

Document Version

Final published version

Licence

CC BY

Citation (APA)

de Winter, J. C. F., Dodou, D., Moorlag, F., & Broekens, J. (2026). Social robots: a meta-analysis of learning outcomes. *Frontiers In Robotics and AI*, 12, 1-16. Article 1735198. <https://doi.org/10.3389/frobt.2025.1735198>

Important note

To cite this publication, please use the final published version (if applicable). Please check the document version above.

Copyright

In case the licence states "Dutch Copyright Act (Article 25fa)", this publication was made available Green Open Access via the TU Delft Institutional Repository pursuant to Dutch Copyright Act (Article 25fa, the Taverne amendment). This provision does not affect copyright ownership. Unless copyright is transferred by contract or statute, it remains with the copyright holder.

Sharing and reuse

Other than for strictly personal use, it is not permitted to download, forward or 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 such as Creative Commons.

Takedown policy

Please contact us and provide details if you believe this document breaches copyrights. We will remove access to the work immediately and investigate your claim.



OPEN ACCESS

EDITED BY

Kostas Karpouzis,
Panteion University, Greece

REVIEWED BY

Erik Lagerstedt,
University of Gothenburg, Sweden
Huanghao Feng,
Changshu Institute of Technology, China

*CORRESPONDENCE

J. C. F. de Winter,
✉ j.c.f.dewinter@tudelft.nl

RECEIVED 29 October 2025

REVISED 16 December 2025

ACCEPTED 17 December 2025

PUBLISHED 23 March 2026

CITATION

de Winter JCF, Dodou D, Moorlag F and
Broekens J (2026) Social robots: a
meta-analysis of learning outcomes.
Front. Robot. AI 12:1735198.
doi: 10.3389/frobt.2025.1735198

COPYRIGHT

© 2026 de Winter, Dodou, Moorlag and
Broekens. This is an open-access article
distributed under the terms of the [Creative
Commons Attribution License \(CC BY\)](#). The
use, distribution or reproduction in other
forums is permitted, provided the original
author(s) and the copyright owner(s) are
credited and that the original publication in
this journal is cited, in accordance with
accepted academic practice. No use,
distribution or reproduction is permitted
which does not comply with these terms.

Social robots: a meta-analysis of learning outcomes

J. C. F. de Winter^{1*}, D. Dodou¹, F. Moorlag¹ and J. Broekens²

¹Faculty of Mechanical Engineering, Delft University of Technology, Delft, Netherlands, ²Leiden Institute of Advanced Computer Science, Leiden University, Leiden, Netherlands

Previous meta-analyses show that social robots aid learning but were often limited in scope or grouped diverse control conditions together. This meta-analysis examined learning outcomes, focusing on control condition type. We retrieved 146 studies (Google Scholar and reference searches) where a physical social robot was used for training cognitive skills, comprising 183 post-test effect sizes between the robot and the control group, and 372 pre-post effect sizes. Analysis of the 78 studies with control groups indicated that robots generally improved learning, most notably when compared to a no-training control group ($d = 0.75$). Comparing robots to human teachers yielded an overall positive effect ($d = 0.31$), although effect sizes varied widely. This variability was explained by the robot's role: robots in a co-teaching capacity showed a strong positive effect ($d = 0.88$), while robots replacing the teacher showed no benefit ($d = -0.06$). LLM-based sentiment analysis indicated that papers from outside Europe received higher positivity scores when describing the robots. We conclude that the effect size is influenced by the robot implementation and the control condition chosen.

KEYWORDS

cognitive skills, educational robotics, human-robot interaction, learning, social robot

Introduction

The integration of artificial intelligence, robotics, and educational sciences has led to the emergence of social robots within educational settings. Social robots, defined as autonomous or semi-autonomous systems capable of interacting with humans in accordance with established social norms (Belpaeme et al., 2018; Neumann, 2020), have been examined in various educational roles, including tutors, peer learners, teaching assistants, and tutees (Belpaeme et al., 2018; Woo et al., 2021). Applications span the instruction of second-language vocabulary, mathematics, and computational thinking, among others (e.g., Derakhshan et al., 2025; Ouyang and Xu, 2024; Papadopoulos et al., 2020).

Proponents of social robots argue that their advantages result from several attributes. A central hypothesis is that physical embodiment, i.e., the ability to gesture, direct gaze, and occupy physical space, enhances student attention and motivation (Donnermann et al., 2020; Lee et al., 2006; Li, 2015). Furthermore, the non-evaluative nature of social robots can mitigate error-related anxiety, creating a safe space for practice, and encouraging help-seeking behaviors (Howley et al., 2014; Wang et al., 2013). These benefits can be augmented through personalization, where robots adapt instructional tactics to individual learners (Baxter et al., 2017; Yang et al., 2024), and by employing novel pedagogical paradigms, such as casting the robot as a tutee to exploit the so-called protégé effect (Tanaka and Matsuzoe, 2012).

Despite the aforementioned optimism, several reservations exist. The literature includes various studies yielding equivocal outcomes, where robotic interventions do

not surpass less complex alternatives such as tablets or standard instructional approaches. For example, a study by Zhexenova et al. (2020) compared a robot-only tutor with a tablet-only application and a human-teacher condition for early literacy (letter learning) and found no significant differences in learning gains. Another factor is that the social attributes designed to augment learning may inadvertently introduce distractions. Kennedy et al. (2015b) reported that excessive social behaviors in a robot diminished its effectiveness relative to a nonsocial counterpart. Other studies indicate that engaging robotic elements can divert focus from primary content (Charpentier et al., 2022; Rosenthal-von der Pütten et al., 2016).

Social robots in educational applications also encounter technical limitations. AI elements, such as speech recognition, frequently exhibit inaccuracies in processing inputs from young children or non-native speakers, yielding user dissatisfaction (Banaeian and Gilanlioglu, 2021; Bhat et al., 2016; Liebens, 2019; Randall, 2019). Text-to-speech systems are often critiqued for lacking natural intonation, which may hinder comprehension and learning (Lee et al., 2010). Perceptual deficiencies, such as failures in user detection, compound these issues, often requiring Wizard-of-Oz experiments instead of an actual autonomous deployment of the robot (e.g., Kennedy et al., 2016b; Komatsubara et al., 2014). High costs, content development, and maintenance also present barriers (Papadopoulou et al., 2020). Certain psychological drawbacks further impede the assimilation of social robots. In particular, the novelty effect, although initially increasing motivation, typically fades quickly (Baxter et al., 2017; Kanda et al., 2007).

A source of the conflicting outcomes in the literature may be the robot's assigned pedagogical function and the chosen control condition. Comparisons against no-intervention controls typically confirm the robots' instructional viability compared to no training (Draper and Clayton, 1992; Kennedy et al., 2016b; Vogt et al., 2019). Although human tutors frequently elicit greater learning gains (Kennedy et al., 2016a; Zhanatkyzy et al., 2022), robots may excel in constrained contexts such as routine exercises (e.g., Chang and Hwang, 2024). For example, for syntax drills, it has been found that a robotic tutor surpassed human teachers, a result that was attributed to the robot's inexhaustible consistency (Iio et al., 2024). The paramount evaluation, however, involves contrasts with virtual interfaces such as tablets. Herein, substantiation of the incremental value of social robots remains contentious. Several studies found no learning benefits for embodied robots relative to screen interfaces (e.g., Verhelst et al., 2024; Vogt et al., 2019), while other studies assert that social robots outperform tablets (e.g., Konijn et al., 2022) or that tangible (touch/gesture) interactions are more effective than non-interactive virtual settings (Al Hakim et al., 2022). In summary, the selection of control conditions shapes the robot's conceptualization as a mere intervention (when compared to nothing), a human co-teacher or surrogate (when compared to a human teacher), or as a sophisticated technology (when compared to a virtual interface).

