social Sciences, University of Groningen, Grote Kruisstraat 2/1, 9712 TS Groningen, The Netherlands.  
Email: h.leeuwestein@rug.nl

## Abstract

Providing first language (L1) translations in L2 vocabulary interventions may be beneficial for L2 vocabulary learning. However, in linguistically diverse L2 classrooms, teachers cannot provide L1 translations to all children. Social robots do offer such opportunities, as they can be programmed to speak any combination of languages. This study investigates whether providing L1 translations in a robot-assisted L2 vocabulary training facilitates children's learning. Participants were Turkish-Dutch kindergartners ( $n = 67$ ) who were taught six Dutch (L2) words for which they knew the L1 (Turkish), but not the L2 Dutch form. Half of these words were taught by a Turkish-Dutch bilingual robot, alongside their Turkish translations; the other half by a monolingual Dutch robot. Children also completed Dutch and Turkish receptive vocabulary tests. Results of generalized linear regression models indicated better performance in the Dutch-only condition than in the Turkish-Dutch condition. Children with well-developed Turkish and Dutch vocabulary knowledge outperformed children with less well-developed vocabulary knowledge. The majority of children preferred working with the bilingual robot, but children's preference did not affect word learning. Thus, contrary to our prediction, we found no evidence for a facilitating effect of providing L1 translations through a robot on bilingual children's L2 word learning.

## KEYWORDS

first language translations, kindergartners, robot-assisted language learning, second language, vocabulary

In many countries across the globe, classrooms are becoming increasingly diverse in terms of students' cultural and linguistic backgrounds (OECD, 2015; Vertovec, 2007). The Netherlands is no exception: Many children grow up with a first language (L1) other than the majority language Dutch. Yet, systematic exposure to Dutch as their second

language (L2) often only starts when these children enter kindergarten. In the Netherlands, the Turkish-Dutch form the largest immigrant group (Statistics Netherlands, 2017). Studies indicate that, on average, Dutch language skills and academic achievements of Turkish-Dutch bilingual children lag behind their monolingual Dutch peers (Hartgers,

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2020 The Authors. *Journal of Computer Assisted Learning* published by John Wiley & Sons Ltd.

Kuipers, & Linder, 2018; Scheele, Leseman, & Mayo, 2010). Therefore, gaining insight into effective strategies for supporting these children's L2 development is highly relevant. Several authors (e.g., Creese & Blackledge, 2015; García, 2009) have proposed to use children's earlier developed and stronger L1 to support L2 learning, following the linguistic interdependence hypothesis (Cummins, 1979). As languages often share their conceptual systems underlying word meaning, this might be particularly helpful in L2 vocabulary learning (Scott & de la Fuente, 2008). Moreover, using children's L1 in L2 instruction might be reassuring and motivating for these bilingual children in an otherwise L2 immersion context (Creese & Blackledge, 2015; Pulinx, Van Avermaet, & Agirdag, 2017).

In the current study, we focused on using children's L1 (Turkish) as a strategy to boost their L2 (Dutch) vocabulary learning. Importantly, policy makers and teachers aiming to implement such a strategy face a major practical issue: In classrooms which are becoming increasingly diverse in terms of students' linguistic backgrounds, teachers simply lack the language skills required to provide L1 input for all students. Several digital technologies have been developed to facilitate bilingual language learning (Golonka, Bowles, Frank, Richardson, & Freynik, 2014), of which the use of social robots has recently gained attention and stands out because of the possibility to provide one-on-one interactions with input in multiple languages (Belpaeme et al., 2018a; Belpaeme, Kennedy, Ramachandran, Scassellati, & Tanaka, 2018b; Kanero et al., 2018; van den Berghe, Verhagen, Oudgenoeg-Paz, van der Ven, & Leseman, 2018). In this study, we aim to explore whether a bilingual robot tutor can indeed facilitate children's L2 vocabulary learning by providing them with L1 translations. Specifically, we investigate whether Turkish-Dutch bilingual kindergartners learn more L2 (Dutch) words from a bilingual robot that provides L1 (Turkish) translations than from a monolingual Dutch robot that does not provide such translations. In so doing, we aim to contribute to the ongoing debate on the effectiveness of L1 use in L2 educational practices, and to the exploration of opportunities for integrating social robots into language education.

## 1 | THEORETICAL BACKGROUND

### 1.1 | Role of L1 in L2 learning

So far, there is no consensus as to whether and how providing children's L1 in educational settings could support children's L2 acquisition (Swain & Lapkin, 2013). Educational policies usually encourage exclusive use of the majority language in schools, reflecting the view that bilingual children should form two independent linguistic systems—one system for each of their languages (as described by Creese & Blackledge, 2015; Extra & Yagmur, 2010). This approach to language teaching has been described as “*parallel monolingualism*” (Heller, 1999), “*bilingualism through monolingualism*” (Swain, 1983) and “*separate bilingualism*” (Creese & Blackledge, 2008). These terms reflect the view of the bilingual speaker as “two monolinguals in one body” (Gravelle, 1996, p. 11).

Conversely, several studies have proposed facilitative effects between L1 and L2 skills (e.g., Bialystok, Luk, & Kwan, 2005; Bouvy, 2000; Cummins, 1981), based on the idea that bilingual children rely on one shared conceptual basis, enabling them to transfer their L1 skills to their L2 (Cummins, 1981). Studies have shown that L1 activation takes place whilst processing L2 vocabulary (Sunderman & Kroll, 2006), likely due to children mapping L2 words to the corresponding L1 representations (Hall, 2002). In cases where there is conceptual similarity between L1 and L2, children's L1 vocabulary skills can thus boost L2 vocabulary learning (Scott & de la Fuente, 2008). This suggests that, in L2 learning classrooms, students use their L1 to relate to L2 concepts and meanings which they have already acquired in their L1.

Previous studies have found indications of such a relation between L1 and L2 skills. For example, Leseman, Henrichs, Blom, and Verhagen (2019) found that L1 vocabulary growth of 3- to 6-year-old Turkish-Dutch children positively predicted their L2 vocabulary size and growth. Moreover, the quantity of rich input these children received in their L1 was associated with their L2 vocabulary size. Similarly, Verhoeven (2007) found concurrent and longitudinal relations between L1 and L2 skills in 5- to 6-year-old Turkish-Dutch children, such that L1 skills were found to positively predict L2 skills in the domain of phonological awareness.

Incorporating learners' L1 in classrooms may also enhance bilingual children's socio-emotional well-being and identity development. Previous studies have shown that belonging to an ethnic or cultural group becomes part of the identity development of 3- to 5-year-old children (Ruble et al., 2004), and that they perceive their L1 to be a crucial marker of their identity (le Page & Tabouret-Keller, 1985). Using children's L1 as a tool in L2 instruction signals that their L1 is valued, also in a context in which another language is more prevalent. In fact, studies have shown that L1 use can indeed enhance engagement (Holmes, 2008) and participation in a task (Probyn, 2005), as well as foster self-esteem, which in turn contributes to positive identity development (Creese & Blackledge, 2015; Pulinx et al., 2017).

### 1.2 | Incorporating L1 in the classroom

Despite the findings indicating that L1 use facilitates both L2 learning and children's socio-emotional well-being, minority languages are rarely used in formal education (Duarte, 2019; Gogolin, 2011). Zooming in on the classroom, Cole (1998) argues how “the struggle to avoid L1 at all costs can lead to bizarre behaviour: One can end up being a contortionist trying to explain the meaning of a language item, where a simple translation would save time and anguish” (p. 2).

The use of direct translation into L1 to support L2 vocabulary learning has been studied in adults, but results are inconclusive. While some studies suggest that providing L1 translations to L2 words results in better learning of these words than providing word definitions or explanations in the L2 (Celik, 2003; Latsanyphone & Bouangeune, 2009; Laufer & Shmueli, 1997; Liu, 2008; Ramachandran & Rahim, 2004), others find no such benefits (Joyce, 2018). However, studies in this field often suffer from multiple methodological

shortcomings, such as the absence of a pre-test (Laufer & Shmueli, 1997) or using different languages in the experimental task and the post-tests (Latsanyphone & Bouangeune, 2009; Ramachandran & Rahim, 2004), making it difficult to draw clear conclusions (for a review, see Joyce, 2018). Moreover, evidence from young children is still missing and studies mostly refer to situations where the teacher and students all share the same L1. In linguistically diverse classrooms, it is virtually impossible for teachers to provide appropriate L1 translations for every child (Gogolin, 2011). Social robots that can be programmed to speak multiple languages could be used to overcome this practical constraint.

### 1.3 | Robot-assisted language learning

In recent years, several digital technologies have been developed to support L2 learning (for a review, see Golonka et al., 2014). These technologies have the advantage of providing learning content in one-on-one interactions while using multiple languages. One type of technology that has recently gained attention is the use of social robots (Belpaeme, Kennedy, et al., 2018b; Belpaeme, Vogt, et al., 2018a; Kanero et al., 2018; van den Berghe et al., 2018). Social robots stand out among other technologies by supporting language learning in a physically grounded situation while being in the world that is shared with the language learner. From an embodied cognition perspective, robots can engage in semi-naturalistic and multimodal interactions using non-verbal cues, such as iconic or deictic gestures, that can facilitate the identification of referents of utterances (Hollich et al., 2000; Tomasello & Todd, 1983; Yu & Smith, 2012), and trigger the activation of several neural pathways that could help strengthen the associations between words and meaning (Glenberg, Goldberg, & Zhu, 2011). In addition, social robots can, by being present, act like a human collocator, being capable of using eye gaze and other behaviours to signal interest and empathy (Belpaeme, Kennedy, et al., 2018b).

