Computers & Education 157 (2020) 103970

Contents lists available at ScienceDirect

Computers & Education

journal homepage: http://www.elsevier.com/locate/compedu

Robot tutor and pupils' educational ability: Teaching the times tables

Elly A. Konijn[a,1], Johan F. Hoorn[a,b,2,*]

[a] VU University, Amsterdam, The Netherlands
[b] The Hong Kong Polytechnic University, Hong Kong SAR

ARTICLE INFO

Keywords:
Social robots
Robot tutor
Multiplication tables
Primary school
Tutoring
Capacity issues

ABSTRACT

Research shows promising results of educational robots in language and STEM tasks. In language, more research is available, occasionally in view of individual differences in pupils' educational ability levels, and learning seems to improve with more expressive robot behaviors. In STEM, variations in robots' behaviors have been examined with inconclusive results and never while systematically investigating how differences in educational abilities match with different robot behaviors. We applied an autonomously tutoring robot (without tablet, partly WOz) in a 2 × 2 experiment of social vs. neutral behavior in above-average vs. below-average schoolchildren ($N = 86$; age 8–10 years) while rehearsing the multiplication tables on a one-to-one basis. The standard school test showed that on average, pupils significantly improved their performance even after 3 occasions of 5-min exercises. Beyond-average pupils profited most from a robot tutor, whereas those below average in multiplication benefited more from a robot that showed neutral rather than more social behavior.

1. Introduction

For about a decade, social robots and teleconference robots have been successful in supporting teaching tasks. In various forms, humanoid robots have been potentially useful as teaching assistants in the classroom (e.g., Kennedy, Baxter, & Belpaeme, 2015; Conti, Di Nuovo, Buono, & Di Nuovo, 2017; Kory-Westlund & Breazeal, 2019). Provided that the educational task is limited, robots are capable of achieving learning outcomes and improving learning performance, up to the level of human educators - vide the meta-analysis by Belpaeme, Kennedy, Ramachandran, Scassellati, and Tanaka (2018a). In the affective domain as well, robots surprisingly achieve similar levels of appreciation as human teachers do Belpaeme et al. (2018a). These beneficial effects are not limited to country or culture and were acquired from different types of robots, educational methods, and content matter (Belpaeme et al., 2018a.

Most STEM-related artificial tutors are on-screen. For example, pedagogical agents were used to teach general science to middle school students in an immersive VR (Makransky, Wismer, & Mayer, 2019). Mathematics with a virtual agent successfully supported learning when it exhibited rapport behaviors (Krämer et al., 2016). An embodied agent helped to improve working memory and math

* Corresponding author. The Hong Kong Polytechnic University, Hong Kong SAR.
E-mail address: johan.f.hoorn@polyu.edu.hk (J.F. Hoorn).

[1] Elly A. Konijn, Ph.D., Department of Communication Science, Media Psychology Program, Vrije Universiteit Amsterdam.
[2] Johan F. Hoorn, Ph.D.mult., Department of Computing and School of Design, The Hong Kong Polytechnic University & Department of Communication Science, Vrije Universiteit Amsterdam.

https://doi.org/10.1016/j.compedu.2020.103970
Received 6 December 2019; Received in revised form 6 May 2020; Accepted 28 June 2020
Available online 16 July 2020
0360-1315/

fluency (the speed to retrieve or calculate answers) while doing basic math operations (Arroyo, Royer, & Park Woolf, 2011). In all, studies with actual physical robots are scarce when it comes to teaching STEM. Challenges and successes in engaging students with general STEM through robots are reported (Gomoll et al., 2018) yet, not with robots as tutors but by learning how to build a robot.

The advantage of physically present humanoid robots over other e-learning devices is their physical presence or embodiment. Research showed that people followed instructions more closely from an embodied robot than from a computer tablet with the same software and voice (e.g., Mann, MacDonald, Kuo, Li, & Broadbent, 2015). Significant learning gains were shown with robot tutoring compared to control groups without a robot or with a computer tablet (Brown, Kerwin, & Howard, 2013; VanLehn, 2011). The advantage of robots over human teachers is that they are patient in rehearsing, never annoyed, and address each pupil the same (i.e., unbiased). Their disadvantage lies in technical (maintenance) issues, not being an educational 'all-rounder,' quite limited in skills, and not capable of keeping discipline in the classroom.

One of the best researched educational areas with robot tutors thus far is teaching (second) language. For example, Belpaeme et al. (2018b) conducted a range of studies into children acquiring a second language through robot tutoring (also Belpaeme et al., 2015; Kennedy, Baxter, Senft, & Belpaeme, 2016). From a viewpoint of developmental psychology, Belpaeme et al. (2018b) offer a number of design guidelines for second-language acquisition supported by robots, including the requirements on the technology. Kory-Westlund and co-workers did a range of studies with preschool children, learning language skills such as new vocabulary with toy-like story-telling robots, adapting the level of teaching to the child's abilities (e.g., Kory-Westlund & Breazeal, 2015, 2019, 2014; Kory-Westlund et al., 2017).

Another area of research into robot teaching is arithmetic and mathematics. For example, Highfield, Mulligan, and Hedberg (2008) studied children who were taught simple mathematics by exploring programmable robots. Janssen, Van der Wal, Neerincx, and Looije (2011) designed a game for long-term educational interaction in which a NAO robot taught pupils arithmetic by adapting the exercises according to the child's performance. Huang and Hoorn (2018) administered an experiment in which a Hanson's Einstein successfully taught simple mathematical equations to pupils of primary schools in the rural areas of mainland China.

The two most frequent forms of robot lecturing that are explored are robots as teachers and robots as tutors. The robot as a teacher is in front of the class, does group instruction, and more-or-less replaces the human teacher. For instance, Hashimoto, Kato, and Kobayashi (2011) used the SAYA life-like human-size robot as a replacement of a real teacher. The study by Huang and Hoorn (2018) also put the robot in front of a group to teach pupils mathematical equations (e.g., $2 = 2y \div 3$). Edwards, Edwards, Spence, Harris, and Gambino (2016) compared a robot as teacher vs a teacher represented in a telepresence robot in a college classroom. Both types of robotics were perceived as credible but the telepresence form was seen as *more* credible and kinder with better learning results.