The present study: a new meta-analytic synthesis

The wide variation in experimental outcomes requires a new meta-analytic review to synthesize findings and identify

the factors that drive learning. Prior meta-analyses have consistently concluded that social robots have a positive overall effect on student learning, for example in language learning (Belpaeme et al., 2018; Derakhshan et al., 2025; Lee and Lee, 2022; Wu and Li, 2024). Some moderators have been identified; for example, effects appear largest for elementary-aged learners (Derakhshan et al., 2025) and elementary/secondary versus tertiary school learners (Lee and Lee, 2022; Wu and Li, 2024), and interventions lasting between 1 and 4 weeks produce larger effects than studies with shorter intervention time spans (Wu and Li, 2024).

However, these prior meta-analyses are constrained by limitations, which indicate the need for a more comprehensive and rigorous synthesis. A primary issue is the limited statistical power caused by the small number of included studies, which has prevented powerful moderator analyses. Previous meta-analyses also suffer from methodological inconsistencies, including the aggregation of diverse control conditions, the conflation of cognitive and affective outcomes (Sukardi et al., 2024; Wang K. et al., 2023; Wu and Li, 2024), the inclusion of non-physical chatbots under the "robot" umbrella (Wu and Li, 2024), and the inclusion of effect sizes without a non-robot control group (Derakhshan et al., 2025; Wu and Li, 2024).

The current meta-analysis includes a larger corpus of studies to address these shortcomings. Our primary goal is not to recalculate an overall effect size but to deconstruct it. We focus on disentangling the robot's effectiveness based on its pedagogical role, with the hypothesis that the distinction between a co-teaching assistant and a full teacher replacement is a primary source of the heterogeneity observed in the literature. Furthermore, this study introduces a novel sentiment analysis using a Large Language Model (LLM). This analysis measures "Positiveness" in research papers to investigate potential reporting biases and examines how this sentiment relates to reported effect sizes and geographical origin to explore whether a tendency toward spin may be distorting the field's conclusions.

Methods

Our literature search and reporting adhered to the PRISMA guidelines (Page et al., 2021), see supplementary material. Literature searches were conducted using Google Scholar (latest search date: 13 September 2024). Instead of fixed search strings, we used combinations of keywords related to the robot (e.g., "social robot", "humanoid robot", "robot tutor") and the outcome (e.g., "learning", "cognitive skills", "education", "post-test"). This approach was needed because of the diverse terminology in the field to achieve comprehensive retrieval of relevant studies. This keyword search was supplemented by a manual review of the reference lists of previous meta-analyses and a prior review on social robots (Johal, 2020) to assess whether there were further relevant studies for inclusion. Furthermore, we performed forward citation tracking for eligible studies and targeted searches for other relevant publications by the authors of these studies. Additional searches were conducted through ResearchGate and using Google Search. This iterative process continued until no new eligible studies were being identified.

The selection process was conducted by two of the authors. One author conducted the initial screening and data extraction, with the

last author verifying all inclusions and resolving ambiguities. No automation tools were used in the process.

Included studies

Studies were included if they used a physical social robot for training cognitive skills in human participants. Studies in which the robot was used as an *additional* teaching medium/assistant to a human teacher (e.g., [Chang and Hwang, 2024](#)) or to a virtual interface (e.g., [Konijn and Hoorn, 2020](#)) were included, as were studies in which the robot was controlled through a Wizard-of-Oz method. Additionally, a number of studies (mostly due to COVID-19 restrictions or inspired by the circumstances at the time) involving the presentation of a physical social robot in video conferencing mode were included ([Donnermann et al., 2021](#); [Donnermann et al., 2024](#); [Kanero et al., 2021](#); [Li et al., 2016](#); [Norman et al., 2022](#); [Steinhausser et al., 2022](#); [Velentza, 2024a](#)). However, studies without a physical robot but with only a virtual-reality representation of a robot or an animated character (e.g., [Davis et al., 2023](#)) were not included. Studies in which participants created or programmed a robot kit (e.g., Lego Robotics) were not included, as these do not concern social robots. Studies in which performance improvement was attempted not through cognitive training but, for example, through emotional training or mindfulness, were also not included (e.g., [Tozadore et al., 2023](#)).

To facilitate a broad analysis of how effect sizes vary by study design, we deliberately adopted an inclusive approach to methodological rigor. Consequently, our sample was not limited to strictly randomized controlled trials; we also included studies with variations such as cluster randomization (e.g., by school class) or non-randomized, self-selected groups. Following this principle, studies with brief cognitive assessments were also accepted. Studies with a large discrepancy in pre-test scores between the groups were not included, but were retained for the post-test vs. pre-test analysis (see below).

In the vast majority of cases, the studies used a between-subjects design. However, a small number of studies used a within-subjects design ([Bravo Perucho and Alimardani, 2023](#); [De Haas et al., 2020](#); [Hu et al., 2023](#); [Kory Westlund et al., 2017](#); [Movellan et al., 2009](#); [Tanaka and Matsuzoe, 2012](#)). These studies were also included.

We did not include studies on autistic persons (e.g., [Alemi et al., 2015](#); [So et al., 2018](#); [Taheri et al., 2016](#)), people with learning disabilities (e.g., [Pérez et al., 2021](#); [Reardon et al., 2019](#)), hearing impairments (e.g., [Akalın et al., 2014](#)), or dementia (e.g., [Feng et al., 2022](#); [Valentí Soler et al., 2015](#)). These groups were excluded because it was expected that their learning abilities differ from the general population and may also involve different types of tasks to be learned.

To be included, the study had to measure learning effectiveness using a test administered after completion of the training, also referred to as a post-test. We excluded studies where test questions were interspersed within the training or where the robot itself asked the test questions, provided hints for solving them, or gave knowledge-of-results feedback on (some of) the test questions (e.g., [Chen Hsieh, 2021](#); [Lee and Specht, 2023](#); [Leyzberg et al., 2014](#); [Wedeborn, 2015](#)). In other words, this

meta-analysis focuses on comparing learning outcomes under equal post-test conditions and does not address how robots impact task performance.

Additionally, to be included in the analysis, the means and standard deviations of performance scores had to be available. In some cases, these values were extracted from figures in the paper, calculated from raw data presented in the figures, obtained from a data repository, or estimated using computer simulation. We sent approximately 70 emails in total to authors requesting further information. In about 40% of the cases, we received a response, and about two-thirds of those were useful for improving our annotations (see also [Derakhshan et al., 2025](#), who indicated that few authors respond to queries).

Other variables collected included the country where the study was conducted, the presence or absence of a human tutor during the training, the total duration of the training, and the similarity of the pre- and post-experiment assessments/tests, where applicable.