Various studies have shown that child-robot interactions can contribute to L2 learning, especially when focusing on the use of feedback and motivational support (Herberg, Feller, Yengin, & Saerbeck, 2015; Kennedy, Baxter, Senft, & Belpaeme, 2016; Saerbeck, Schut, Bartneck, & Janse, 2010). For instance, social robots that use iconic gestures to depict some aspect of the target meanings in teaching L2 vocabulary (e.g., mimicking the flapping of wings when teaching the word "bird") have been shown to positively affect word retention in Dutch kindergartners (de Wit et al., 2018). However, a recent study (de Wit, Brandse, Kraemer, & Vogt, 2020) has suggested that a beneficial effect of iconic gestures is found only for children above the age of 5.5 years. Other reported advantages of social robots are that generally children find interacting with them engaging and motivating (Gordon et al., 2016; Hong, Huang, Hsu, & Shen, 2016; Kory Westlund & Breazeal, 2015), and that robots can help reduce children's anxiety to talk in their L2 (Alemi, Meghdari, & Ghazisaedy, 2015).

With regard to vocabulary teaching, research on robots' effectiveness has thus far produced mixed results (for a review, see van den Berghe et al., 2018). Several studies show no or only small effects of robots on L1 and L2 learning (Gordon et al., 2016; Kanda, Hirano,

Eaton, & Ishiguro, 2004; Movellan, Eckhardt, Virnes, & Rodriguez, 2009; Vogt et al., 2019). Other studies found larger learning gains when using methods such as storytelling (Kory Westlund et al., 2015), playing "I spy with my little eye" with a robot that uses iconic gestures (de Wit et al., 2018), and a teaching paradigm in which children learn L2 words by teaching these words to a robot (Tanaka & Matsuzoe, 2012). Similarly, a review by Kanero et al. (2018) concluded that robots can successfully teach vocabulary to young children, though not better than human teachers. However, robots are not meant to replace human teachers. In fact, they are often employed as teaching aids, to complement human teachers (van den Berghe et al., 2018). In the case of using L1 to support L2 learning, robots have a clear advantage: They can be programmed to speak virtually any combination of languages, while human teachers are not always able to provide input in children's L1.

### 1.4 | The present study

#### Research goals

In this study, we investigate whether using a social robot that translates L2 words into Turkish-Dutch bilingual children's L1 (Turkish) enhances L2 (Dutch) word learning, compared to a social robot that only instructs L2 words without L1 translations. To the best of our knowledge, this study is the first to investigate the effects of providing translations in L1 on the learning of L2 in a vocabulary learning experiment using social robots.

#### Research design

In an experimental design, participating children were taught Dutch words for which they knew the Turkish, but not the Dutch label, using a social robot. Half of the target words were taught by a bilingual robot which provided Turkish translations (L2-L1 condition); the other half by a monolingual robot using only Dutch (L2-only condition).

#### Research questions and hypotheses

First, we examined children's learning gains in both conditions. As several studies suggested that the use of L1 translations is beneficial for L2 vocabulary learning (Celik, 2003; Latsanyphone & Bouangeune, 2009; Laufer & Shmueli, 1997; Liu, 2008; Ramachandran & Rahim, 2004), we hypothesized that learning gains would be higher when L1 translations were provided (L2-L1 condition). Third, as is common in vocabulary learning interventions, learning gains were measured with both an immediate and a delayed post-test. This method allows for the assessment of both direct learning and word retention (Marulis & Neuman, 2010). We hypothesized that differences in learning gains between conditions may only appear on the delayed post-test, due to consolidation effects (for a review, see Axelsson, Williams, and Horst, 2016). Fourth, previous studies (e.g.,

Joyce, 2018; Ramachandran & Rahim, 2004) suggested future research to look into differential effects for low- and high-proficiency language learners. Therefore we tested whether children's existing vocabulary knowledge in Turkish and Dutch moderated the effects. This analysis was exploratory. Finally, based on the idea that L1 use would foster children's cultural identity and their engagement and, in turn, positively affect children's learning gains (Holmes, 2008; Pulinx et al., 2017), we expected most children to prefer the bilingual robot. Moreover, we tentatively predicted that children who preferred the bilingual robot would benefit more from the L1 translations than children who preferred the monolingual Dutch robot.

## 2 | METHOD

### 2.1 | Participants

We recruited 67 Dutch children with a Turkish immigration background. These children (34 girls) were aged between 4 and 6 years ( $M = 4$  years and 9 months,  $SD = 6$  months). They were recruited from kindergartens of nine primary schools in different cities in the Netherlands. Six were public schools ( $n = 32$  children), and three were Islamic schools ( $n = 35$  children). A passive informed consent procedure was used, as approved by the ethics committees of the universities involved.

All children's parents were either born in the Netherlands or migrated to the Netherlands before their children were born. Children were therefore labelled as second (83%) or third (17%) generation immigrants with a Turkish background. Three children were born in Germany or Belgium, but had moved to the Netherlands before enrollment in kindergarten. Of the participating children, 10% were from families in which both parents had attended primary education only, 65% came from families in which at least one of the parents had attended secondary or vocational education, and 25% of the children came from families in which at least one parent had attended education at college or university level.

Information on home language use was acquired via a written questionnaire filled in by children's parents (see Table 1). None of the parents reported speaking only Dutch to their child. Hence, all children had been exposed to Turkish. This is in line with the general finding that families with a Turkish background living in the Netherlands show relatively high patterns of maintaining their Turkish language (Extra, Aarts, van der Avoird, Broeder, & Yağmur, 2002; Hartgers, 2012). More than

half of the parents spoke mostly Turkish with their child and only 8% used mostly Dutch, which justifies referring to Turkish as L1 and to Dutch as L2 of these children. We are however aware that language use and input may vary across contexts and the L1-L2 division is not always straightforward.

### 2.2 | Procedures

A Softbank Robotics NAO humanoid robot (58 cm tall) was used in conjunction with a tablet as an intermediate device. As previous research showed high levels of heterogeneity in language proficiencies among Turkish-Dutch kindergartners (e.g., Demir-Vegter, Aarts, & Kurvers, 2014; Mayo & Leseman, 2008), a counterbalanced within-participant design was used to make sure these individual differences would not affect the results. The experiment with two conditions (L1 only condition, L1-L2 condition) started with a group introduction to the robots. Next, a pre-test took place on the same day up to 10 days prior to the experiment ( $M = 2.40$  days,  $SD = 2.90$  days). After the pre-test was completed, children participated in the main experiment, which was videotaped. Children were individually picked up from their classrooms and took part in the experiment in a separate room at their school. They were seated in front of the tablet, next to the robot, and started the familiarization phase. Then, they started their lesson with either robot Robin (L2-only condition) or robot Deniz (L1-L2 condition) with the only visible difference between the robots being the different shirts they were wearing. Once the lesson for the first condition was finished, the robot was switched (by changing the shirt out of sight). Then the lesson for the second condition started, followed by an assessment of children's robot preference. Finally, an immediate post-test was administered to measure target word knowledge. This complete session including post-tests lasted approximately 40 minutes. About 1 week after the experiment ( $M = 8.07$  days,  $SD = 2.70$  days, range = 6–18 days), the target word knowledge test was administered again, followed by Dutch and Turkish receptive vocabulary tests. An overview of all steps in the experiment is presented in Table 2.

#### 2.2.1 | Robot introduction in groups

Previous studies on child-robot interaction have stressed the importance of an introduction prior to experimental sessions, to reduce children's shyness or anxiety when being confronted with a robot for the first time (e.g., Belpaeme, Vogt, et al., 2018a; Han, Jo, Jones, & Jo, 2008). Therefore, children were introduced to respectively the monolingual and the bilingual robot in small groups of 3 to 10 children between 1 week to a few hours prior to the experiment. In both introductions, children listened to a story about the robot visiting the children's hometown and then performed a dance with the robot. Five children were not able to attend the robot introductions. The L2-only robot was introduced to the children as Robin, and they were explicitly told that Robin could speak only Dutch. The L2-L1 robot was

**TABLE 1** Frequencies of the language(s) parents communicate in with their child ( $n = 62^a$ )

What language(s) do you use when speaking to your child?	Number of children
Turkish only	5
Mostly Turkish, and a little bit of Dutch	30
Equally both Turkish and Dutch	22
Mostly Dutch, and a little bit of Turkish	5

<sup>a</sup>Questionnaire data were missing for five children.

**TABLE 2** Overview of all steps in the experimental design

Design step	Description	Average time allocated
Robot introduction in groups	Introducing children to the monolingual robot and the bilingual robot prior to the experiment.	20 min
Pre-test: Target word selection	Testing children's existing knowledge of all possible target words in both Turkish and Dutch to personalize the target words of the experiment for each child.	10 min
Main experiment		30 min
Familiarization phase	Providing children with an opportunity to practice the experimental tasks.	
Experimental phase		
Condition 1	Presenting children with either the L2-only condition or the L2-L1 condition (randomly assigned).	
Short break to change robots	Changing the robot's shirt for the second condition.	
Condition 2	Presenting children with the other condition.	
Assessment: Robot preference	Asking children whether they would prefer playing again with either the monolingual or the bilingual robot.	1 min
Immediate post-test: Target word knowledge	Assessing children's receptive knowledge of their individual target words immediately after the experiment.	10 min
Delayed post-test: Target word knowledge	Assessing children's receptive knowledge of their personal target words a week after the main experiment.	10 min
Dutch and Turkish receptive vocabulary tests	Standardized task to assess children's receptive vocabulary knowledge in Dutch and Turkish a week after the experiment.	10 min

introduced as Deniz, a robot that could speak both Turkish and Dutch.

### 2.2.2 | Pre-test: Target word selection

As earlier works with our target group reported enormous variation in terms of language proficiency (Demir-Vegter et al., 2014; Mayo & Leseman, 2008), it was not possible to find target words that *all* children would know in Turkish, but not in Dutch. Therefore, we conducted a pre-test to measure whether children knew a set of 20 words in Turkish and in Dutch. The results of this pre-test were then used to make an individual selection of 6 target words for each child. The pre-test contained 16 motion verbs and 4 spatial prepositions (postpositions in Turkish) that had all been retrieved from the basic word list for kindergartners in Amsterdam ("*Basiswoordenlijst Amsterdamse Kleuters*"; Mulder, Timman, & Verhallen, 2009). An overview of the target words and the frequency with which each word was selected for the main experiment is included in Table A1).