The most frequently encountered form of robot teaching is the robot as a tutor, instructing the student individually (e.g., Ramachandran, Sebo, & Scassellati, 2019). For instance, Janssen et al. (2011) found that personal benchmarks in performance worked better than setting a group target. In teaching vocabulary during a storytelling game to pre-schoolers, furry DragonBots that accommodated the child's oral language skills had more success than robots that did not (Kory-Westlund & Breazeal, 2015). Pupils who received one-to-one tutoring significantly improved in learning outcomes compared to those in traditional classroom settings (on average two sigma's above the mean; cf. Bloom, 1984; Leyzberg, Spaulding, & Scassellati, 2014). Meta-analyses reported significant improvements of one-to-one tutoring, including computerized tutoring, although closer to one than two sigma's (VanLehn, 2011; Giannandrea; Sansoni, 2013). Baxter, Ashurst, Read, Kennedy, and Belpaeme (2017) did a longitudinal study into unsupervised personalized learning with a social robot. Positive results for new content matters radiated to performance on other topics. Acceptance also seemed higher compared to non-personalized interaction. Huang and Hoorn (2018) found that both boys and girls benefitted from robot instruction but that girls gained more from the robot than boys, who seemed to be more distracted by the technology.

Clearly, different teachers behave differently and some are more effective than others. An important differential factor is the teacher's behavior or communication style (Wayne & Youngs, 2003). Research shows that teachers who communicate in a more personalized and socially supportive way are more successful, which is generally supported for human teachers (VanLehn, 2011; Farrell, 2010; Skinner & Belmont, 1993). An important asset of robot tutors is that they can be programmed to communicate in different ways according to preferred styles. Therefore, this has been explored in research in robot tutoring on various aspects of behavior. For robots expressing different gestures, gazes, and behavioral expressions, mixed results were reported on learning outcomes (Leyzberg et al., 2014; Kennedy et al., 2015; Castellano et al., 2013; Kory-Westlund & Breazeal, 2014). For example, Saerbeck, Schut, Bartneck, and Janse (2010) and Kennedy et al. (2015) compared robots that behaved in more social versus neutral ways in a tutoring experiment but did not find the expected results. In all, mixed results have been obtained thus far regarding the social behavior of robot tutoring on learning outcomes (Belpaeme et al., 2018a). The current study aims to explain these mixed results by individual differences.

Individual differences among learners largely influence their progress and a robot perhaps should level its teaching to match the child's abilities (Kory-Westlund & Breazeal, 2015). A prominent predictor is that individuals differ in their educational or academic ability (e.g., Chamorro-Premuzic & Furhnam, 2006; Farrell, 2010) or in terms of attention deficits (e.g., ADHD, Rutter, 1978; Ferrara & Hill, 1980). For example, whereas pupils in general might benefit from socially supportive communication in human teachers, typically and atypically developing children may differ in this respect. Research shows that children with attention deficits and hyperactivity disorders are at increased risk for poor school performance, including poor reading and poor math standardized test scores (Loe & Feldman, 2007). They achieved substantially lower results on various school tasks in a large-sale comparison study, independent of socioeconomic background factors (Jangmo, Stålhandske, Chang, et al., 2019). Among the various symptoms and problems in school are that such children have low levels of attention and are easily distracted (Loe & Feldman, 2007), for example, by unexpected behaviors, sounds, or competitive cues, especially under low task demands (Friedman-Hill, Wagman; Gex, Pine, Leibenluft, & Ungerleider, 2010). Therefore, we reasoned that children with attention deficits in particular and lower achievers in general might *not* benefit from a socially behaving robot, following the research with human teachers, whereas they may benefit from the one-to-one tutoring with a robot which increases task demand. It is important to unravel how robot tutors might (not) support children differing in educational ability levels.

Thus, in the current study, we examined how social behavior of a robot tutor versus neutral behavior would (not) support children differing in their level of education thus far. We tested our assumptions in an experiment as research design to assess causal and possible interaction effects of the variations in robot behaviors in view of children's educational ability levels. The setting, however, was not a traditional lab but the experiment took place in a regular school where children could rehearse with the robot tutor in their own environment (cf. remedial teaching). A school large enough for the number of same-aged pupils needed, willingly and fully cooperated as further detailed below.

For various reasons, we chose tutoring the multiplication tables to elementary school children as the study's educational focus because: 1) Robots are particularly suited for repetitive tasks such as rehearsing multiplication tables at the primary school level and, to our knowledge, this is the first study focusing on robot tutoring of multiplication tables; 2) Individual differences among elementary school children in math and mastery of the times tables are clearly pronounced, with longer lasting implications, and can unequivocally be assessed through objective tests (Entwisle, Alexander, & Olson, 1994); 3) Speech recognition in humanoid robots is still limited, in particular for certain national languages. With rehearsing the times tables not much verbal language is needed and the task is relatively simple in interaction with a social robot. This allows actual one-to-one tutoring whereas most language tutoring with robots depend on accompanying tablets which seriously limits interaction (Van den Berghe et al., 2019; Belpaeme et al., 2018a); Thus, 4) to create optimal study conditions in view of our aim, we considered it important that the child would interact on a one-to-one basis and autonomously with the robot (i.e., without an additional tablet).

2. Method

2.1. Participants, design, and measures

Participants were primary school pupils aged 8–10 years old ($N = 86$; $M_{age} = 8.76$, $SD = 0.87$; 45 boys), who participated voluntarily after having received active consent from their parents. Pupils were randomly assigned in a mixed factorial design with the robot's social behavior (social vs. neutral) as a between factor and task performance as a within factor (i.e. pre-vs posttest learning outcome in terms of the test score). Regarding the individual educational ability differences, which was a second between factor, we used the pre-test scores (cf. Kory-Westlund & Breazeal, 2014) to indicate half of the participants as generally below average (i.e., below average on test scores) vs. 'advanced' or beyond average in their achievements calculating multiplication times tables.