Since this study focuses on cognitive skills, the following types of effects were not included:

- Affective outcomes.
- Self-reported knowledge or perceived knowledge gain rather than knowledge measured through a test (e.g., [Ahmad et al., 2020](#); [Serholt et al., 2022](#)).
- Physical and motor performance (e.g., dancing, wayfinding, theater performance, walking speed, rehabilitation; [Al Hakim et al., 2020](#); [Benvenuti and Mazzoni, 2020](#)). However, handwriting skills were included. It was reasoned that handwriting is a typical cognitive skill learned in schools.
- Attitude change (e.g., attitudes towards waste recycling; [Castellano et al., 2021](#)).
- Social skills, wellbeing, anxiety (e.g., [Jeong et al., 2023](#)).

Effect size calculation—robot group vs. control group

Cohen's *d* was calculated between the post-test performance of the robot group and the control group. Specifically, for each study, we calculated Cohen's *d* per score on the outcome measure in the post-test (e.g., number of correctly answered questions, number of answers generated, task completion time) for the robot condition *versus* the control group featuring the same post-test and the same time interval between the training and post-test. In a number of studies ([Alves-Oliveira, 2020](#); [Gruson, 2019](#); [Howley et al., 2014](#); [Kennedy et al., 2015a](#); [Mubin et al., 2019](#); [Suzuki and Kanoh, 2015](#); [2017](#); [Zhanatkyzy et al., 2022](#)), no post-test score was available, but a learning gain relative to a pre-test score was reported; in these cases, we calculated Cohen's *d* between the learning gains instead of Cohen's *d* between the post-test scores.

The effect sizes per study were categorized into six different control groups:

1. Audio recording.
2. Human teacher.
3. Nothing (i.e., no training/instruction).
4. Paper (e.g., a list of material to be learned).
5. Sham task, such as interacting with the robot but not engaging with the robot's educational material ([Alves-Oliveira, 2020](#);

Vogt et al., 2019), or engaging in unrelated (cognitive) tasks (Resing et al., 2019; Resing et al., 2020).

6. Virtual interface. This could involve any digital medium, such as interacting with a tablet (e.g., Rintjema et al., 2018; Zhaxenova et al., 2020) or laptop (e.g., Miyauchi et al., 2020), but it could also involve a representation of a robot, i.e., a virtual robot (Kennedy et al., 2015a; Rosenthal-von der Pütten et al., 2016).

If, for a given comparison between the robot group and control group, multiple robot groups were available, Cohen's d was weighted according to sample size. Since the effects often involved a (weighted) average of multiple measures (with different standard deviations of the scores) and potentially multiple robot conditions, we used a simplified approach to calculate Cohen's d by assuming equal variance and equal sample size for the robot group and the control group.

Typically, in meta-analyses, either within-study variance (or sample size) determines the weight (fixed-effects meta-analysis), or studies are weighted more equally depending on the degree of between-study variance (random-effects meta-analysis). In determining the overall effect size, we assigned equal weight to all studies, regardless of their sample sizes. However, for studies reporting multiple robot-control comparisons, these were first aggregated internally by weighting them according to their respective sample sizes.

There were a number of reasons for unit-weighting the studies:

1. An exploratory analysis showed substantial between-study variance, with some studies showing a large negative effect of the robot (e.g., Zhanatkyzy et al., 2022) and others a positive effect (e.g., Alemi et al., 2014). This variability could not be explained by sample size alone but rather by the study design.
2. Some studies featured large sample sizes, but should not necessarily be seen as more important. For example, McIntosh et al. (2022) had 995 people interact briefly with a social robot and subsequently answer three health-related questions. Such a study should not necessarily carry more weight than a smaller study where participants interacted with the robot for a longer period or where their skills were assessed in a more comprehensive manner. This idea is consistent with various works indicating that assigning equal weight to effects is sensible, especially when sample sizes vary substantially (Bobko et al., 2007; Osburn and Callender, 1992; Shuster, 2010). Underlying the idea of assigning equal weights to studies is also the notion that sample sizes should not be seen as fixed parameters (as is typically done in meta-analyses) but are themselves subject to random sampling (Shuster, 2010).
3. Using equal weights allowed us to examine the effect of moderator variables in correlational analyses in a simple and transparent manner.

In addition to categorizing studies by control group, we examined two other potential moderator variables to explore the heterogeneity in the robot-versus-control effect sizes. Specifically, we compared the Cohen's d values from studies conducted within versus outside Europe. We also investigated the influence of participant age by calculating the correlation between the mean age reported in each study and its corresponding Cohen's d .

Effect size calculation—post-test vs. pre-test

A large number of studies reported pre-test scores in addition to post-test scores. For explanatory purposes, Cohen's d between the pre-test and post-test was computed, regardless of whether the study featured a control group. For each pre-post effect, we extracted the total duration of the training sessions in minutes (if a range was mentioned, the midpoint was taken). We calculated the Pearson product-moment correlation coefficient between the logarithm of the training duration and the pre-post effect. The logarithmic transformation was used to prevent the data from being overly influenced by a small number of training sessions with very long durations. Additionally, a learning curve typically follows a pattern where performance increases rapidly at first and then levels off as training duration increases, which is another reason to apply the logarithmic transformation.

Robot effectiveness—text analysis

Recently, it has become possible to summarize and evaluate texts, including entire research papers, using LLMs (e.g., De Winter, 2024; Tabone and De Winter, 2023). Our analysis protocol did not include a component-level assessment for specific internal validity threats (e.g., allocation concealment, participant blinding). Consequently, applying a traditional risk-of-bias tool was not appropriate. Instead, we chose to investigate potential reporting bias across the literature through a supplementary sentiment analysis. Specifically, we used Gemini to analyze the degree of positive wording in the text. Each paper's text was separately submitted to Google's Gemini 2.5 Pro (June 2025 version).

We applied a self-consistency method (e.g., Tabone and De Winter, 2023; Tang et al., 2024; Wang X. et al., 2023), in which the same question was posed multiple times with variations in the ordering of items. The average of these outputs was then calculated. The goal of this sentiment analysis was to derive an overall score reflecting the 'positivity towards the robot' expressed in each paper.

For each paper, we had Gemini 2.5 Pro answer the following questions:

On a scale of 0–10, ...

1. *How positive is the article about the social robot being studied?*
2. *How negative is the article about the social robot being studied?*
3. *How effective is the social robot under evaluation according to the article?*
4. *How ineffective is the social robot under evaluation according to the article?*
5. *How well does the social robot under evaluation perform according to the article?*
6. *How poorly does the social robot under evaluation perform according to the article?*
7. *Based on the introduction, how positive is this article about social robots in general?*
8. *Based on the introduction, how negative is this article about social robots in general?*
9. *Based on the discussion, how elaborately are the limitations of the experiment pointed out?*

10. *Based on the discussion, how elaborately are the strengths of the experiment pointed out?*
11. *To what extent are the authors biased in the positive direction, i.e. overly optimistic about social robots?*
12. *To what extent are the authors biased in the negative direction, i.e., overly skeptical about social robots?*

These 12 items, along with the automatically extracted text from the PDF of the paper, were submitted to the Gemini API. This process was repeated five times, with the 12 items presented in a different random order each time. Subsequently, we calculated a mean score per study for each item.

We then extracted a total score by performing a principal component analysis on the first six items. The six additional items about the paper's introduction, discussion, and potential bias were collected for exploratory purposes (see Supplementary Material). For each included study, a Positiveness score was calculated; across studies, these scores had a mean of 0 and a standard deviation of 1. We analyzed the Positiveness score in a similar manner to how we assessed Cohen's d effect sizes. Additionally, we calculated associations between the Positiveness score and the country of the research publication.