During the pre-test, children saw four pictures on a tablet screen and heard one of the target words. They were then asked to choose the corresponding picture. The words had been pre-recorded by native Dutch and Turkish speakers. For each child, we selected the words that the child knew in Turkish (i.e., the child was able to choose the correct picture), but not in Dutch. Children with less than four possible target words ( $n = 28$ ) were excluded from participation. For

the majority of children, six target words could be selected ( $n = 38$ ). To keep the length of the experimental session equal for all children, children for whom four ( $n = 20$ ) or five ( $n = 9$ ) target words could be selected received one or two extra "filler words," to arrive at a total of six words for all children during the experiment. These fillers were words children already knew in both languages or in Dutch only. Each target word and each filler word was alternately assigned to the conditions, resulting in an equal number of occurrences of each target word across the two conditions and across children. The order in which target words were taught within each condition was randomized.

### 2.2.3 | Main experiment

All children took part in a one-on-one session with the robot. The learning task involved an animated game, presented on a Microsoft Surface Pro tablet with a 12.3-inch display. The experiment started with a familiarization phase in which the experimenter was actively involved to make sure children understood what they were supposed to do. During this phase, children practiced the three tasks of the main experiment: (1) selecting pictures on the tablet by tapping on them, (2) repeating words produced by the robot and (3) acting out the target word using Sesame Street plush toys. The plush toys were added to the experiment, as this allowed children to physically enact the reference of target words. From an embodied cognition perspective, this



could, potentially, benefit the learning of these words (Glenberg et al., 2011).

The familiarization phase was followed by the experimental phase, which consisted of the L2-only and L2-L1 lessons (in counterbalanced order), and a short break in between to change the robot's shirt. In the experiment, children were instructed to look for animals in a big and colourful picture of a forest displayed on the tablet. Each target word was separately introduced through either an animation in which an animal performed a specific action such as *shaking* or *clapping* (for verbs) or a static picture showing a specific position relative to trees (for prepositions), such as *behind* or *in front of* a tree.

The whole robot-child interaction was pre-programmed in a script, illustrated in Table 3 for the two conditions. The script was the

**TABLE 3** Model of the interaction script for L2-only and L2-L1 condition

L2-only condition—Monolingual robot	
Robot speech	Hey, look! I see a monkey. The monkey is <u>shaking</u> the tree. Touch the monkey that is <u>shaking</u> the tree on the screen.
Child action	<i>Children select monkey on the tablet.</i>
Tablet speech	<u>Shaking</u> . The monkey is <u>shaking</u> .
Robot speech	Did you hear that? So, in Dutch we say " <u>shaking</u> ." The monkey is <u>shaking</u> the tree. Say after me: <u>Shaking</u> .
Child action	<i>Children repeat after the robot.</i>
Robot speech	Can you <u>shake</u> the dolls? Get them out of the bag. Come on, show me how you can <u>shake</u> the dolls.
Child action	<i>Children act out target word with the dolls.</i>
Robot speech	Okay, let us put them in the bag again.
End task	Hey, where do you see a monkey <u>shaking</u> ?
L2-L1 condition—Bilingual robot	
Robot speech	Look! Do you see the monkey that is <u>shaking</u> the tree? In Turkish we say <u>sallamak</u> . Touch the monkey on the screen.
Child action	<i>Children select monkey on the tablet.</i>
Tablet speech	<u>Shaking</u> . The monkey is <u>shaking</u> .
Robot speech	Oh, so it is <u>sallamak</u> in Turkish and <u>shaking</u> in Dutch. Say after me: <u>Shaking</u> .
Child action	<i>Children repeat after the robot.</i>
Robot speech	Can you <u>shake</u> the dolls? Get them out of the bag. Come on, show me how you can <u>shake</u> the dolls.
Child action	<i>Children act out target word with the dolls.</i>
Robot	Okay, let us put them in the bag again.
End task	Hey, where do you see a monkey <u>shaking</u> ?

Note: In the original (Dutch) interaction, all target words were presented as infinitives to avoid different forms due to inflection.

same across both conditions, except for the phrase "in Turkish we say X", which was repeated twice for each target word in the L2-L1 condition to provide the Turkish translations of the target words. In the L2-only condition, children were exposed to each Dutch target word at least 10 times. In the L2-L1 condition, children were exposed to the Dutch target word at least 8 times, complemented with two Turkish translations. Differences in the number of exposures were due to the variable amount of feedback children received. Three percent of all target words in total were presented less often due to technical issues; only 8 or 9 times in the L2-only condition, or 6 or 7 times in the L2-L1 condition. Thus, in total, children heard each target word at least 6 and at most 19 times ( $M = 10.81$ ,  $SD = 1.69$  in the L2-only condition, and  $M = 8.94$ ,  $SD = 1.65$  in the L2-L1 condition). Target word exposure served as a control variable in all analyses.

## 2.3 | Data collection

### 2.3.1 | Target word knowledge post-tests

Immediate and delayed post-tests were administered to assess children's receptive knowledge of the six target words, using a picture selection task. In this task, children saw three videos or pictures on a tablet screen and heard a target word in Dutch. They were then asked to choose the video or picture corresponding to the target word. All the words had been pre-recorded by a native speaker of Dutch. Videos were used when the target word was a verb, and pictures were used for prepositions. To compensate for children selecting the target words randomly, three trials per target word were presented, resulting in 18 trials. Using a program script in Python, we created individual tests for each child based on the different sets of target words that they had been exposed to during the experiment. These tests were composed as follows: For each trial, one distractor depicted another target word that the particular child had been exposed to in the experiment and one distractor depicted a word that was not included in the experiment for that particular child. This was done in order to avoid children being able to determine the correct answer based on the distractors. For children with only four or five target words who had been given additional (filler) words, these filler words were also tested to keep testing time equally long for all children. However, responses to these filler words were not included in the analyses.

### 2.3.2 | Dutch and Turkish receptive vocabulary tests

The Diagnostic Test of Bilingualism was used to measure children's receptive vocabulary knowledge in both Dutch and Turkish (Verhoeven, Narrain, Extra, Konak, & Zerrouk, 1995). In this standardized task, children were presented with a test word and four different pictures and asked to select the picture that best matched the target word. The official test contains 60 words of increasing difficulty, to be tested in both languages in two separate sessions. Following previous

studies (Messer, Leseman, Boom, & Mayo, 2010), testing time was reduced by splitting the test into two parts of 30 words for each language. Specifically, two versions of the test were created, one in which the odd items were assessed in Dutch and the even items in Turkish, another version which was the other way around. Following the standard protocol, testing was stopped after three consecutive incorrect answers. Correct answers were summed, resulting in two final scores (maximum = 30), one for each language. Test scores did not significantly differ between the two versions for Dutch and Turkish. Cronbach's alpha values for the different test parts ranged from .89 to .93.

### 2.3.3 | Robot preference

Immediately after the experiment, children were shown pictures of the monolingual and bilingual robot (i.e., two robots wearing different shirts) (see Figure 1). The experimenter then asked the children which robot they would prefer to play with again, as a proxy for children's preference for either one of the robots (forced choice).

## 2.4 | Data analyses

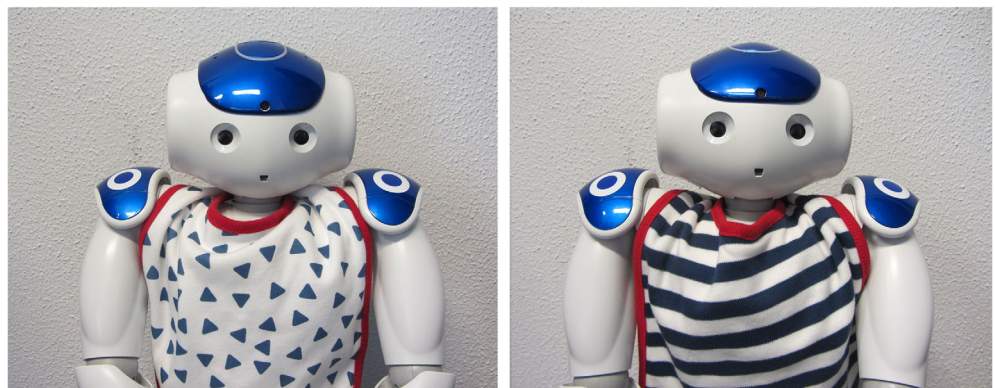
As a first step, we wanted to know whether children learned the vocabulary items at all. To test this, we conducted one sample *t*-tests in which we compared children's performance on the task to chance level (33%). For our main research questions, three separate generalized linear regression models with mixed effects were carried out. These analyses were carried out in R (R Core Team, 2015), using the `lme4` package (Bates, Maechler, Bolker, & Walker, 2015). In all models, orthogonal sum-to-zero contrast coding was applied to our binary fixed effects (i.e., condition, time, robot preference) and all continuous variables were centered around zero (Baguley, 2012, pp. 590–621). Furthermore, to avoid problems with non-converging models, we rescaled our continuous variables by dividing them by 10 (Babyak, 2009) and we increased the number of possible iterations to 100,000 (Powell, 2009). We aimed to keep our models as fully specified as possible by including random intercepts for participants and items as well as all within-participant and within-item factors and

their possible interactions as random slopes for participant and item (Barr, Levy, Scheepers, & Tily, 2013). However, because this was not always supported by our relatively small data set, we always reported on the maximal random effect justified by the data (Jaeger, 2009). We reported simple rather than standardized effect sizes (Baguley, 2009) and Wald confidence intervals (Agresti & Coull, 1998). All analyses are posted on [osf.io/uq2gy](https://osf.io/uq2gy) (Leeuwstein & Spit, 2020).

To answer our main research question on the effect of providing L1 translations of the L2 target words, a mixed effects model was run to investigate the effect of condition (monolingual vs. bilingual robot) and time (post-test 1 vs. post-test 2) on children's scores in the target word knowledge post-tests. We conducted a generalized linear regression model, with children's scores on the target word knowledge test (on item level) as a dependent variable (0 = incorrect, 1 = correct), condition (L2-only, L2-L1) and time (post-test 1, post-test 2) as within-participants fixed effects, and the number of exposures as a fixed controlling factor. By-participant and by-item random slopes for condition, time and number of exposures, but not their interactions, were included, because they were both within-participant and within-item fixed effects.