Task performance was measured through the school's standard tests, the so-called tempo-test in the classroom which was done by the teachers as a normal test in the classroom, both a pre- and posttest. The tempo-test comprised of completing 140 multiplications, randomly chosen from tab 1–10, within 4 min with the actual number of correct answers as result. The pre-test took place approximately one week before the actual tutoring started and the post-test was administered in the week after the final tutoring with the robot (i.e., depending on the day the child interacted with the robot, the class-wise post-test was a few days to max a week later for the individual child). Over the course of three weeks, each child interacted three times, once a week, with the robot in a similar way, based on its condition (i.e. either a social or neutral robot). From the test scores, we calculated the mean difference score as the post-test score minus the pre-test score to assess task performance level as the dependent variable in the statistical analyses.

2.2. Materials and procedure

We used a NAO robot, which children generally experience as a smart, non-threatening educational support (Nalin, Bergamini, Giusti, Baroni, & Sanna, 2011; Shamsuddin et al., 2012). NAO is a 58 cm tall humanoid (Gouaillier et al., 2009) with two MT9M114 video cameras of the type SOC Image Sensor in its forehead, offering up to 1280 × 960 output resolution at 30fps with 2.24V/Lux-sec (550 nm) responsivity. The top camera was used to identify the children's faces in the visual field of view (72.6°DFOV (60.9°HFOV, 47.6°VFOV), 30 cm ~ infinity). Face detection was conducted with ALFaceDetection, NAO's vision module, providing face position based on the angular coordinates of facial features such as eyes, eyebrows, nose, and mouth (~45–~160 pixels in a QVGA image), working sufficiently well with standard office lighting (100–500 lux).

For interaction purposes, NAO was installed with Zorabots QBMT's Zora software[3] and programmed to tutor the times tables autonomously in direct one-to-one interaction with the child. A research assistant followed the tutoring through a webcam from an adjacent room, and interrupted through WOz when needed, for example, in case of speech recognition issues.

Based on a random assignment procedure, a detailed data collection scheme was developed when to ask which child out of the classroom. Each child was individually invited to sit with the robot alone for a 5-min session rehearsing the times tables with the robot, repeated three times over the course of three weeks (i.e., one time a week). Pupils only trained the difficult tables 6–9 with the robot on a one-to-one basis (i.e., one pupil-to-robot) and in random order to optimize the training time. In that sense, our set up resembled remedial teaching rather than instruction on a new topic.

The NAO-robot had been introduced in class beforehand to get the children used to it and to allow them to ask any question they

[3] A NAO robot (Aldebaran/Softbank) is named Zora as it is equipped with Zora-software, programmed by Zorabots/QBMT, Belgium: https://www.ald.softbankrobotics.com/en/zora-bots.

had. Furthermore, we could check for children who might not feel comfortable with a robot. The individual meetings with the robot took place in a separate small room at school. During their daily teaching program, the pupils were picked up one by one from the classroom to take part in the one-to-one tutoring session with the robot. A desk and a chair were placed in the test room. The robot was on the desk (Fig. 1). The pupils were asked to sit in the chair or stand in front of the robot. After the task was explained and understood by the pupil, the research assistant left the room and closed the door. Then, the one-to-one tutoring session with the robot started.

In the more social condition ($n = 45$), the robot tutor showed behavior that is indicated as socially supportive in the literature such as verbal encouragement (Brown & Howard, 2014), personalized speech and appropriate gestures (Kennedy et al., 2015), and expression of enthusiasm (Saerbeck et al., 2010). Therefore, the robot greeted the pupil by saying the pupil's name. From the first moment on, it made eye contact and followed the gaze. In addition, its eyes blinked. During the whole session, the robot gestured while talking. In the case of a correct answer, it clapped his hands at times, cheered "Fantastic" or "Well done." In the case of an incorrect answer from the pupil, its feedback was "That is not correct, we will try again" or "Too bad, that is wrong, let's try again." At the end of the session, the robot thanked the student and applauded.

In the less-social, neutral condition ($n = 41$), the robot tutor did not show the aforementioned socially supportive behavior. It did not greet the student in the beginning and just started with "We are going to practice several multiplication tables." It did not follow the pupil's face and its eyes did not blink. During the session, the robot did not make hand gestures and the body was in a fixed position. In the case of a correct answer by the student, it replied with "correct" or "right" as feedback, and some variations thereof, and in the case of an incorrect answer variations of "wrong" or "incorrect." At the end of the session the robot said "That was it for today. Bye," indicating that the session was over.

Each pupil participated in three one-to-one tutoring sessions with the robot. During each of these sessions, the robot randomly asked the pupil 20 multiplications from the tables of 6, 7, 8 and 9. For each multiplication, the pupil was given two chances for a good answer. After two failed attempts, the robot gave the correct answer. Pupils were also allowed to say "I do not know" (or push the button on the robot's right foot). Then, the robot gave the correct answer and continued to the next multiplication. After all 20 multiplications were done, the robot repeated the number of multiplications the pupil had answered correctly.

Thus, the robot's social behavior included personalized greeting, eye gaze, eye blinking, gestures, and behaviors when (in)correct answers were given. The neutral communication style avoided such behaviors and kept communication limited and strictly focused on the task at hand (e.g., "We will now rehearse times tabs 6 and 7"; "Correct"; "Try again"). Feedback delay was about 1 s in both conditions. The research assistant observed the pupils through a webcam.