Results

We retrieved 146 studies, comprising 183 effect sizes between robot group and control group and 372 pre-post effects. Among the 146 studies, effect sizes between the robot and control groups were available for 78 studies, based on 4,819 participants. Pre-post effect sizes were available for 108 studies based on 6,232 participants.

Across the 146 included studies, publications were concentrated in the last decade (2015–2024: 129/146), indicating rapid growth of empirical work on social robots for learning outcomes. Conference proceedings/symposia/workshops (55) and journal articles (76) were represented in high numbers, with the remainder consisting of other formats (15; e.g., theses, reports, preprints).

The mean Cohen's d of the 372 pre-post effects was 1.08 ($SD = 1.32$). Figure 1 shows the effects for studies for which a pre-test score and post-test score were available for the robot condition as well as for the control group. It can be seen that in nearly all studies, participants demonstrated learning, indicated by a positive effect ($d > 0$) for both the control group and the robot group.

Figure 2 illustrates that the degree of learning relative to the pre-test was higher when the experimental condition featured prolonged training (>500 min). The correlation between training duration and pre-post effects is positive, at $r = 0.46$ (based on $m = 332$ pre-post effects for which the training duration was available; all control conditions combined).

The pre-post effects in Figures 1, 2 are indicative of the amount of learning relative to the pre-test. However, to assess the learning outcome of a social robot, it must be compared with a control group (i.e., the degree to which the marker is above the diagonal in Figure 1).

The Cohen's d values for the robot group versus the control group are shown in Figure 3, for six types of control groups. Training by means of social robots was found to be more effective than no training (Nothing: $M = 0.75$; $SD = 0.62$; Sham task: $M = 0.38$;

$SD = 0.40$). Whether training with a social robot is more effective than training with a virtual interface is less clear ($M = 0.20$; $SD = 0.37$). The comparison between a robot and a human teacher was overall positive ($M = 0.31$; $SD = 0.70$). According to an independent-samples t -test (treating comparisons as independent observations), Cohen's d was larger for 'Nothing' control groups compared to 'Virtual interface' control groups, $t(32) = 3.03$, $p = 0.005$. In other words, compared to no training, the robot appeared more effective than when it was compared with a virtual interface. The difference in Cohen's d between Human control groups and Nothing control groups was not significant, $t(51) = 1.64$, $p = 0.108$.

It can also be observed that, particularly when the control group involved a human teacher, the heterogeneity was high. This heterogeneity can be seen from the high standard deviation of the Cohen's d values ($SD = 0.70$), as well as from the non-overlapping 95% confidence intervals, with some studies showing a negative effect and others a positive effect.

When examining the strongest effects shown in Figure 3, the robot was accompanied by a human teacher (e.g., Tsai, 2019, d compared to Human = 2.07; Chang and Hwang, 2024, d compared to Human = 1.86; Alemi et al., 2023, d compared to Human = 1.80; Wu et al., 2015, d compared to Human = 1.62). This is remarkable when considering that in the majority of the studies included in this meta-analysis (59 out of 78, or 76% of studies), the robot was *not* accompanied by a human teacher. The Cohen's d for robot versus human comparisons was larger for studies where the robot was accompanied by a human teacher, i.e., co-teaching ($M = 0.88$; $SD = 0.68$; see 18 white markers in the 'Robot vs. Human' column in Figure 3) compared to robot-only conditions ($M = -0.06$; $SD = 0.40$; see 27 solid markers in Figure 3), $t(43) = 5.83$, $p < 0.001$.

We repeated the above analysis, this time focusing on the Positiveness scores. The results in Figure 4 show some similarities to those in Figure 3. For instance, the mean Positiveness score for Robot versus Nothing papers ($M = 0.50$; $SD = 0.34$) is higher than for Robot versus Virtual papers ($M = -0.55$; $SD = 1.35$). However, this difference was not statistically significant, $t(21) = 1.70$, $p = 0.105$.

Additionally, we examined whether the Gemini 2.5 Pro-based scores of the studies were related to the effect sizes as shown in Figure 3. If the robot was compared to more than one control condition (e.g., Human and Nothing, see Rosi et al., 2016), these effects were averaged across these control conditions. Thus, for the correlational analysis, multiple effects within a single study were averaged to produce one effect size per study. It was found that the Positiveness score was positively correlated with effect size ($r = 0.47$, $p < 0.001$, $k = 78$; Spearman's rank-order correlation = 0.62), as illustrated in Figure 5.

In a further attempt to explain under what conditions positive effect sizes and high Positiveness scores occur, we made a distinction between studies conducted within and outside Europe. This distinction was made after an initial exploration suggested that there might be national differences in the tone of papers, namely papers that seem to promote robots and those that critically evaluate robots.

We found that the effect size of the robot versus the control group was weaker for studies from Europe ($M = 0.17$; $SD = 0.39$; $k = 36$) than for studies outside Europe ($M = 0.53$; $SD = 0.69$; $k = 42$), a statistically significant difference, $t(76) = -2.74$, $p = 0.008$. A multivariable linear regression (fitted on 91 study-control-condition pairs) confirmed that this regional difference remained

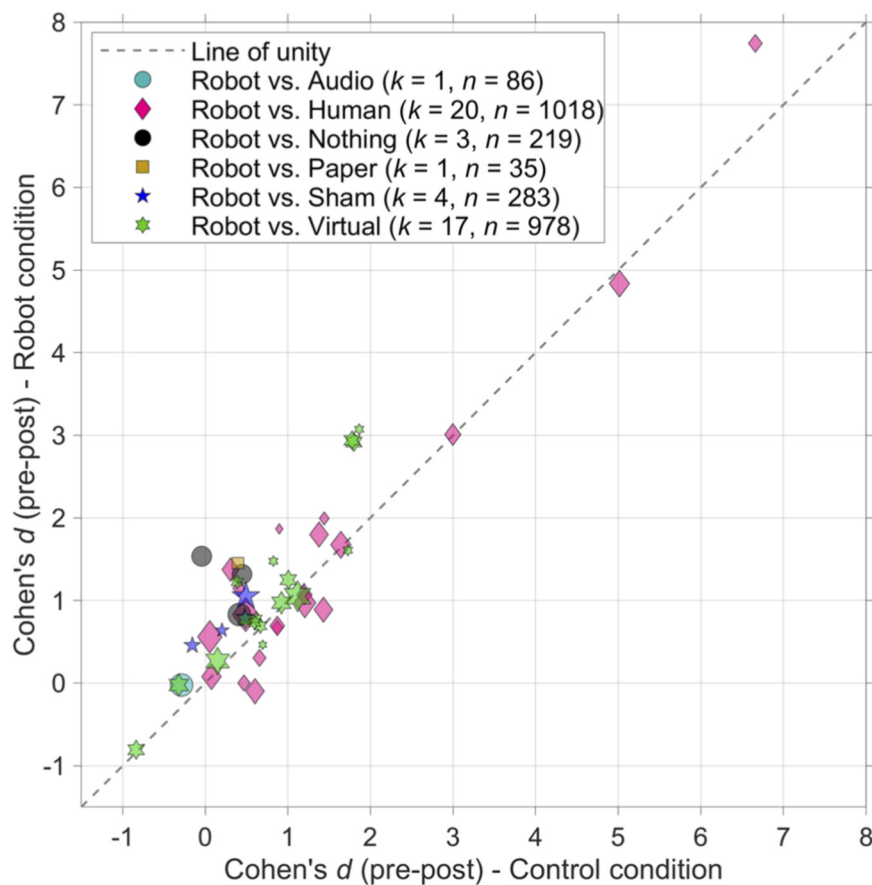


FIGURE 1

Scatter plot showing the pre-post effect sizes for the robot condition compared to the respective control group. Points above the unity line suggest greater effectiveness in the robot condition relative to the control group. The areas of the markers are linearly related to the sample sizes. k = number of studies; n = total sample size of the studies combined.

significant ($\beta = -0.37$, $p = 0.005$) after controlling for the type of control condition used. Furthermore, a Chi-square test indicated no significant difference in the selection of control conditions between regions ($p = 0.105$).