In a subsequent analysis, Turkish and Dutch vocabulary scores were added as covariates to the model, because we aimed to explore possible moderation effects of children's vocabulary knowledge in these languages on their L2 learning outcomes. In this model, scores on the target word knowledge test (0 or 1) were entered as the dependent variable, condition (L2-only, L2-L1) and time (post-test 1, post-test 2) as within-participants fixed effects, and the number of exposures, Dutch vocabulary score, and Turkish vocabulary score as fixed controlling factors. By-participant and by-item random slopes for condition, time and number of exposures, but not their interactions, were included, because they were both within-participant and within-item fixed effects.

A final model also included children's preference for either the monolingual or bilingual robot, to investigate whether children's preference for a particular robot modulated any effect of condition on children's L2 learning gains. Specifically, to investigate whether children's robot preference affected learning gains differentially between the conditions, preference was added as a between-participant factor in this regression model. As before, the dependent variable in this model involved children's scores on the items in the target word



**FIGURE 1** Stimuli of robot preference: Monolingual robot (left) and bilingual robot (right) [Colour figure can be viewed at [wileyonlinelibrary.com](https://wileyonlinelibrary.com)]



knowledge test (0 or 1). Within-participant fixed effects were condition (L2-only, L2-L1) and time (post-test 1, post-test 2). Robot preference (preference for monolingual robot, preference for bilingual robot) was entered as a between-participants fixed effect, and the number of exposures, the Dutch vocabulary score and the Turkish vocabulary score were entered as fixed controlling factors. Condition, time and number of exposures, but not their possible interactions, were included as random slopes for participant, because they were within-participant fixed effects. Only the number of exposures was included as a random slope for item, as it was a within-item fixed effect.

### 3 | RESULTS

Table 4 displays descriptive statistics for the target word knowledge scores for both conditions. Children performed above chance level (33%) in all post-tests: ( $t(65) = 9.62, p < .001$ ) and ( $t(66) = 9.88, p < .001$ ) for post-tests 1 and 2 of the L2-only condition, respectively, and ( $t(65) = 9.74, p < .001$ ) and ( $t(65) = 11.58, p < .001$ ) for the L2-L1 condition, meaning that children learned new L2 words through the one-on-one sessions with the robot. Table 4 also shows the average number of exposures to a target word in Dutch during the training. Recall that in the L2-L1 condition children were exposed to a Turkish translation of the Dutch target word twice. Data of one child were missing for post-test 1 due to refusal to participate, data of one other child were missing for post-test 2 in the L2-L1 condition due to technical issues.

#### 3.1 | Learning gains in L2-only and L2-L1 condition

A first generalized regression model was run to investigate whether learning gains were higher when children were provided with L1 translations by the robot (L2-L1 condition) than when they did not receive such translations (L2-only condition). The results of this analysis are shown in Table 5. We found a main effect of condition in the opposite direction, such that children performed significantly better with the L2-only robot than with the L2-L1 robot ( $OR = 2.11, 95\% CI = [1.13, 3.94], z = 2.35, p = .019$ ). No effect of time was found: Although children performed better a week later at post-test 2, this difference was not significant ( $OR = 1.29, 95\% CI = [0.94, 1.79],$

$z = 1.56, p = .119$ ). There was a main effect of target word exposure, which indicated that performance on the target word knowledge test decreased with *increasing* exposure ( $OR = 0.76, 95\% CI = [0.63, 0.92], z = -2.84, p = .005$ ). Recall that target word exposure increased when children needed additional instructions to complete the task. None of the interactions in this model were significant.

#### 3.2 | Examining possible moderation effects of dutch and turkish vocabulary skills

Next, we ran a generalized linear regression model, in which Dutch and Turkish vocabulary scores, as assessed with the Diagnostic Test of Bilingualism, were added as covariates to explore possible moderation effects on the above-reported main effect of condition (L2-only vs. L2-L1). As in the first model, the results of this model, presented in Table 6, showed a main effect of condition, indicating that children performed significantly better with the monolingual robot than with the bilingual robot ( $OR = 2.91, 95\% CI = [1.55, 5.46], z = 3.33, p = .001$ ). We found no main effect of time. In line with the first model, there was a main effect of exposures: Performance on the

**TABLE 5** Results of the generalized linear regression model with scores from the target word knowledge test as a dependent variable, condition and time as fixed effects and number of exposures as a fixed controlling factor

	$\beta$	SE	z	p
Condition	0.75	0.32	2.35	.019
Time	0.26	0.17	1.56	.119
Exposures	-2.70	0.95	-2.84	.005
Time * Condition	-0.32	0.26	-1.23	.219
Time * Exposures	1.18	0.66	1.78	.074
Condition * Exposures	1.12	1.33	0.85	.396
Time * Condition * Exposures	0.69	1.31	0.52	.601

Note: As all continuous variables were rescaled,  $\beta$ -values are not in an interpretable scale either. To get sensible values, values for effects with one rescaled variable should be divided by 10, values for effects with two rescaled variables by 100, and values for effects with three rescaled variables by 1,000. This holds for all three reported models and their outcomes.

**TABLE 4** Mean proportions correct on the target word knowledge test for both conditions in the immediate (post-test 1) and delayed post-test (post-test 2), and mean number of exposures to the target words

	L2-only condition			L2-L1 condition		
	M	SD	Range	M	SD	Range
Post-test 1	0.66 (n = 66)	0.27	0-1	0.62 (n = 66)	0.24	0-1
Post-test 2	0.69 (n = 67)	0.29	0-1	0.69 (n = 66)	0.25	0-1
Target word exposure in Dutch	10.80	1.20	8.33-14.00	8.92	1.33	6.67-15.33

**TABLE 6** Results of the generalized linear regression model with scores from the target word knowledge as a dependent variable, condition and time as within-participants fixed effects, and the number of exposures and Dutch and Turkish vocabulary scores as fixed controlling factors

	$\beta$	SE	Z	p
Time	0.11	0.20	0.53	.595
Condition	1.07	0.32	3.33	.001
Dutch vocabulary	0.57	0.27	2.12	.034
Turkish vocabulary	0.70	0.24	2.93	.003
Exposures	-3.90	0.84	-4.63	.001
Condition * Dutch vocabulary	0.24	0.50	0.49	.628
Condition * Turkish vocabulary	-0.05	0.46	-0.12	.906

Note: As all continuous variables were rescaled,  $\beta$ -values are not in an interpretable scale either. To get sensible values, values for effects with one rescaled variable should be divided by 10, values for effects with two rescaled variables by 100, and values for effects with three rescaled variables by 1,000. This holds for all three reported models and their outcomes.

knowledge test decreased, when the number of exposures increased ( $OR = 0.68$ , 95% CI = [.57, .80],  $z = -4.63$ ,  $p < .001$ ). Furthermore, a main effect of Turkish vocabulary indicated higher performance for children with higher Turkish vocabulary scores than for children with lower Turkish vocabulary scores ( $OR = 1.07$ , 95% CI = [1.02, 1.12],  $z = 2.93$ ,  $p = .003$ ). Likewise, there was a main effect of Dutch vocabulary, signalling higher performance for children with higher Dutch vocabulary scores than for children with lower Dutch vocabulary scores ( $OR = 1.06$ , 95% CI = [1.00, 1.12],  $z = 2.12$ ,  $p = .034$ ). The interaction effects between condition and Dutch vocabulary, and between condition and Turkish vocabulary were not significant. Several other interactions reached significance, but because of the large number of comparisons these findings should be interpreted with caution. For the full results of the model, see Appendix Table A2.

### 3.3 | Robot preferences

Finally, we examined children's preference for either the monolingual or bilingual robot, and whether their preference for one of the robots affected their learning of the vocabulary items, and whether any such effect interacted with the robot type. The results showed that a majority of the children ( $n = 48$  children, 72%) expressed their preference for the bilingual robot, and only 19 children (28%) stated their preference for the monolingual robot. We ran a final model that included robot preference in addition to the earlier described variables. The model showed a main effect of condition, as children learned significantly more L2 words with the monolingual robot than with the bilingual robot ( $OR = 2.67$ , 95% CI = [1.49, 4.78],  $z = 3.31$ ,  $p = .001$ ). No significant effect of time was found ( $OR = 1.24$ , 95% CI = [0.84, 1.84],  $z = 1.08$ ,  $p = .279$ ). Again, we found a main effect of exposures, which indicated that performance on the target word knowledge test decreased, with more exposures ( $OR = 0.63$ , 95% CI = [0.50,

0.79],  $z = -3.97$ ,  $p < .001$ ). Furthermore, we found positive effects of Turkish vocabulary ( $OR = 1.09$ , 95% CI = [1.03, 1.15],  $z = 3.15$ ,  $p = .002$ ) and Dutch vocabulary ( $OR = 1.10$ , 95% CI = [1.03, 1.16],  $z = 3.08$ ,  $p = .002$ ) on target word knowledge. Finally, the results showed no significant main effect of robot preference ( $OR = 1.57$ , 95% CI = [0.86, 2.84],  $z = 1.48$ ,  $p = .139$ ). Moreover, children's robot preference did not affect their learning gains between conditions ( $OR = 1.29$ , 95% CI = [0.42, 3.97],  $z = 0.44$ ,  $p = .662$ ). Because of the very large number of comparisons in this model, which increases the chance of finding a significant effect, we report no further interaction effects. For the full results of this model, including all interactions, see Appendix Table A3.

In sum, above chance level scores on all post-tests showed that children learned new L2 words through the one-on-one sessions with the robot. Subsequently, across all models, main effects were found for condition, number of exposures, Turkish and Dutch vocabularies, but no significant interaction effects between these factors. Specifically, children obtained higher scores on the target word knowledge test for the words they had learned with the monolingual robot, as compared to the bilingual robot at both time points. Moreover, children's Dutch and Turkish vocabulary scores were positively related to performance on both target word knowledge tests. Last, a vast majority of the children preferred to play with the bilingual robot over the monolingual robot, but children's preference did not affect the difference in learning gains between the L2-only and the L2-L1 conditions.