3. Results

Mean difference scores were calculated from the correctly answered multiplications. Results of a paired-samples *t*-test showed that the pupils' performance on the multiplication tables significantly increased in their mean difference score (post-test minus pre-test scores), $t_{(85)} = 2.76$, $p < .007$, from $M = 55.73$ ($SD = 31.80$) on the pre-test to $M = 62.12$ ($SD = 36.4$) on the post-test, for the group as a whole regardless of conditions. Mean difference score was $\Delta M = 6.40$, $SD = 21.44$. In an independent samples *t*-test, the

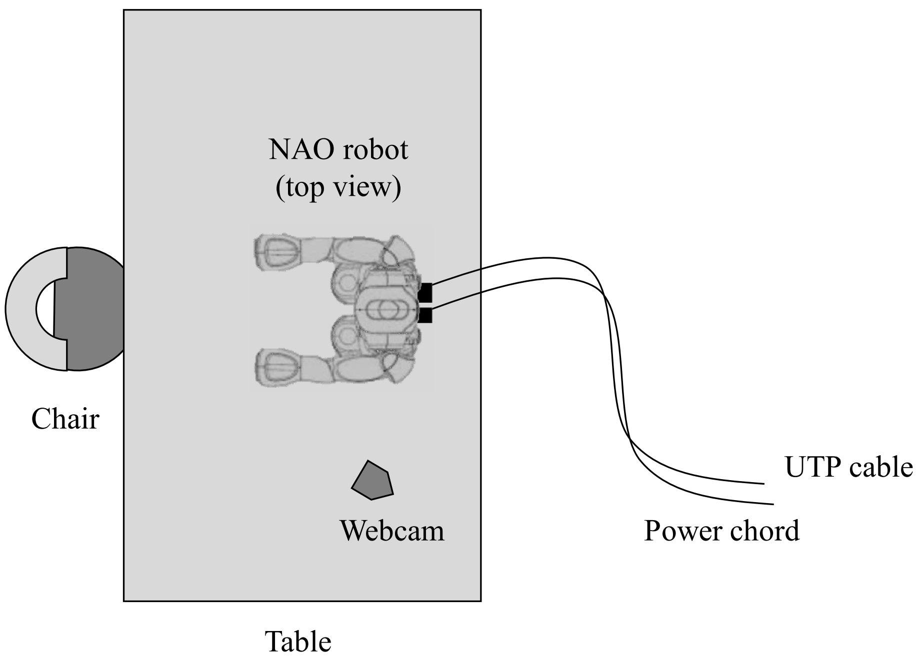

Fig. 1. Experiment-room layout.

effect of Gender on the difference score (learning gain) was not significant: $t_{(84)} = -0.22$, $p = .825$.

Results of a 2 (robot behavior) × 2 (individual ability) ANOVA did not show significant differences in the mean difference score (post-test minus pre-test) between the more social and the neutral behavior of the robot, $F_{(1,82)} = 0.03$, $p = .87$. We found no significant difference between individual ability levels, $F_{(1,82)} = 1.91$, $p > .17$, and no significant interaction, $F_{(1,82)} = 0.09$, $p > .77$.

Based on the difference scores, we additionally created four groups for level of advancement: $\Delta M <$ above average $< \Delta M + 1SD$ ($n = 19$), $\Delta M >$ below average $> \Delta M – 1SD$ ($n = 46$), advanced $> \Delta M + 1SD$ ($n = 12$), and challenged $< \Delta M – 1SD$ ($n = 9$). Then we ran a 2 (robot behavior) × 4 (advancement) ANOVA on the mean difference scores but the interaction was not significant: $F_{(3,78)} = 2.55$, $p > .05$. The significant effect of level of advancement was a trivial finding, obviously.

Detailed inspection of data and further analyses showed that overall, from pre-test to post-test, the advanced pupils (as indicated beforehand based on the tempo-test scores) profited most from tutoring with a robot, $t_{(44)} = -2.53$, $p = .015$. Separate analyses *within* the factor robot behavior further showed a significant improvement for the neutral robot overall, $t_{(40)} = 2.15$, $p = .038$. However, a separate analysis within educational ability level showed that the *lower* ability pupils profited more from a robot tutoring in a neutral way than from a more socially acting robot, $t_{(25)} = -2.81$, $p = .010$ (see Table 1 for details and Fig. 2 for visual illustration).

4. Discussion

In this study, we wanted to see if robots are capable of sorting a positive learning effect and so they did. A social robot performed a remedial teaching task by autonomously tutoring school children hard-to-remember multiplication tables without the support of a tablet. Inspired by language-related studies (e.g., Kory-Westlund et al., 2017), we wondered if a robot's more social behavior adds to learning a STEM-oriented task such as rehearsing multiplications, which it did not. Note that we were not teasing out a relationship per se but simply looked if increasing the number of social cues was helpful in remembering the multiplications.

In line with previous research, our results support the idea of a humanoid robot tutor as an effective support tool in education. The current study assessed the robot's either social or neutral (i.e. less social) behavior in coalescence with the educational ability level of its pupils. In contrast to expectations from the pedagogical literature on human teachers, the neutral robot seemed more effective overall. On average, the more advanced pupils tended to benefit most from robot tutoring, the under-averaged pupils tended to gain most from tutoring with a neutral, not specifically socially behaving robot. Perhaps, the robot's social behavior was distracting for them.

In a systematic literature review, expressive robots were beneficial for learning vocabulary and oral language skills (Hein & Nathan-Roberts, 2018). Expressive robots seemed to be better suited for communicative tasks than technical skills such as learning grammar (Hein & Nathan-Roberts, 2018) although expressive robots might enhance a pupil's interest in and motivation for learning language (Hein & Nathan-Roberts, 2018). Perhaps this is why for technical skills such as maths and arithmetic, less expressive robots function just as well. For instance, a robot that taught basic maths to school children introduced itself, welcomed the child, and gave compliments such as "You're doing great!" and "You're the best!" (Hindriks & Liebens, 2019). Yet, this kind of feedback did not exert any learning or motivational effects on task performance, affect, or interaction in maths (Hindriks & Liebens, 2019).

Our findings also are in line with the literature regarding poor school performance in children with attention deficits (Friedman-Hill et al., 2010; Jangmo et al., 2019; Loe and Feldman, 2007). Likewise, Kennedy et al. (2015) did not find the expected effects of a socially behaving robot and discussed, but did not study, that it might have been distracting for the pupils in general. Our results add to that and may explain the thus far mixed results found for socially behaving robots in education (Belpaeme et al., 2018a). Our results suggest that the robot's social behavior seemed distracting in particular for the subsample of under-averaged pupils. Thus, children with attention deficits and those with under-average school performance who are easily distracted may benefit from a more neutral, task-focused communication style.