Furthermore, it was found that studies from Europe had a lower Positiveness score ($M = -0.40$, $SD = 1.09$, $k = 65$) than studies outside Europe ($M = 0.32$, $SD = 0.79$, $k = 81$), $t(144) = -4.61$, $p < 0.001$. A multivariable regression analysis performed on the subset of studies with a single control condition ($k = 66$) confirmed that this regional disparity remained significant ($\beta = -0.71$, $p = 0.006$) after controlling for the control condition type.

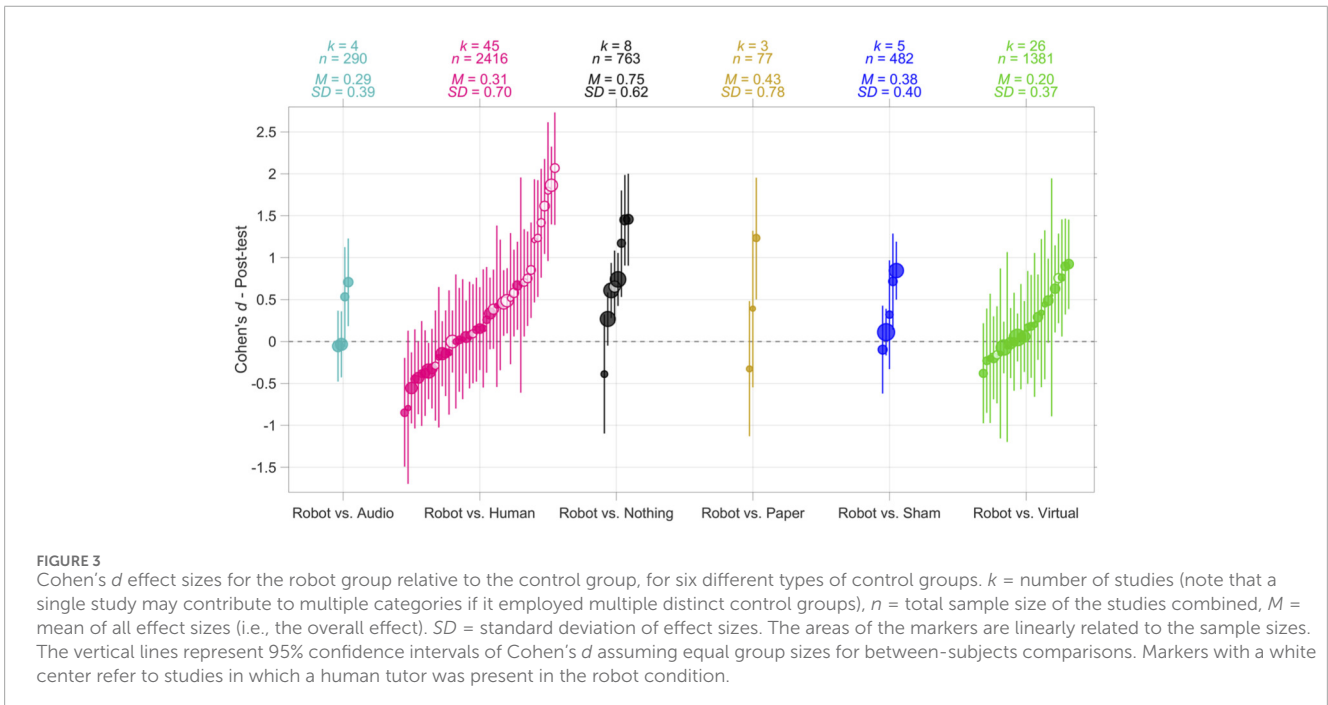
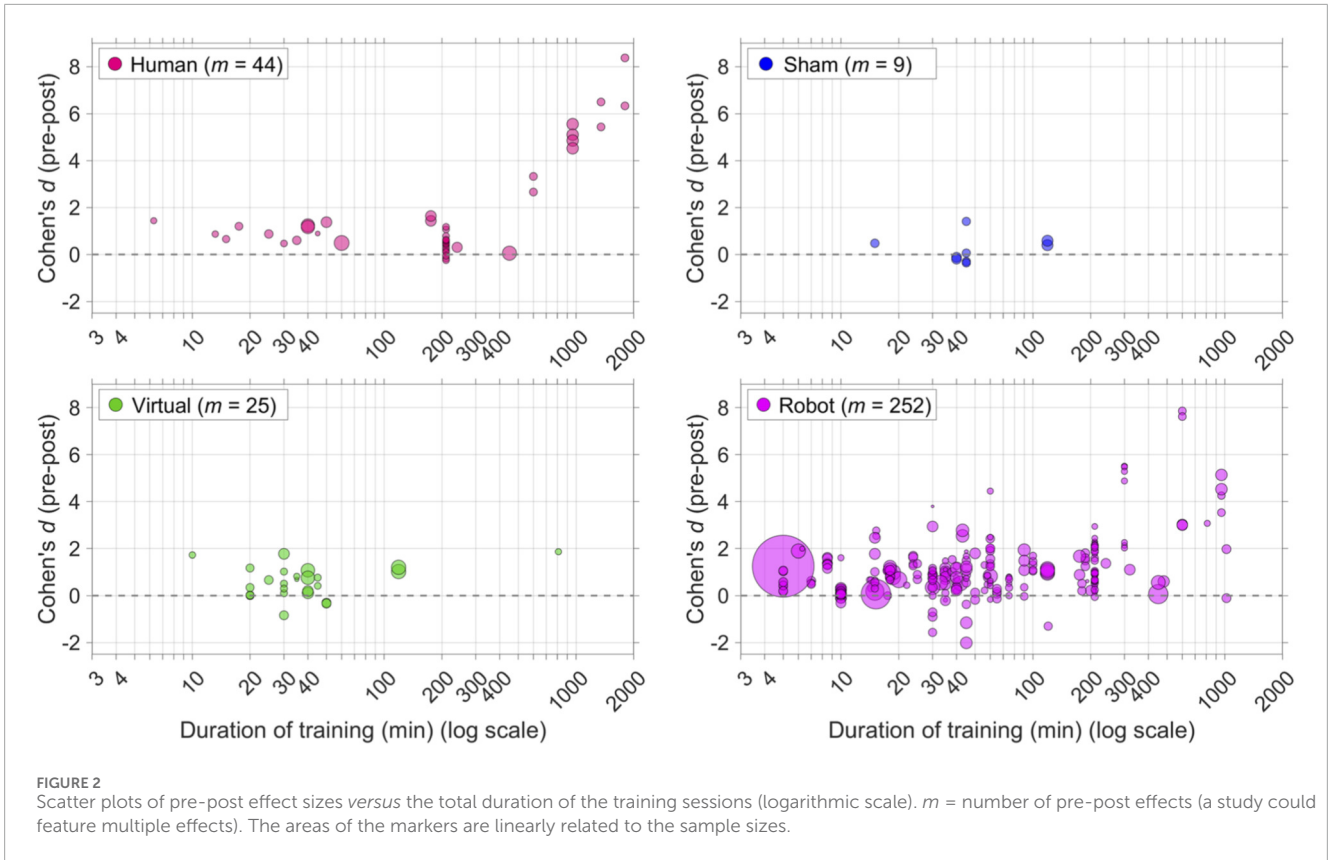
In response to the moderation effects that remained unclear in previous meta-analyses, we investigated whether the age of participants was associated with learning outcomes relative to the control group. We found no significant association between age (sample-size-weighted mean participant age in the study) and Cohen's d ($r = -0.05$, $p = 0.686$; Spearman's rank-order correlation = -0.13 , $p = 0.270$, $k = 78$ studies). To determine if age effects varied depending on the control condition, we performed a linear regression analysis with interaction terms ($\text{Age} \times \text{Control Type}$). The interaction terms were not statistically significant (all $p > 0.238$), indicating that the relationship between age and effect size did