## 4 | DISCUSSION

The aim of the current study was to investigate whether providing L1 translations during an L2 vocabulary training by a social robot facilitated L2 word learning. A bilingual Turkish-Dutch and a monolingual Dutch social robot were used to teach 67 Turkish-Dutch kindergartners six words for which they knew the Turkish (L1) but not the Dutch (L2) label. The bilingual robot provided Turkish translations of the target words, whereas the monolingual robot used only Dutch.

### 4.1 | Using L2-L1 translations?

First, our findings demonstrate that the experiment using robot tutors to teach L2 words to Turkish-Dutch kindergartners contributed to both direct vocabulary learning and target word retention, meaning that the L2 words were integrated into children's memory (Axelsson et al., 2016). This is in line with findings from related studies in the field of robot-assisted language learning (for a review, see van den Berghe et al., 2018). Second, and contrary to the hypothesis of our main research question, learning gains were higher in the L2-only condition than in the L2-L1 condition. Thus, the robot's provision of Turkish translations of the target words did not additionally improve children's Dutch vocabulary learning, and in fact, resulted in significantly lower vocabulary learning.

There are several possible explanations for this unexpected result. One explanation is that children were not fully prepared for the fact

that the robot provided Turkish translations for the target words, as the use of Turkish words in a Dutch school context is highly uncommon (Extra & Yagmur, 2010). Anecdotal observations during the experiment indeed showed that children were oftentimes surprised by the Turkish translations, especially at the beginning of the L2-L1 condition. Moreover, switching between languages might have placed extra cognitive load on the children. While studies show that bilingual children may be better at tasks requiring cognitive flexibility than monolingual children (Adesope, Lavin, Thompson, & Ungerleider, 2010), in our study, the use of translations did not additionally benefit learning. This might be because the use of L1 in this study was minimal and the setup was rather artificial. Perhaps, a more naturalistic use of L1 within the context of teaching L2 might prove to be more beneficial (for a discussion of this issue, see Ticheloven, Blom, Leseman, & McMonagle, 2019); for a meta-analysis showing modest effects of bilingual education, see Reljic, Ferring, & Martin, 2014).

Additionally, we only measured receptive knowledge and not deep word knowledge in Turkish and Dutch in our pre-test in order to select the target words that children were presented with in the experiment. It is possible that using L1 to support L2 learning is beneficial especially for words for which the concept in L1 is already deeply mapped (Ellis, 2006). It is also possible that the use of L1 may be more beneficial for children with lower levels of L2 proficiency. Although our sample showed variation in L2 levels, most children were born in the Netherlands and already had some knowledge of Dutch. The use of L1 might be more beneficial for first generation immigrant children with very little knowledge of the L2.

## 4.2 | Effects of prior vocabulary knowledge

Our findings demonstrate that children with well-developed Turkish and Dutch receptive vocabulary knowledge outperformed those with less receptive vocabulary knowledge. A similar effect was found in a large-scale study on robot-assisted language learning (Vogt et al., 2019), where children who scored higher on L1 receptive vocabulary also scored higher on the L2 post-test. This seems to be in line with our finding that children with more exposure to the target words (indicating that they made more errors during the experiment), scored lower on the target word knowledge tests. Children with well-developed vocabulary skills in both languages performed better in the vocabulary learning task, needed less additional instructions, were less often exposed to the target words, and performed better at the target word knowledge tests.

## 4.3 | Children's robot preferences

In our study, the vast majority of children showed a preference for the bilingual Turkish-Dutch robot. This is in line with our hypothesis and may be explained along the lines of earlier work stating that the acknowledgement of children's cultural identity by using their L1 in education increases their enjoyment and well-being (Holmes, 2008; Pulinx et al., 2017). Similar effects of robots adapting to specific

cultures were found in a study by Trovato et al. (2013), where adult Egyptian participants preferred interacting with a robot that behaved according to Egyptian cultural standards, compared to a robot that behaved according to Japanese cultural standards. It is also in line with an interview study of Ahmad, Mubin, and Orlando (2016), which showed that primary and high school language teachers stress the importance of culture-based adaptation when using social robots for L2 teaching. Despite children's preference for the Turkish-Dutch bilingual robot in our study, this did not affect their learning gains differentially across conditions. This may be an indication of the complex relation between engagement and learning (Iten & Petko, 2016). For instance, children's enthusiasm towards the bilingual robot may also have distracted them from learning, as also found in a study of Kennedy, Baxter, and Belpaeme (2015) where social and adaptive robot behaviours negatively affected language learning gains. More research is needed to address if and how children's preference may increase motivation and word learning for future interactions with the robot.

## 4.4 | Limitations

This study has several limitations. First, because all children were taught different target words, a variety of factors related to the individual target words may have influenced the results. Due to the heterogeneity of proficiency in both languages in our sample, this design limitation was inevitable. Using mixed-effects regression models, we tried, as much as possible, to reduce the effect of each child learning different target words. Additionally, the counterbalanced within-participant design was chosen because of the high heterogeneity in children's language proficiency in both Turkish and Dutch in general, as well as clear differences in the specific words they knew in Turkish but not in Dutch that could serve as target words across children. Adopting a between-participant design would require careful matching between groups or even at the level of individual children, which was deemed impossible here because of this individual variation. A possible drawback of the within-participant design is that children were presented with both the L2-only and the L1-L2 robot, which raises the question whether this affected their learning outcomes in any way. We think that this was not the case, for two reasons. First, we counterbalanced the conditions, such that any effect of presentation order was cancelled out in our results. Second, we made a clear effort to make children believe there were two robots by changing their shirt out of children's sight and consequently referring to them with different names. In addition, anecdotal evidence indicates that children genuinely believed that there were two different robots, as indicated by some of their statements during or after the experiment (e.g., "Deniz [the bilingual robot] is sleeping now [while playing with Robin the monolingual robot]"). Taken together, we think there are no reasons to believe that our within-participant design affected our results in a substantial way.

Furthermore, the so-called novelty effect may have affected our results. This effect refers to the often observed relatively high

interest when children are introduced to new technology, which may have caused an increase in learning gains due to increased engagement. Single, brief interventions like ours may simply offer participants too little time to get used to the new technology (Han et al., 2008; van den Berghe et al., 2018), as evidenced by various previous robot-assisted language learning studies showing a clear decrease in engagement and learning gains with increasing exposure to a robot (e.g., Fernaeus, Håkansson, Jacobsson, & Ljungblad, 2010; Kanda et al., 2004). As an attempt to reduce a possible novelty effect in our study, we included the robot group introductions, following recommendations made in earlier work (e.g., Han et al., 2008). However, the introduction session was relatively short and the time between the introductions and the experiment varied substantially between children due to planning issues. Hence, the question whether children's unfamiliarity with the robot affected their learning cannot be answered based on our results. Importantly, however, since we counterbalanced children's exposure to the monolingual and bilingual robots in a within-participants design, we can rule out the possibility that the novelty effect influenced learning across our two conditions differently. Yet ideally, future research should contain multiple training sessions and aim to achieve sustained engagement by for example technical improvements such as adaptive behaviour of the robot.

To enable the robot to function autonomously, the experiment session in our study was highly structured with little room for individual variation based on children's progress. Moreover, given the lack of well-functioning Automatic Speech Recognition (Kennedy et al., 2017), the amount of interaction between the children and the robot was limited. Also, optimizing the pronunciation, both in Dutch and Turkish, appeared to be a technological challenge. Slightly different pronunciations of the same target word in different sentences were inevitable. Future technological improvements making the child-robot interaction more naturalistic will likely contribute to the effectiveness of such training programmes. However, these limitations are evident for robot-assisted language learning in general (Van den Berghe et al., 2018), and therefore, apply on the same scale in both conditions in the current study. Thus, these limitations have no consequences for the results of the current study.

## 5 | CONCLUSIONS

Despite the limitations, this study contributes to a growing body of literature on L1 use in L2 teaching, while also exploring the use of social robots in this field. Our results demonstrate that, across both conditions, the experimental task using social robots enhanced L2 word learning among Turkish-Dutch kindergartners. This is an important finding because in the future, social robots could complement human teachers and provide opportunities for L2 learning by one-on-one interactions (Belpaeme, Kennedy, et al., 2018b; Kanero et al., 2018; van den Berghe et al., 2018).

The current study offers no support for providing L1 translations to Turkish-Dutch kindergartners. Future research should further

explore how L1 use could be incorporated into L2 learning programmes to optimally support bilingual children by taking into account the differences in language use and proficiency in both children's L1 and L2. Moreover, it is important for future studies to extend this work to other bilingual groups.

Concluding, the current study asserts that integrating multilingual social robots into language learning programmes in linguistically diverse classrooms holds promises for the future. However, more research is needed to fully understand how to best utilize the possibilities offered by robots to support children's L2 learning. Further technological advancements are needed to enable robots to realize their true educational potential.

## ACKNOWLEDGEMENTS

The authors have declared that no competing interests exist. This study was carried out within the L2TOR project, funded by the European Union's Horizon 2020 research and innovation programme. Grant agreement no. is 688014. All co-authors had complete access to data supporting the manuscript. Ethical approval was granted by the ethics board of the Faculty of Social and Behavioral Sciences at Utrecht University and the ethics board of the University of Tilburg.

## PEER REVIEW

The peer review history for this article is available at <https://publons.com/publon/10.1111/jcal.12510>.