Thus far, many studies with educational robots focused on storytelling and language acquisition, primarily in toddlers and preschoolers, often with 'toy-like' robots in combination with tablets (Kennedy et al., 2016; Kory-Westlund et al., 2017; Kory-Westlund & Breazeal, 2014; Fridin, 2014). Importantly, the robot is then often accompanied by a tablet to compensate for limited speech technology in current humanoid robots (Van den Berghe, Verhagen, Oudgenoeg-Paz, Van der Ven, & Leseman, 2019). It is concluded that the tablet may actually distract from interacting with the robot itself and undermine its possible potential (Konijn, Smakman, & Van

Table 1
Mean difference scores on multiplication test per condition.

Robot behavior	Pupil's ability level	*M*	*SD*	*N*
Neutral	Beyond average	9.52	22.83	25
	Below average	4.38	22.09	16
	Total	7.51	22.41	41
Social	Beyond average	10.17	25.53	18
	Below average	2.18	16.54	27
	Total	5.38	20.72	45
Total	Beyond average	9.79	23.70	43
	Below average	3.00	18.57	43
	Total	6.40	21.44	86

Note. *M* = mean difference scores, *SD* = standard deviation. Groups organized by robot's behavior (social vs. neutral) and individual educational ability levels (beyond average vs below average).

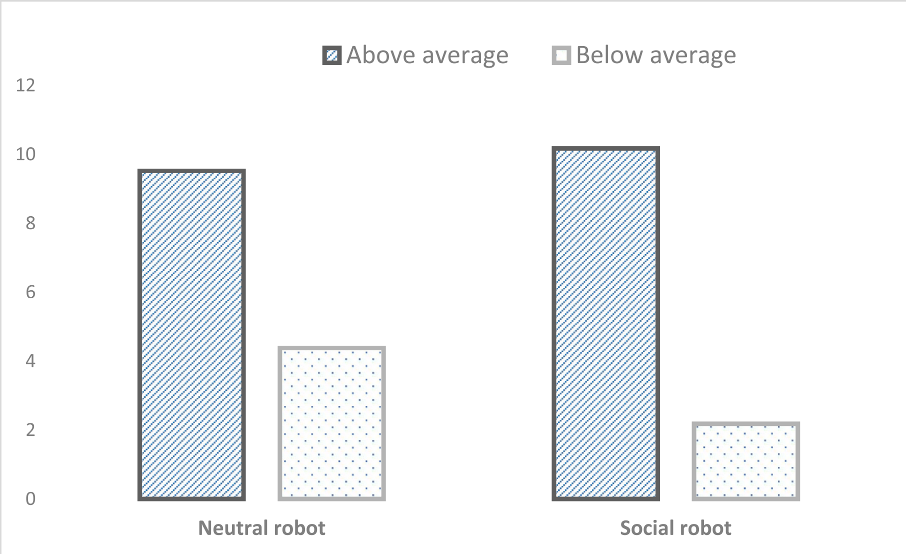

Fig. 2. Pupils' mean difference scores on the multiplication test per condition, indicating post-test improvements, organized by the robot's behavior (left: neutral; right: social). Filling of bars indicate pupils' educational ability level, according to their test scores (dark fill: 'above average'; light fill: 'below average').

den Berghe, 2020).

With regards to STEM teaching, studies commonly employ virtual on-screen tutors (e.g., Makransky, Wismer, & Mayer, 2019) but for teaching mathematics or basic arithmetic, few studies are available (notable exceptions are Krämer et al., 2016 and Arroyo et al., 2011). Studies that use physically present robots to teach mathematics are even scarcer.

Our study looked into robot tutoring in primary school children rehearsing the times tables and systematically compared individual educational ability differences in view of different communication styles of a humanlike robot tutor, interacting largely autonomously on a one-to-one basis. Results showed a significant improvement of learning outcomes with only very little training with the robot tutor. Noting that the training consisted of merely 3 one-weekly sessions of 5 min each, the significant improvement on the standard multiplication 'tempo-test' is remarkable, both among the 'advanced' pupils and those who were below average. In all, these findings support largely that receiving one-to-one tutoring significantly improves learning outcomes (Bloom, 1984), even with a robot. This is in line with previous research comparing various robot- and avatar-tutoring contexts (Leyzberg et al., 2014; Giannandrea; Sansoni, 2013; VanLehn, 2011). Our results add that it matters to match the pupil's proficiency to improve learning outcomes. For future research, it is important to make a distinction between different types of tasks such as acquisition of new materials or, as in the current study, improvement of tasks that are already known but not yet fully mastered.

The current study showed the effectiveness of social robots as autonomous tutors in one-to-one instruction, rehearsing the times tables while varying their communication styles with differences in educational ability. Although pupils in general appear to benefit from socially supportive behavior in learning from humans, when it comes to robot tutors communicating in a socially supportive way (e.g., speech supportive gestures, eye gaze, personalized greeting), thus far, mixed results were found on learning outcomes (Leyzberg et al., 2014; Castellano et al., 2013; Kennedy et al., 2015; Kory-Westlund & Breazeal, 2014). In our case, one might even wonder if continuously following the child's face was socially appropriate.

Our findings present a possible explanation in that the more advanced pupils seem to benefit from instruction that is more varied. They welcome some 'distraction' while repetitively rehearsing something that is not difficult to them. Thus, studies finding positive results of a socially behaving robot tutor might have sampled advanced or higher ability pupils or applied tasks that were relatively easy to them.

In contrast, the lower ability pupils may need stricter tutoring without too much of a distraction (i.e., a neutral robot just rehearsing the exercises). This aligns with the psychological and pedagogical literature discussing poor school performance in children with attention deficits (Friedman-Hill et al., 2010; Jangmo et al., 2019; Loe & Feldman, 2007). A key strength of robots is that they can adapt to the individual child in ways that is hard to ask from human teachers, specifically when it comes to behavioral and facial expressions. Clearly, this hypothesis needs to be further substantiated in future research.