not differ across control conditions. Furthermore, examining the groups individually revealed no significant correlations between age and effect size for any control type (e.g., Human: $r = -0.10$, $p = 0.525$; Virtual: $r = 0.23$, $p = 0.257$). Figure 6 presents a scatter plot illustrating the substantial heterogeneity of the effects, where age does not provide an obvious explanatory factor.

Discussion

This meta-analysis was conducted to synthesize the learning outcomes of educational interventions that used social robots. We focused on how the robot's pedagogical role and the type of control condition influence effectiveness. Our findings show that the effectiveness of social robots is dependent on their method of implementation. The most important factor is whether the robot is used as a teaching assistant or as a full teacher replacement. In addition, a sentiment analysis suggests that the reported effectiveness of robots may be connected to reporting biases that appear to differ by geographical region.

Our results show that the measured benefit of a social robot is directly dependent on the control condition against which it is



compared. The largest effect size appeared when robot interventions were compared to a “nothing” control group ($d = 0.75$). This outcome confirms that a structured educational activity is more effective than no activity.

A central debate in the literature concerns the added value of a robot's physical embodiment. Our analysis provides evidence to inform this debate, by revealing a modest but positive effect size for robots versus virtual interfaces ($d = 0.20$). At first glance, this

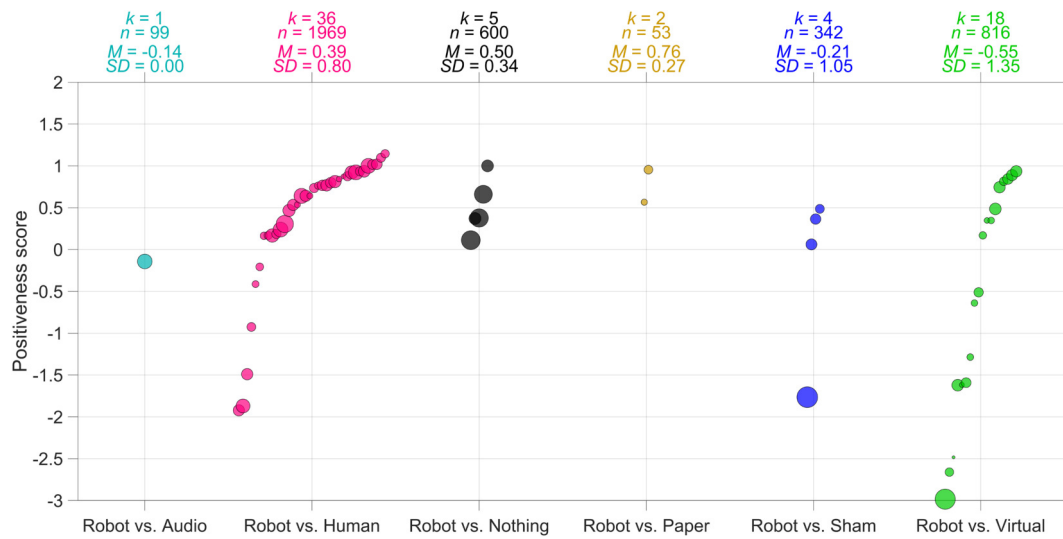


FIGURE 4 Gemini 2.5 Pro-based Positiveness scores of the research paper for different types of control groups. k = number of studies (papers), n = total sample size of the studies combined, M = mean Positiveness score. SD = standard deviation of Positiveness scores. The areas of the markers are linearly related to the sample sizes. Scores are only shown for papers in which only one type of control condition was present. For example, when the paper compared the robot with both a human teacher and a virtual interface, the Positiveness score is not shown.

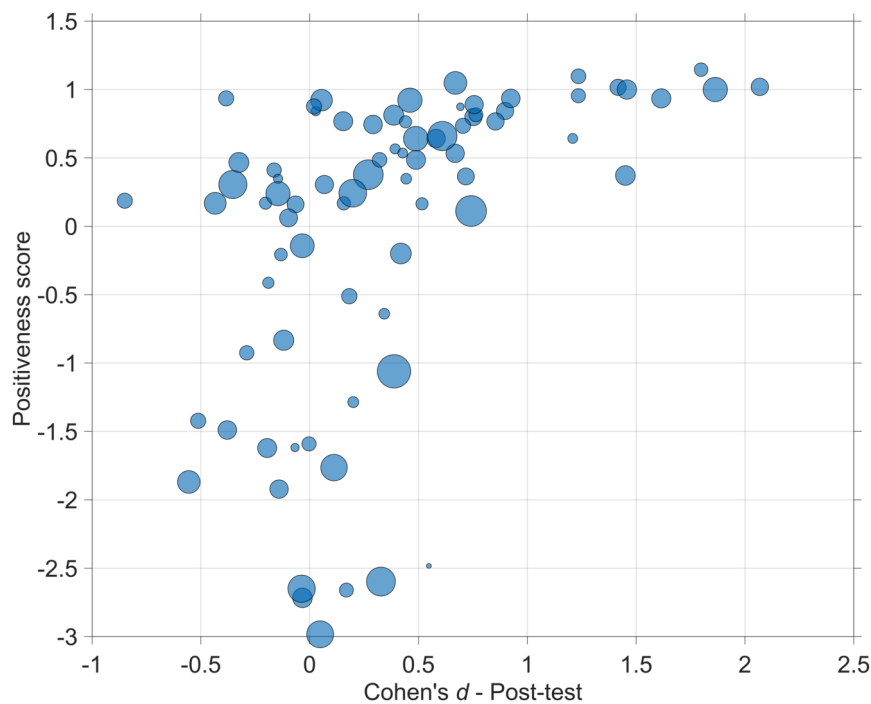


FIGURE 5 Scatter plot of the Positiveness score versus the Cohen's d effect size ($k = 78$). The areas of the markers are linearly related to the sample sizes of the studies.

result suggests that embodiment does confer a learning advantage, and lends support to theories that link physical presence to social presence and attention (Lee et al., 2006; Li, 2015). However, embodiment is not a panacea. Some work even reported that

a physical robot hindered learning compared to a virtual agent (Rosenthal-von der Pütten et al., 2016) or that its engaging social features negatively correlated with learning gains, possibly by acting as a distraction that increases cognitive load (Kennedy et al., 2015b).