## DATA AVAILABILITY STATEMENT

The data set that supports the findings of this study are openly available on <https://osf.io/uq2gy/>

## ORCID

Hanneke Leeuwestein  <https://orcid.org/0000-0003-0989-5253>

Hande Sodaci  <https://orcid.org/0000-0003-1995-6005>

Ora Oudgenoeg-Paz  <https://orcid.org/0000-0003-4271-7504>

Josje Verhagen  <https://orcid.org/0000-0002-6662-1657>

Paul Vogt  <https://orcid.org/0000-0002-9446-4425>

Rian Aarts  <https://orcid.org/0000-0003-0632-5853>

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

Jan de Wit  <https://orcid.org/0000-0002-7299-7992>

Paul Leseman  <https://orcid.org/0000-0003-1286-0510>

## REFERENCES

- Adesope, O. O., Lavin, T., Thompson, T., & Ungerleider, C. (2010). A systematic review and meta-analysis of the cognitive correlates of bilingualism. *Review of Educational Research, 80*(2), 207–245. <https://doi.org/10.3102/0034654310368803>
- Agresti, A., & Coull, B. A. (1998). Approximate is better than "exact" for interval estimation of binomial proportions. *The American Statistician, 52*(2), 119–126. <https://doi.org/10.1080/00031305.1998.10480550>
- Ahmad, M. I., Mubin, O., & Orlando, J. (2016, October). *Understanding behaviours and roles for social and adaptive robots in education: Teacher's perspective*. Paper presented at Proceedings of the Fourth International Conference on Human Agent Interaction. pp. 297–304. <https://doi.org/10.1145/2974804.2974829>

- 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(4), 523–535. <https://doi.org/10.1007/s12369-015-0286-y>
- Axelsson, E. L., Williams, S. E., & Horst, J. S. (2016). The effect of sleep on children's word retention and generalization. *Frontiers in Psychology*, 7(1192), 1–9. <https://doi.org/10.3389/fpsyg.2016.01192>.
- Babyak, M.A. (2009). *Post to statistical tips from the editors of psychosomatic medicine*. Retrieved from <http://stattips.blogspot.com/2009/08/rescaling-continuous-predictors-in.html>
- Baguley, T. (2009). Standardized or simple effect size: What should be reported? *British Journal of Psychology*, 100(3), 603–617. <https://doi.org/10.1348/000712608X377117>
- Baguley, T. (2012). *Serious stats*. Basingstoke, UK: Palgrave Macmillan.
- Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, 68(3), 255–278. <https://doi.org/10.1016/j.jml.2012.11.001>
- Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1), 1–48.
- Belpaeme, T., Vogt, P., van den Berghe, R., Bergman, K., Göksun, T., De Haas, M., & Kumar, & Pandey, A. (2018a). Guidelines for designing social robots as second language tutors. *International Journal of Social Robotics*, 10(3), 325–341. <https://doi.org/10.1007/s12369-018-0467-6>
- Belpaeme, T., Kennedy, J., Ramachandran, A., Scassellati, B., & Tanaka, F. (2018b). Social robots for education: A review. *Science Robotics*, 3(21), 1. <https://doi.org/10.1126/scirobotics.aat5954-9>.
- Bialystok, E., Luk, G., & Kwan, E. (2005). Bilingualism, biliteracy, and learning to read: Interactions among languages and writing systems. *Scientific Studies of Reading*, 9(1), 43–61. [https://doi.org/10.1207/s1532799xssr0901\\_4](https://doi.org/10.1207/s1532799xssr0901_4)
- Bouvy, C. (2000). Towards the construction of a theory of cross-linguistic transfer. In J. Cenoz & U. Jessner (Eds.), *English in Europe: The acquisition of a third language* (pp. 143–156). Bristol: Multilingual Matters.
- Celik, M. (2003). Teaching vocabulary through code-mixing. *English Language Teaching*, 57(4), 361–369. <https://doi.org/10.1093/elt/57.4.361>
- Cole, S. (1998). The use of L1 in communicative English classrooms. *Japan Association for Language Teaching*, 22, 11–14.
- Creese, A., & Blackledge, A. (2008). *Flexible bilingualism in heritage language schools*. Paper presented at Urban Multilingualism and Intercultural Communication, Antwerp, Belgium.
- Creese, A., & Blackledge, A. (2015). Translanguaging and identity in educational settings. *Annual Review of Applied Linguistics*, 35, 20–35. <https://doi.org/10.1017/S0267190514000233>
- Cummins, J. (1979). Linguistic interdependence and the educational development of bilingual children. *Review of Educational Research*, 49(2), 222–251. <https://doi.org/10.3102/00346543049002222>
- Cummins, J. (1981). *Bilingualism and minority language children*. Ontario, Canada: Ontario Institute for Studies in Education.
- Demir-Vegter, S., Aarts, R., & Kurvers, J. (2014). Lexical richness in maternal input and vocabulary development of Turkish preschoolers in The Netherlands. *Journal of Psycholinguistic Research*, 43(2), 149–165. <https://doi.org/10.1007/s10936-013-92457>
- de Wit, J., Schodde, T., Willemsen, B., Bergmann, K., de Haas, M., Kopp, S., ... & Vogt, P. (2018). *The effect of a robot's gestures and adaptive tutoring on children's acquisition of second language vocabularies*. Paper presented at Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction. ACM. pp. 50–58. <https://doi.org/10.1145/3171221.3171277>
- de Wit, J., Brandse, A., Krahmer, E., & Vogt, P. (2020, March). *Varied human-like gestures for social robots: Investigating the effects on children's engagement and language learning*. Paper presented at Proceedings of the 2020 ACM/IEEE International Conference on Human-Robot Interaction. pp. 359–367. <https://doi.org/10.1145/3319502.3374815>
- Duarte, J. (2019). Translanguaging in mainstream education: A sociocultural approach. *International Journal of Bilingual Education and Bilingualism*, 22(2), 150–164. <https://doi.org/10.1080/13670050.2016.1231774>
- Ellis, N. C. (2006). Selective attention and transfer phenomena in L2 acquisition: Contingency, cue competition, saliency, interference, overshadowing, blocking, and perceptual learning. *Applied Linguistics*, 27(2), 164–194. <https://doi.org/10.1093/applin/aml015>
- Extra, G., Aarts, R., van der Avoird, T., Broeder, P., & Yağmur, K. (2002). *De andere talen van Nederland: Thuis en op school [The other languages of The Netherlands: At home and at school]*. Bussum, The Netherlands: Coutinho.
- Extra, G., & Yagmur, K. (2010). Language proficiency and socio-cultural orientation of Turkish and Moroccan youngsters in The Netherlands. *Language and Education*, 24(2), 117–132. <https://doi.org/10.1080/09500780903096561>
- Fernaues, Y., Håkansson, M., Jacobsson, M., & Ljungblad, S. (2010). *How do you play with a robotic toy animal? A long-term study of Pleo*. Paper presented at Proceedings of the 9th International Conference on Interaction Design and Children. pp. 39–48. <https://doi.org/10.1145/1810543.1810549>
- García, O. (2009). *Bilingual education in the 21<sup>st</sup> century: A global perspective*. Malden, MA: Wiley, Blackwell.
- Glenberg, A. M., Goldberg, A. B., & Zhu, X. (2011). Improving early reading comprehension using embodied CAL. *Instructional Science*, 39(1), 27–39. <https://doi.org/10.1007/s11251-009-9096-7>
- Gogolin, I. (2011). The challenge of super diversity for education in Europe. *Education Inquiry*, 2(2), 239–249. <https://doi.org/10.3402/edui.v2i2.21976>
- 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(1), 70–105. <https://doi.org/10.1080/09588221.2012.700315>
- Gordon, G., Spaulding, S., Westlund, J. K., Lee, J. J., Plummer, L., Martinez, M., ... & Breazeal, C. (2016). *Affective personalization of a social robot tutor for children's second language skills*. Paper presented at Proceedings of the 30th AAAI Conference on Artificial Intelligence. pp. 3951–3957.
- Gravelle, M. (1996). *Supporting bilingual learners in schools*. London, UK: Trentham Books.
- Hall, C. (2002). The automatic cognate form assumption: Evidence for the parasitic model of vocabulary development. *International Review of Applied Linguistics*, 40(2), 69–87. <https://doi.org/10.1515/iral.2002.008>
- Han, J.-H., 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(4), 159–168. <https://doi.org/10.3745/JIPS.2008.4.4.159>
- Hartgers, M. (2012). Onderwijs [education]. In R. van der Vliet, J. Ooijevaar, & R. van der Bie (Eds.), *Jaarrapport Integratie 2012* (pp. 69–92). Den Haag/Heerlen, The Netherlands: Centraal Bureau voor de Statistiek.
- Hartgers, M., Kuipers, N., & Linder, F. (2018). Onderwijs [education]. Chap. 2. In *Jaarrapport Integratie 2018*. Den Haag, The Netherlands: Centraal Bureau voor de Statistiek.
- Heller, M. (1999). *Linguistic minorities and modernity: A sociolinguistic ethnography*. London, UK: Longman.
- Herberg, J. S., Feller, S., Yengin, I., & Saerbeck, M. (2015). *Robot watchfulness hinders learning performance*. Paper presented at 24th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN). pp. 153–160. <https://doi.org/10.1109/ROMAN.2015.7333620>