In a similar vein, Huang and Hoorn (2018) found that boys were distracted by the robot and performed even better without. Note, however, that the reported studies addressed different arithmetical and mathematical topics (i.e. learning prime numbers, multiplication, solving equations). Perhaps that learning languages is a more socially oriented effort. In all, both content of robot tutoring and individual differences need to be taken into account in future research.

4.1. Limitations

The sample size for this type of research was relatively large, in particular in view of related robot tutoring research. Nevertheless, a limitation to the study is noted in view of the relatively large variations among the pupils (i.e. large *SD*s), which makes it difficult for statistical testing. In this regard, subsamples were too small to reach solid conclusions regarding a possible match between the social behavior of the robot and individual ability levels. Additional data collection is needed and in longer term perspectives. For future research, testing a lower grade where the children are just beginning with the multiplication tables and so start at a lower level might further our insights in a robot tutor for arithmetic tasks. Furthermore, the robot should have more technical sophistication to create a clearer distinction between a more social (e.g., speech supportive behavior) and less social or neutral robot (cf. O'Keefe, 2003). In addition, it is recommendable to add a no-robot baseline condition, take measures to deal with potential test threat, and examine possible mediating states such as improved social support, trust, credibility, or distraction to validate that children in the social condition had better rapport with the robot or, as suggested, might have been distracted.

To avoid bias in test results and increase ecologically validity, we used the standardized tempo test as common in the Dutch primary school system. This test assesses the total performance on all tables 1–10 and not solely the improvement in the 6–9 tables that were trained with the robot. Therefore, strictly speaking, children could have improved in the 1–5 and 10 tables and not so in the 6–9 tables. Although we believe that any improvement in multiplication should be welcomed, and the pre- and post-test was the same, we cannot conclude that precisely the difficult tables had improved the children's performance. However, if training the difficult tables did not matter, pupils should have shown about the same scores before-and-after robot tutoring as based on the tables they already did well. With the robot training of the difficult tables alone, overall performance improved, which is useful in its own right. For future research, it might be an interesting suggestion to look more closely into the improvements on the single tables that have been specifically trained. On that note, it may also be worthwhile to let the robot rehearse tables 6 and 8 to a random subset of children and 7 and 9 to another random subset and see where the robot tutoring was most effective.

While in general, robots seem to be effective teaching supports, what are the pedagogical and societal implications if a robot is actually integrated in educational practice? Questions arise about the profession of the teacher (Reich-Stiebert, Eyssel, 2016) as well as how introducing robots in the classroom may affect the public debate and classroom dynamics. The prevalent adagio of robots as surrogates (cf. Reich-Stiebert, Eyssel, 2016) seems to become gradually replaced by enthusiasm for robots – at least by the children. This instigates the societal debate on how to implement robots in education best, to design them in responsible ways and to define the social role the robot will take in the educational system. What teaching role will work best for which tasks: tutor, tutee, peer (Chen, Park Hae, & Breazeal, 2020)? How could the robot be employed in peer-learning (e.g., Kory-Westlund & Breazeal, 2019), learning-by-teaching, or as collaborative learning companion?

The physical embodiment of robots adds to their value compared to other technologies. Importantly, we also have to look at the disadvantages. For example, Smakman and Konijn (2019) review the moral and ethical considerations that come with the implementation of robot tutors in education. In all, many parties and varying perspectives are at stake and should be taken into account to best define the social role a robot may take in the educational system.

Author credit

Both authors are responsible for the reported research and both have participated in the concept and design of the theory, drafting and revising the manuscript. Both approved the manuscript as submitted.

Decalartion of competing interest

The authors have no competing interests to declare.

Acknowledgments

We are extremely grateful to the director, teachers, parents, and pupils of primary school "De Ontdekking" (Almere, NL), who enthusiastically supported our research by full cooperation. We thank our students Pouya Zarchin and Ruben Schneider for their indispensable help in data collection. We kindly acknowledge ZoraBots, QBMT, Belgium, in making available their robots and realizing the software used in this study complimentary. We would like to thank the anonymous reviewer for commenting on an earlier draft of this paper. Correspondence concerning this article can be addressed to the first and second author: elly.konijn@vu.nl; csjfhoorn@comp.polyu.edu.hk; j.f.hoorn@vu.nl.

References

Arroyo, I., Royer, J. M., & Park Woolf, B. (2011). Using an intelligent tutor and math fluency training to improve math performance. *International Journal of Artificial Intelligence in Education, 21*(1–2), 135–152. https://doi.org/10.3233/JAI-2011-020.

Baxter, P., Ashurst, E., Read, R., Kennedy, J., & Belpaeme, T. (2017). Robot education peers in a situated primary school study: Personalisation promotes child learning. *PloS One, 12*(5), Article e0178126.

Belpaeme, T., Kennedy, J., Baxter, P., Vogt, P., Krahmer, E. E., Kopp, S., & Pandey, A. K. (2015). L2TOR-second language tutoring using social robots. In *Proceedings of the ICSR 2015 WONDER workshop*.

Belpaeme, T., Kennedy, J., Ramachandran, A., Scassellati, B., & Tanaka, F. (2018a). Social robots for education: A review. eaat5954 *Science Robotics, 3*(21). Available from https://www.researchgate.net/profile/Tony_Belpaeme/publication/327046544_Social_robots_for_education_A_review/links/5bef5602299bf1124fd82287/Social-robots-for-education-A-review.pdf.

Belpaeme, T., Vogt, P., Van den Berghe, R., Bergmann, K., Göksun, T., De Haas, M., et al. (2018b). 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.

Bloom, B. S. (1984). The 2 sigma problem: The search for methods of group instruction as effective as one-to-one tutoring. *Educational Researcher, 13*(6), 4–16.

Brown, L. N., & Howard, A. M. (2014). The positive effects of verbal encouragement in mathematics education using a social robot. In *Integrated STEM education conference (ISEC)* (pp. 1–5). 2014 IEEE.