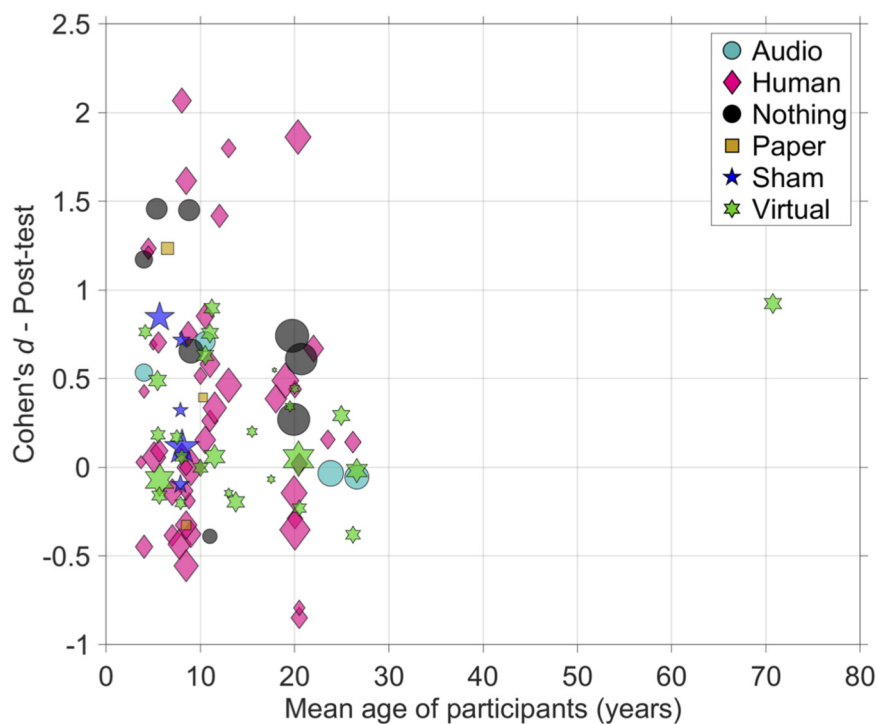


FIGURE 6

Scatter plot depicting the effect size between social robot and control group versus the mean participant age ($k = 78$; 91 effects), stratified by control condition. The areas of the markers are linearly related to the sample sizes of the studies.

The value of embodiment is likely dependent on task and implementation and not a guaranteed benefit. The type of feedback of the social robot matters; in three of the four studies where the robot was most effective against a virtual interface ($d > 0.75$), the robot offered more interactive and application-focused feedback, i.e., a question-and-answer module or real-life examples (Wei et al., 2021; $d = 0.92$), a variety of interactive functions (Hyun et al., 2008; $d = 0.76$), and more apps (Darmawansah and Hwang, 2024; $d = 0.76$).

This meta-analysis found a positive learning effect for robots versus human teachers ($d = 0.31$). The introduction also outlined several mechanisms through which robots might aid learning. One presumed mechanism is the mitigation of anxiety, a concept related to Krashen's Affective Filter Hypothesis (Hong et al., 2016). Students report feeling less pressure and anxiety with a robot because it is perceived as a non-judgmental partner (Donnermann et al., 2020; Peura et al., 2023; Suzuki and Kanoh, 2017; Wang et al., 2013). One student noted, "I felt comfortable practicing English with her because she is just a toy and I can say whatever I want with her" (Wang et al., 2013, p. 306). Some pedagogical strategies deliberately leverage this, programming the robot to make its own mistakes to normalize errors for the students (Alemi and Haeri, 2020; Alemi et al., 2014; Jimenez and Kanoh, 2013). Personalization was also noted as a potential benefit and can be connected to Self-Determination Theory (Lu et al., 2023; Van Minkelen et al., 2020). Our results include a number of studies where personalized robots that adapted to student performance led to better learning outcomes than non-personalized versions (Baxter et al., 2017; Kory-Westlund

and Breazeal, 2019; Leyzberg et al., 2018; Park et al., 2019; Ramachandran et al., 2017; Ramachandran et al., 2019b).

The most noteworthy finding of this meta-analysis is the powerful moderating effect of the robot's pedagogical role. When a robot acts as a teaching assistant in a co-teaching capacity with a human, it produces a strong positive effect on learning ($d = 0.88$). In contrast, when a robot is positioned as a complete replacement for the teacher, it offers no benefit ($d = -0.06$). This result indicates that the most effective use of current social robots in education is as a supportive tool for human teachers, not as a substitute. Current robots often cannot manage classroom dynamics, respond to spontaneous student questions, or handle unforeseen occurrences (Komatsubara et al., 2014; Ponce et al., 2019; Randall, 2019). As interactions can break down, a human may be required to repair these interactions (Bhat et al., 2016; Ponce et al., 2019; Woo et al., 2021). The efficacy of the robot is further constrained by the context gap between the design environment and the deployment environment. Singh et al. (2023) provided evidence from an under-resourced community school where the robot's Westernized accent forced students to rely on a human moderator for validation. In an assistive role, however, the robot can excel at structured and repetitive tasks. These tasks include leading drills or running quizzes (Alemi et al., 2023; Chang and Hwang, 2024; Wu et al., 2015). This frees the human teacher to provide higher-level support to students (Chang and Hwang, 2024; Wu et al., 2015). This synergy, which combines the robot's consistency with the human's adaptability, may be the underlying mechanism that creates an effective learning environment.

TABLE 1 Selected references with extreme scores on effect sizes and Positiveness scores.

References	Country	Robot vs. control effect size (rank out of 78)	Positiveness score (rank out of 146)	Selected quote from the paper	Our evaluation
Suzuki and Kanoh (2015)	Japan	$d = -0.85$ (1st, lowest overall)	0.19 (60th)	<i>“The results show that the robot provides the same effectiveness when engaging in expression education as does a human teacher”</i> (p. 498)	The authors claim equivalence despite a large (or inconclusive) negative effect size.
Alemi et al. (2023)	Iran	$d = 1.80$ (76th)	1.15 (146th, highest overall)	<i>“RALL can be a very appropriate alternative needing more attention by policymakers and language practitioners”</i> (p. 13)	The extremely positive sentiment score is consistent with the large effect size ($d = 1.80$) reported in the study.
Riedmann et al. (2024)	Germany	$d = 0.05$ (28th)	-2.98 (1st, lowest overall)	<i>“The generally high performance in test and post-test might be an indication of a ceiling effect with the potential problem to allow no sufficient differentiation between individual learners”</i> (p. 381), <i>“Due to the absence of desired effects, it is questionable to what extent the social robot was suitable for the implemented learning scenario. Even though ...participants in the present study even reported the robot as an interesting and fun learning partner, there was no discernible benefit”</i> (p. 382)	The authors candidly acknowledge the null result and explicitly question the utility of the robot.
Van den Berghe (2019)	Netherlands	$d = -0.03$ (24th)	-2.72 (2nd)	<i>“We tried to stimulate the interaction between the robot and the child ...but perhaps more interaction is required ...the robot was controlled via an assistant in real time, which entailed that its responses were rather slow”</i> (p. 121)	The tone is critical, aligning the null learning result with a transparent discussion of technical limitations (e.g., slow response times).
Bravo Perucho and Alimardani (2023)	Netherlands	$d = 0.55$ (54th)	-2.48 (7th)	<i>“students reported a significantly lower impression of robot’s usefulness for their activity”</i> (p. 358)	Despite finding a moderate positive learning effect, the authors focus the narrative on negative user perceptions.
Chandra et al. (2015)	Portugal/Switzerland	$d = -0.38$ (5th)	0.94 (134th)	<i>“The results of the study suggest that the teacher-children felt more responsible over the learner-children’s performance in the presence of robot facilitator”</i> (p. 172).	The abstract mentions positive social outcomes (responsibility) while not mentioning the null result in the abstract.