- Hollich, G. J., Hirsh-Pasek, K., Golinkoff, R. M., Brand, R. J., Brown, E., Chung, H. L., & Bloom, L. (2000). Breaking the language barrier: An Emergentist coalition model for the origins of word learning. *Monographs of the Society for Research in Child Development*, 65(3), 1–135.
- Holmes, J. (2008). *An introduction to sociolinguistics* (3rd ed.). England: Pearson Longman.
- 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. *Journal of Educational Technology & Society*, 19(1), 337–349.
- Iten, N., & Petko, D. (2016). Learning with serious games: Is fun playing the game a predictor of learning success? *British Journal of Educational Technology*, 47(1), 151–163. <https://doi.org/10.1111/bjet.12226>
- Jaeger, T. F. (2009). *Random effect: Should I stay or should I go? Post to HLP/Jaeger lab blog*. Retrieved from <https://hlplab.wordpress.com/2009/05/14/random-effect-structure/>
- Joyce, P. (2018). L2 vocabulary learning and testing: The use of L1 translation versus L2 definition. *The Language Learning Journal*, 46(3), 217–227. <https://doi.org/10.1080/09571736.2015.1028088>
- Kanda, T., Hirano, T., Eaton, D., & Ishiguro, H. (2004). Interactive robots as social partners and peer tutors for children: A field trial. *Human Computer Interaction*, 19(1), 61–84.
- Kanero, J., Geçkin, V., Oranç, C., Mamus, E., Küntay, A. C., & Göksun, T. (2018). Social robots for early language learning: Current evidence and future directions. *Child Development Perspectives*, 12(3), 146–151. <https://doi.org/10.1111/cdep.12277>
- Kennedy, J., Baxter, P., & Belpaeme, T. (2015, March). *The robot who tried too hard: Social behaviour of a robot tutor can negatively affect child learning*. Paper presented at 2015 10th ACM/IEEE International Conference on Human-Robot Interaction (HRI). IEEE. pp. 67–74.
- Kennedy, J., Baxter, P., Senft, E., & Belpaeme, T. (2016). Social robot tutoring for child second language learning. *11th Annual ACM/IEEE International Conference on Human Robot Interaction*, <https://doi.org/10.1109/HRI.2016.7451757>.
- Kennedy, J., Lemaignan, S., Montassier, C., Lavalade, P., Irfan, B., Papadopoulou, F., Senft, E., & Belpaeme, T. (2017). *Child speech recognition in human-robot interaction: Evaluations and recommendations*. Paper presented at Proceedings of the 12th ACM/IEEE International Conference on Human-Robot Interaction. pp. 82–90. <https://doi.org/10.1145/2909824.3020229>
- Kory Westlund, J. & Breazeal, C. (2015). *The interplay of robot language level with children's language learning during storytelling*. Paper presented at Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction Extended Abstracts. pp. 65–66. <https://doi.org/10.1145/2701973.2701989>
- Kory Westlund, J., Dickens, L., Jeong, S., Harris, P., Desteno, D., & Breazeal, C. (2015). *A comparison of children learning new words from robots, tablets, and people*. Paper presented at Conference Proceedings New Friends 2015. pp. 7–8.
- Latsanyphone, S., & Bouangeune, S. (2009). Using L1 in teaching vocabulary to low proficiency level students: A case study at the National University of Laos. *English Language Teaching*, 2(3), 186–193.
- Laufer, B., & Shmueli, K. (1997). Memorizing new words: Does teaching have anything to do with it? *RELC Journal*, 28(1), 89–108. <https://doi.org/10.1177/003368829702800106>
- le Page, R. B., & Tabouret-Keller, A. (1985). *Acts of identity: Creole-based approaches to language and ethnicity*. England: Cambridge University Press.
- Leeuwestein, H., & Spit, S. (2020). *Teaching Turkish-Dutch kindergartners Dutch vocabulary with a social robot*. Retrieved from [osf.io/uq2gy](https://osf.io/uq2gy)
- Leseman, P. P. M., Henrichs, L. F., Blom, E., & Verhagen, J. (2019). Young mono- and bilingual children's exposure to academic language as related to language development and school achievement. In V. Grøver, P. Ucelli, M. Rowe, & E. Lieven (Eds.), *Learning through language*. Cambridge, MA: Cambridge University Press.
- Liu, J. (2008). L1 use in L2 vocabulary learning: Facilitator or barrier. *International Education Studies*, 1(2), 65–69.
- Marulis, L. M., & Neuman, S. B. (2010). The effects of vocabulary intervention on young children's word learning: A meta-analysis. *Review of Educational Research*, 80(3), 300–335. <https://doi.org/10.3102/0034654310377087>
- Mayo, A. Y., & Leseman, P. P. M. (2008). Off to a good start? Vocabulary development and characteristics of early family and classroom experiences of children from native Dutch speaking and bilingual minority families in The Netherlands. *Educational and Child Psychology*, 25(3), 66–78.
- Messer, M. H., Leseman, P. P. M., Boom, J., & Mayo, A. Y. (2010). Phonotactic probability effects in nonword recall and its relationship with vocabulary in monolingual and bilingual preschoolers. *Journal of Experimental Child Psychology*, 105(4), 306–323. <https://doi.org/10.1016/j.jecp.2009.12.006>
- Movellan, J. R., Eckhardt, M., Virnes, M., & Rodriguez, A. (2009). *Sociable robot improves toddler vocabulary skills*. Paper presented at Proceedings of the 4th ACM/IEEE International Conference on Human Robot Interaction. pp. 307–308. <https://doi.org/10.1145/1514095.1514189>
- Mulder, F., Timman, Y., & Verhallen, S. (2009). *Handreiking bij de Basiswoordenlijst Amsterdamse Kleuters (BAK) [Guidelines for the basic word list for kindergartners in Amsterdam]*. Amsterdam, The Netherlands: ITTA and UvA.
- OECD. (2015). *Immigrant students at school: Easing the journey towards integration*, *OECD Reviews of Migrant Education*. Paris, France: OECD Publishing.
- Powell, M. J. D. (2009). *The BOBYQA algorithm for bound constrained optimization without derivatives*. Technical Report, Cambridge: Department of Applied Mathematics and Theoretical Physics, University of Cambridge.
- Probyn, M. (2005). Language and the struggle to learn: The intersection of classroom realities, language policy, and neocolonial and globalisation discourses in South African schools. In A. M. Y. Lin & P. W. Martin (Eds.), *Decolonisation, globalisation: Language-in-education policy and practice* (pp. 153–172). Bristol: Multilingual Matters.
- Pulinx, R., Van Avermaet, P., & Agirdag, O. (2017). Silencing linguistic diversity: The extent, the determinants and consequences of the monolingual beliefs of Flemish teachers. *International Journal of Bilingual Education and Bilingualism*, 20(5), 542–556. <https://doi.org/10.1080/13670050.2015.1102860>
- R Core Team. (2015). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. Retrieved from <https://www.R-project.org/>.
- Ramachandran, S. D., & Rahim, H. A. (2004). Meaning recall and retention: The impact of the translation method on elementary level learners' vocabulary learning. *RELC Journal*, 35(2), 161–178. <https://doi.org/10.1177/003368820403500205>
- Reljic, G., Ferring, D., & Martin, R. (2014). A meta-analysis on the effectiveness of bilingual programs in Europe. *Review of Educational Research*, 85(1), 92–128. <https://doi.org/10.3102/0034654314548514>
- Ruble, D. N., Alvarez, J., Bachman, M., Cameron, J., Fuligni, A., & García Coll, C. (2004). The development of a sense of 'we': The emergence and implications of children's collective identity. In M. Bennett & F. Sani (Eds.), *The development of the social self* (pp. 29–76). London, UK: Psychology Press. [https://doi.org/10.4324/9780203391099\\_chapter\\_2](https://doi.org/10.4324/9780203391099_chapter_2)
- 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*. Paper presented at Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. ACM. pp. 1613–1622. <https://doi.org/10.1145/1753326.1753567>
- Scheele, A. F., Leseman, P. P. M., & Mayo, A. Y. (2010). The home language environment of monolingual and bilingual children and their language proficiency. *Applied Psycholinguistics*, 31(1), 117–140. <https://doi.org/10.1017/S0142716409990191>

- Scott, V. M., & de la Fuente, M. J. (2008). What is the problem? L2 learners' use of the L1 during consciousness-raising form focused tasks. *The Modern Language Journal*, 92(1), 100–113. <https://doi.org/10.1111/j.1540-4781.2008.00689.x>
- Statistics Netherlands. (2017). *Bevolking; kerncijfers [Population; key data]*. Retrieved from <http://statline.cbs.nl>
- Sunderman, G., & Kroll, J. F. (2006). First language activation during second language lexical processing: An investigation of lexical form, meaning, and grammatical class. *Studies in Second Language Acquisition*, 28(3), 387–422.
- Swain, M. (1983). Bilingualism without tears. *TESOL Journal*, 82, 35–46.
- Swain, M., & Lapkin, S. (2013). A Vygotskian sociocultural perspective on immersion education: The L1/L2 debate. *Journal of Immersion and Content-Based Language Education*, 1(1), 101–129. <https://doi.org/10.1075/jicb.1.1.05swa>
- 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(1), 78–95. <https://doi.org/10.5898/JHRI.1.1.Tanaka>
- Ticheloven, A., Blom, E., Leseman, P., & McMonagle, S. (2019). Translanguaging challenges in multilingual classrooms: Scholar, teacher and student perspectives. *International Journal of Multilingualism*, 1–24. <https://doi-org.proxy-ub.rug.nl/10.1080/14790718.2019.1686002>.
- Tomasello, M., & Todd, J. (1983). Joint attention and lexical acquisition style. *First Language*, 4(12), 197–211. <https://doi.org/10.1177/014272378300401202>
- Trovato, G., Zecca, M., Sessa, S., Jamone, L., Ham, J., Hashimoto, K., & Takanishi, A. (2013, August). *Towards culture-specific robot customisation: A study on greeting interaction with Egyptians*. Paper presented at 2013 IEEE RO-MAN. IEEE. pp. 447–452. <https://doi.org/10.1109/ROMAN.2013.6628520>
- van den Berghe, R., Verhagen, J., Oudgenoeg-Paz, van der Ven, S., & Leseman, P. P. M. (2018). Social robots for language learning: A review. *Review of Educational Research*, 89(2), 259–295. <https://doi.org/10.3102/0034654318821286>
- Verhoeven, L. (2007). Early bilingualism, language transfer, and phonological awareness. *Applied Psycholinguistics*, 28(3), 425–439.
- Verhoeven, L., Narrain, G., Extra, G., Konak, O. A., & Zerrouk, R. (1995). *Diagnostische toets tweetaligheid (DTT) [Diagnostic test of bilingualism]*. Arnhem, The Netherlands: CITO.
- Vertovec, S. (2007). Super-diversity and its implications. *Ethnic and Racial Studies*, 30(6), 1024–1054. <https://doi.org/10.1080/01419870701599465>
- Vogt, P., van den Berghe, R., de Haas, M., Hoffmann, L., Kanero, J., Mamus, E., ... Pandey, A. K. (2019). Second language tutoring using social robots. A large-scale study. *14th ACM/IEEE International Conference on Human-Robot Interaction (HRI)* (pp. 497–505). New Jersey, NJ: IEEE. <https://doi.org/10.1109/HRI.2019.8673077>.
- Yu, C., & Smith, L. B. (2012). Embodied attention and word learning by toddlers. *Cognition*, 125(2), 244–262. <https://doi.org/10.1016/j.cognition.2012.06.016>