October Brown, L., Kerwin, R., & Howard, A. M. (2013). Applying behavioral strategies for student engagement using a robotic educational agent. In *Proceedings of the 2013 IEEE international conference on systems, man, and cybernetics* (pp. 4360–4365). https://doi.org/10.1109/SMC.2013.744.

Castellano, G., Paiva, A., Kappas, A., Aylett, R., Hastie, H., Barendregt, W., et al. (2013). Towards empathic virtual and robotic tutors. In *Artificial intelligence in education* (Vol. 7926, pp. 733–736). Berlin, Heidelberg: Springer. https://doi.org/10.1007/978-3-642-39112-5_100.

Chamorro-Premuzic, T., & Furhnam, A. (2006). Intellectual competence and the intelligent personality: A third way in differential psychology. *Review of General Psychology, 10*(3), 251–267. https://doi.org/10.1037/1089-2680.10.3.251.

Chen, H., Park Hae, W., & Breazeal, C. L. (2020). Teaching and learning with children: Impact of reciprocal peer learning with a social robot on children's learning and emotive engagement. *Computers & Education, 150*, 103836. https://doi.org/10.1016/j.compedu.2020.103836.

Conti, D., Di Nuovo, S., Buono, S., & Di Nuovo, A. (2017). Robots in education and care of children with developmental disabilities: A study on acceptance by experienced and future professionals. *International Journal of Social Robotics, 9*(1), 51–62. doi: 10.100712369-016-0359-6.

Edwards, A., Edwards, A., Spence, P. R., Harris, C., & Gambino, A. (2016). Robots in the classroom: Differences in students' perceptions of credibility and learning between "teacher as robot" and "robot as teacher". *Computers in Human Behavior, 65*, 627–634. https://doi.org/10.1016/j.chb.2016.06.005.

Entwisle, D., Alexander, K., & Olson, L. (1994). The gender gap in math: Its possible origins in neighborhood effects. *American Sociological Review, 59*(6), 822–838. http://www.jstor.org/stable/2096370.

Farrell, P. (2010). School psychology: Learning lessons from history and moving forward. *School Psychology International, 31*(6), 581–598.

Ferrara, C., & Hill, S. D. (1980). The responsiveness of autistic children to the predictability of social and nonsocial toys. *Journal of Autism and Developmental Disorders, 10*(1), 51–57.

Fridin, M. (2014). Storytelling by a kindergarten social assistive robot: A tool for constructive learning in preschool education. *Computers & Education, 70*, 53–64.

Friedman-Hill, S. R., Wagman, M. R., Gex, S. E., Pine, D. S., Leibenluft, E., & Ungerleider, L. G. (2010). What does distractibility in ADHD reveal about mechanisms for top-down attentional control? *Cognition, 115*(1), 93–103. https://doi.org/10.1016/j.cognition.2009.11.013.

Giannandrea, L., & Sansoni, M. (2013). A literature review on intelligent tutoring systems and on student profiling. *Learning & Teaching with Media & Technology, 287*.

Gomoll, A., Šabanović, S., Tolar, E., Hmelo-Silver, C. E., Francisco, M., & Lawlor, O. (2018). Between the social and the technical: Negotiation of human-centered robotics design in a middle school classroom. *International Journal of Social Robotics, 10*(3), 309–324.

May Gouaillier, D., Hugel, V., Blazevic, P., Kilner, C., Monceaux, J. O., Lafourcade, P., & Maisonnier, B. (2009). Mechatronic design of NAO humanoid. In *IEEE international conference on robotics and automation, 2009 (ICRA'09)* (pp. 769–774). IEEE. https://doi.org/10.1109/ROBOT.2009.5152516.

Hashimoto, T., Kato, N., & Kobayashi, H. (2011). Development of educational system with the android robot SAYA and evaluation. *International Journal of Advanced Robotic Systems, 8*(3), 51–61.

Hein, M., & Nathan-Roberts, D. (2018). Socially interactive robots can teach young students language skills; a systematic review. *Proceedings of the Human Factors and Ergonomics Society - Annual Meeting, 62*(1), 1083–1087. https://doi.org/10.1177/1541931218621249.

Highfield, K., Mulligan, J., & Hedberg, J. (2008). Early mathematics learning through exploration with programable toys. In *Proceedings of the joint conference on psychology and mathematics* (Vol. 2008, pp. 17–21).

Hindriks, K. V., & Liebens, S. (2019). A robot math tutor that gives feedback. In M. Salichs, & et al (Eds.), *Social robotics. ICSR 2019. Lecture notes in computer science* (Vol. 11876, pp. 601–610). Cham, CH: Springer. https://doi.org/10.1007/978-3-030-35888-4_56.

Huang, I. S., & Hoorn, J. F. (2018). Having an Einstein in class. Teaching maths with robots is different for boys and girls. July 4-8, 2018. In X. Wang, Z. Wang, J. Wu, & L. Wang (Eds.), *Proceedings of the 13th world congress on intelligent control and automation (WCICA 2018)* (pp. 424–427). Changsha, China: New York: IEEE. https://doi.org/10.1109/WCICA.2018.8630584. Available from https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=8630584.

Jangmo, A., Stålhandske, A., Chang, Z., Chen, Q., Almqvist, C., Feldman, I., et al. (2019). Attention-Deficit/Hyperactivity Disorder, school performance, and effect of medication. *Journal of the American Academy of Child & Adolescent Psychiatry, 58*(4), 423–432. https://doi.org/10.1016/j.jaac.2018.11.014.

Janssen, J. B., Van der Wal, C. C., Neerincx, M. A., & Looije, R. (2011). Motivating children to learn arithmetic with an adaptive robot game. In B. Mutlu, C. Bartneck, J. Ham, V. Evers, & T. Kanda (Eds.), *Social robotics. ICSR 2011. Lecture notes in computer science* (Vol. 7072). Berlin, Heidelberg: Springer. https://doi.org/10.1007/978-3-642-25504-5_16.

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* (pp. 67–74). https://doi.org/10.1145/2696454.2696457. ACM.

Kennedy, J., Baxter, P., Senft, E., & Belpaeme, T. (2016). Social robot tutoring for child second language learning. In *11th ACM/IEEE international conference on human-robot interaction (HRI)* (Vol. 2016, pp. 231–238). IEEE. https://doi.org/10.1109/HRI.2016.7451757. March.

Konijn, E. A., Smakman, M., & Van den Berghe, R. (2020). Use of robots in education. In J. Van den Bulck, E. Sharrer, D. Ewoldsen, & M.-L. Mares (Eds.), *The international encyclopedia of Media psychology*. New York: Wiley.

Kory-Westlund, J. M., & Breazeal, C. L. (2014). Storytelling with robots: Learning companions for preschool children's language development. In *23rd IEEE international symposium on robot and human interactive communication* (Vol. 2014, pp. 643–648). RO-MAN), Edinburgh: IEEE. https://doi.org/10.1109/roman.2014.6926325.

Kory-Westlund, J. M., & Breazeal, C. L. (2015). The interplay of robot language level with children's language learning during storytelling. In J. A. Adams, W. Smart, B. Mutlu, & L. Takayama (Eds.), *Proceedings of the 10th annual ACM/IEEE international conference on human-robot interaction: Extended abstracts* (pp. 65–66).

Kory-Westlund, J. M., & Breazeal, C. L. (2019). A long-term study of young children's rapport, social emulation, and language learning with a peer-like robot playmate in preschool. *Frontiers in Robotics and AI, 6*, 81. https://doi.org/10.3389/frobt.2019.00081.

Kory-Westlund, J. M., Jeong, S., Park Hae, W., Ronfard, S., Adhikari, A., Harris, P. L., et al. (2017). Flat versus expressive storytelling: Learning and retention of a robot's narrative. *Frontiers in Human Neuroscience, 11*, 295. https://doi.org/10.3389/fnhum.2017.00295.

Krämer, N. C., Karacora, B., Lucas, G., Dehghani, M., Rüther, G., & Gratch, J. (2016). Closing the gender gap in STEM with friendly male instructors? On the effects of rapport behavior and gender of a virtual agent in an instructional interaction. *Computers & Education, 99*, 1–13. https://doi.org/10.1016/j.compedu.2016.04.002.

Leyzberg, D., Spaulding, S., & Scassellati, B. (2014). Personalizing robot tutors to individuals' learning differences. In *Proceedings of the 2014 ACM/IEEE international conference on Human-robot interaction* (pp. 423–430). ACM. https://doi.org/10.1145/2559636.2559671.

Loe, I. M., & Feldman, H. M. (2007). Academic and educational outcomes of children with ADHD. *Journal of Pediatric Psychology, 32*(6), 643–654. https://doi.org/10.1093/jpepsy/jsl054.

Makransky, G., Wismer, P., & Mayer, R. E. (2019). A gender matching effect in learning with pedagogical agents in an immersive virtual reality science simulation. *Journal of Computer Assisted Learning, 35*(3), 349–358.

Mann, J. A., MacDonald, B. A., Kuo, I., Li, X., & Broadbent, E. (2015). People respond better to robots than computer tablets delivering healthcare instructions. *Computers in Human Behavior, 43*, 112–117.

Nalin, M., Bergamini, L., Giusti, A., Baroni, I., & Sanna, A. (2011). Children's perception of a robotic companion in a mildly constrained setting. In *IEEE/ACM human-robot interaction 2011 conference (robots with children workshop) proceedings*.

O'Keefe, D. J. (2003). Message properties, mediating states, and manipulation checks: Claims, evidence, and data analysis in experimental persuasive message effects research. *Communication Theory, 13*(3), 251–274. https://doi.org/10.1111/j.1468-2885.2003.tb00292.x.

Ramachandran, A., Sebo, S. S., & Scassellati, B. (2019). Personalized robot tutoring using the assistive tutor POMDP (AT-POMDP). *Association for the Advancement of Artificial Intelligence*.

Reich-Stiebert, N., & Eyssel, F. (2016). Robots in the classroom: What teachers think about teaching and learning with education robots. *Lecture Notes in Computer Science, 9979*, 671–680. https://doi.org/10.1007/978-3-319-47437-3_66.

Rutter, M. (1978). Diagnosis and definition of childhood autism. *Journal of Autism & Childhood Schizophrenia, 8*(2), 139–161. https://doi.org/10.1007/BF01537863.

Saerbeck, M., Schut, T., Bartneck, C., & Janse, M. D. (2010). Expressive robots in education: Varying the degree of social supportive behavior of a robotic tutor. In *Proceedings of the SIGCHI conference on human factors in computing systems* (pp. 1613–1622). Atlanta: ACM. April.

Shamsuddin, S., Yussof, H., Ismail, L., Hanapiah, F. A., Mohamed, S., Piah, H. A., et al. (2012). Initial response of autistic children in human-robot interaction therapy with humanoid robot NAO. In *8th international colloquium on signal processing and its applications (CSPA), 2012 IEEE* (pp. 188–193). IEEE. https://doi.org/10.1109/CSPA.2012.6194716.

Skinner, E. A., & Belmont, M. J. (1993). Motivation in the classroom: Reciprocal effects of teacher behavior and student engagement across the school year. *Journal of Educational Psychology, 85*(4), 571. https://doi.org/10.1037/0022-0663.85.4.571.

Smakman, M., & Konijn, E. A. (2019). Robots tutors: Welcome or ethically questionable? Paper presented at the. In *International conference on robotics in education* (pp. 10–12). Vienna: published in the Springer Series: Advances in Intelligent Systems and Computing. April, 2019.

Van den Berghe, R., Verhagen, J., Oudgenoeg-Paz, O., Van der Ven, S., & Leseman, P. (2019). Social robots for language learning: A review. *Review of Educational Research, 89*(2), 259–295. https://doi.org/10.3102/0034654318821286.

VanLehn, K. (2011). The relative effectiveness of human tutoring, intelligent tutoring systems, and other tutoring systems. *Educational Psychologist, 46*(4), 197–221.

Wayne, A., & Youngs, P. (2003). Teacher characteristics and student achievement gains: A review. *Review of Educational Research, 73*(1), 89.