**How to cite this article:** Leeuwestein H, Barking M, Sodacı H, et al. Teaching Turkish-Dutch kindergartners Dutch vocabulary with a social robot: Does the robot's use of Turkish translations benefit children's Dutch vocabulary learning? *J Comput Assist Learn*. 2021;37:603–620. <https://doi.org/10.1111/jcal.12510>

## APPENDIX

**TABLE A1** Overview of the 20 target words, including the amount each word was selected for the main experiment

Target word English translation	Target word in Dutch	Target word Turkish	Total occurrences	L2-only condition	L2-L1 condition	Filler word in L2-only condition	Filler word in L2-L1 condition
In front of	Voor	Önünde	5	3	2	0	0
Behind	Achter	Arkasında	15	7	8	0	0
Next to	Naast	Yanında	12	6	6	0	0
Between	Tussen	Arasında	5	3	2	0	0
To shiver	Bibberen	Titremek	18	9	9	0	0
To descend	Dalen	Alçalmak	19	9	9	1	0
To close	Dichtdoen	Kapatmak	12	4	4	2	2
To dive	Duiken	Dalmak	12	6	6	0	0
To clap	Klappen	Alkışlamak	14	2	2	4	6
To empty	Legen	Boşaltmak	34	15	15	3	1
To stand up	Opstaan	Kalkmak	16	7	8	1	0
To staple	Opstapelen	Yığmak	8	4	4	0	0
To row	Roeien	Kürek çekmek	23	12	11	0	0
To roll	Rollen	Yuvarlamak	29	7	7	7	8
To shake	Schudden	Sallamak	33	12	12	4	5
To swing	Slingeren	Sallanmak	31	14	14	2	1
To ascend	Stijgen	Yükselmek	42	21	21	0	0
To trip	Struikelen	Ayağı takılmak	36	17	18	1	0
To follow	Volgen	Takip etmek	11	6	5	0	0
To fill	Vullen	Doldurmak	27	13	13	0	1

**TABLE A2** Results of the generalized linear regression model with scores from the target word knowledge as a dependent variable, condition and time as within-participants fixed effects and the number of exposures, a Dutch vocabulary score and a Turkish vocabulary score as fixed controlling factors

	$\beta$	SE	z	p
Time	0.11	0.20	0.53	.595
Condition	1.07	0.32	3.33	.001
Dutch	0.57	0.27	2.12	.034
Turkish	0.70	0.24	2.93	.003
Exposures	-3.90	0.84	-4.63	.001
Time * Condition	-0.04	0.33	-0.134	.892
Time * Dutch	-0.59	0.32	-1.85	.064
Condition * Dutch	0.24	0.50	0.49	.628
Time * Turkish	0.12	0.28	0.43	.667
Condition * Turkish	-0.05	0.46	-0.12	.906
Dutch * Turkish	1.02	0.44	2.32	.021
Time * Exposures	0.18	0.94	0.20	.844
Condition * Exposures	1.07	1.24	0.87	.386
Dutch * Exposures	-2.67	1.30	-2.06	.040
Turkish * Exposures	2.44	1.48	1.65	.098
Time * Condition * Dutch	0.19	0.61	0.31	.754
Time * Condition * Turkish	0.14	0.56	0.25	.799
Time * Dutch * Turkish	0.59	0.50	1.18	.239
Condition * Dutch * Turkish	0.52	0.82	0.64	.524
Time * Condition * Exposures	2.04	1.90	1.08	.282
Time * Dutch * Exposures	-0.29	1.67	-0.18	.860
Condition * Dutch * Exposures	4.54	2.21	2.05	.040
Time * Turkish * Exposures	-1.31	1.81	-0.72	.469
Condition * Turkish * Exposures	-6.68	2.48	-2.70	.007
Dutch * Turkish * Exposures	1.85	2.58	0.72	.473
Time * Condition * Dutch * Turkish	-0.83	1.01	-0.83	.408
Time * Condition * Dutch * Exposures	5.41	3.37	1.61	.108
Time * Condition * Turkish * Exposures	-2.35	3.55	-0.66	.507
Time * Dutch * Turkish * Exposures	2.51	3.37	0.75	.456
Condition * Dutch * Turkish * Exposures	-9.51	4.90	-1.94	.052
Time * Condition * Dutch * Turkish * Exposures	1.11	6.65	0.17	.868

Note: As all continuous variables were rescaled,  $\beta$ -values are not in an interpretable scale either. To get sensible values, values for effects with one rescaled variable should be divided by 10, values for effects with two rescaled variables by 100, and values for effects with three rescaled variables by 1,000. This holds for all three reported models and their outcomes.

**TABLE A3** Results of the generalized linear regression model with scores from the target word knowledge as a dependent variable, condition as within-participants fixed effects, robot preference as a between-participants fixed effect and the number of exposures, a Dutch vocabulary score and a Turkish vocabulary score as fixed controlling factors

	$\beta$	SE	z	p
Time	0.22	0.20	1.08	.279
Condition	0.98	0.30	3.31	.001
Dutch	0.91	0.30	3.08	.002
Turkish	0.84	0.27	3.15	.002
Preference	0.45	0.30	1.48	.139
Exposures	-4.58	1.16	-3.97	.001
Time * Condition	0.02	0.39	0.04	.967
Time * Dutch	-0.85	0.38	-2.22	.026
Condition * Dutch	-0.33	0.55	-0.61	.544
Time * Turkish	-0.04	0.34	-0.13	.896
Condition * Turkish	-0.27	0.50	-0.54	.587
Dutch * Turkish	0.87	0.49	1.78	.076
Time * Preference	0.20	0.39	0.50	.615
Condition * Preference	0.25	0.58	0.44	.662
Dutch * Preference	0.63	0.58	1.09	.277
Turkish * Preference	-0.28	0.53	-0.52	.602
Time * Exposures	1.27	1.61	0.79	.429
Condition * Exposures	-1.33	1.89	-0.71	.481
Dutch * Exposures	0.39	1.67	0.23	.816
Turkish * Exposures	1.22	1.70	0.72	.473
Preference * Exposures	-0.95	2.01	-0.48	.635
Time * Condition * Dutch	0.74	0.74	0.99	.320
Time * Condition * Turkish	-0.46	0.68	-0.67	.502
Time * Dutch * Turkish	1.14	0.66	1.74	.082
Condition * Dutch * Turkish	-0.07	0.93	-0.07	.942
Time * Condition * Preference	0.77	0.78	0.98	.326
Time * Dutch * Preference	-0.45	0.76	-0.59	.555
Condition * Dutch * Preference	0.81	1.10	0.74	.459
Time * Turkish * Preference	-0.42	0.67	-0.62	.536
Condition * Turkish * Preference	-0.26	0.98	-0.27	.788
Dutch * Turkish * Preference	-3.23	0.98	-3.31	.001
Time * Condition * Exposures	5.04	3.15	1.60	.110
Time * Dutch * Exposures	-3.22	2.46	-1.31	.191
Condition * Dutch * Exposures	1.89	3.09	0.61	.541
Time * Turkish * Exposures	-1.01	2.70	-0.37	.708
Condition * Turkish * Exposures	11.15	3.24	-3.45	.001
Dutch * Turkish * Exposures	8.35	4.28	1.95	.051
Time * Preference * Exposures	1.07	3.23	0.33	.740
Condition * Preference * Exposures	-2.17	3.69	-0.59	.556
Dutch * Preference * Exposures	3.85	3.35	1.15	.250
Turkish * Preference * Exposures	-5.34	3.34	-1.60	.110
Time * Condition * Dutch * Turkish	-0.87	1.32	-0.66	.507

(Continues)



**TABLE A3** (Continued)

	$\beta$	SE	z	p
Time * Condition * Dutch * Preference	2.48	1.51	1.65	.100
Time * Condition * Turkish * Preference	-2.23	1.35	-1.66	.098
Time * Dutch * Turkish * Preference	1.99	1.25	1.59	.113
Condition * Dutch * Turkish * Preference	-0.47	1.85	-0.25	.799
Time * Condition * Dutch * Exposures	5.30	5.21	1.02	.309
Time * Condition * Turkish * Exposures	2.83	5.30	0.53	.594
Time * Dutch * Turkish * Exposures	4.92	7.10	0.69	.488
Condition * Dutch * Turkish * Exposures	3.91	8.25	0.47	.636
Time * Condition * Preference * Exposures	7.13	6.25	1.14	.254
Time * Dutch * Preference * Exposures	-8.86	4.97	-1.78	.075
Condition * Dutch * Preference * Exposures	5.12	6.18	0.83	.408
Time * Turkish * Preference * Exposures	3.20	5.39	0.59	.553
Condition * Turkish * Preference * Exposures	-3.52	6.57	-0.54	.592
Dutch * Turkish * Preference * Exposures	7.27	8.71	0.84	.403
Time * Condition * Dutch * Turkish * Preference	-0.29	2.61	-0.11	.911
Time * Condition * Dutch * Turkish * Exposures	13.81	3.25	-1.04	.297
Time * Condition * Dutch * Preference * Exposures	-2.78	0.48	-0.27	.791
Time * Condition * Turkish * Preference * Exposures	14.51	0.64	1.36	.173
Time * Dutch * Turkish * Preference * Exposures	7.75	4.24	0.54	.586
Condition * Dutch * Turkish * Preference * Exposures	52.17	6.16	3.23	.001
Time * Condition * Dutch * Turkish * Preference * Exposures	28.99	4.88	-1.17	.244

Note: As all continuous variables were rescaled,  $\beta$ -values are not in an interpretable scale either. To get sensible values, values for effects with one rescaled variable should be divided by 10, values for effects with two rescaled variables by 100, and values for effects with three rescaled variables by 1,000. This holds for all three reported models and their outcomes.