Our automated sentiment analysis using a LLM showed that Positiveness scores correlated strongly with Cohen's d effect sizes ($r = 0.47$; Spearman's rank-order correlation = 0.62). In a further attempt to explain the Positiveness scores, we reasoned that these scores are not just based on the effects found, but possibly also influenced by bias. In other words, the Positiveness scores may be partially attributable to the authors' inclination to present their findings in a more favorable manner, a practice commonly referred to as 'spin' (Boutron et al., 2010). An exploration of extreme effects or extremely high or low Positiveness scores suggests that different authors have different approaches, from introducing a positive spin to applying a critical tone (see Table 1 for an overview). Previous research supports the notion that biases are not isolated but are embedded in the culture of the research team, which can, in part, be influenced by the department, university (Köse and Korkmaz, 2019), or nationality (Vickers et al., 1998; Zhang et al., 2011). While our sentiment analysis attempted to quantify 'spin', publication bias is another threat. Although we retrieved many null results, the tendency toward positive sentiment, as shown in Table 1, suggests that successful or positively framed studies may be overrepresented in the present meta-analysis.

Finally, our analysis found no reliable relationship between participant age and learning effect sizes, a finding that remained consistent also when disaggregating the data by control condition. The wide scatter of data visible in our analysis (Figure 6) suggests that factors such as the robot's specific role, the nature of the learning task, and the quality of the instructional design are more powerful predictors of success than the learner's age alone.

Conclusion and recommendations

This meta-analysis shows that social robots can be effective educational tools, but it also qualifies this finding: the effectiveness of social robots is not a property of the technology but a function of their pedagogical implementation. Robots acting as assistants in a co-teaching relationship with a human educator produce large learning gains, while robots positioned as replacements for teachers are ineffective. Therefore, this meta-analysis suggests that the most promising path forward is not to develop robots that can supplant human teachers, but rather to design them as tools that support human-led instruction.

This study has several limitations. Our broad inclusion criteria allowed for a broad analysis but means that the overall effect sizes are averaged across studies of varying quality. The decision to weight all studies equally, although justified by the heterogeneity of designs and sample sizes, is a departure from traditional meta-analytic practice. Additionally, the sentiment analysis, while novel, is an exploratory method for investigating reporting bias and should be interpreted with caution. Finally, because we excluded affective outcomes, our findings apply only to cognitive performance. A robot that fails to improve test scores may still offer value through mechanisms not measured here.

In conclusion, the debate on social robotics should not focus on whether robots can teach but on *how* they can best be used to support learning. For practitioners, this meta-analysis provides a clear recommendation: social robots should be used as assistive

tools to enhance, not replace, human instruction. As AI capabilities advance, the robot may eventually be able to function as a complete substitute for a human teacher; however, being able to replace a teacher does not mean one should. Future research must guide this transition based on learning outcomes, rather than technological availability alone.

Data availability statement

The data and analysis scripts supporting the findings of this study are available online (<https://doi.org/10.4121/a78b6b99-3fdf-4ae0-97b0-3b618b00805e>), including the coded study data, Gemini-based sentiment scores, and MATLAB scripts for reproducing all analyses and figures.

Author contributions

JdW: Conceptualization, Investigation, Writing – original draft, Writing – review and editing, Methodology, Data curation, Visualization, Formal Analysis, Supervision, Validation. DD: Data curation, Investigation, Writing – review and editing. FM: Conceptualization, Investigation, Writing – review and editing. JB: Writing – review and editing, Conceptualization, Supervision.

Funding

The author(s) declared that financial support was not received for this work and/or its publication.

Conflict of interest

The author(s) declared that this work was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The author(s) declared that generative AI was used in the creation of this manuscript. The authors declare that generative AI was used for the sentiment analysis, as described in the Methods section.

Any alternative text (alt text) provided alongside figures in this article has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated

organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim

that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

- References marked with an asterisk (*) indicate studies included in the meta-analysis.
- Ahmad, M. I., Khordi-Moodi, M., and Lohan, K. S. (2020). "Social robot for STEM education," Companion of the 2020 ACM/IEEE international conference on human-robot interaction. Cambridge, UK, 90–92. doi:10.1145/3371382.3378291
- Akalin, N., Uluer, P., and Köse, H. (2014). "Non-verbal communication with a social robot peer: towards robot assisted interactive sign language tutoring," in Proceedings of the 2014 IEEE-RAS international conference on humanoid robots. Madrid, Spain, 1122–1127. doi:10.1109/HUMANOIDS.2014.7041509
- Al Hakim, V. G., Yang, S.-H., Tsai, T.-H., Lo, W.-H., Wang, J.-H., Hsu, T.-C., et al. (2020). "Interactive robot as classroom learning host to enhance audience participation in digital learning theater," in Proceedings of the 2020 IEEE 20th international conference on advanced learning technologies. Tartu, Estonia, 95–97. doi:10.1109/ICALT49669.2020.00036
- *Al Hakim, V. G., Yang, S.-H., Liyanawatta, M., Wang, J.-H., and Chen, G.-D. (2022). Robots in situated learning classrooms with immediate feedback mechanisms to improve students' learning performance. *Comput. and Educ.* 182, 104483. doi:10.1016/j.compedu.2022.104483
- *Alemi, M., and Haeri, N. S. (2017). "How to develop learners' politeness: a study of RALL's impact on learning greeting by young Iranian EFL learners," in Proceedings of the 2017 5th RSI international conference on robotics and mechatronics. Tehran, Iran, 88–94. doi:10.1109/ICRoM.2017.8466206
- Alemi, M., and Haeri, N. S. (2020). Robot-assisted instruction of L2 pragmatics: effects on young EFL learners' speech act performance. *Lang. Learn. and Technol.* 24, 86–103. doi:10.64152/10125/44727
- *Alemi, M., Meghdari, A., and Ghazisaedy, M. (2014). Employing humanoid robots for teaching english language in Iranian young high-schools. *Int. J. Humanoid Robot.* 11, 1450022. doi:10.1142/S0219843614500224
- Alemi, M., Meghdari, A., and Ghazisaedy, M. (2015). The impact of social robotics on L2 learners' anxiety and attitude in English vocabulary acquisition. *Int. J. Soc. Robotics* 7, 523–535. doi:10.1007/s12369-015-0286-y
- *Alemi, M., Bakhtiari Fard, N., and Rezanejad, A. (2023). The triumph of human-robot interaction: employing humanoid robot to boost grammar acquisition and retention among Iranian young EFL learners. *Mextesol J.* 47, 1–15. doi:10.61871/mj.v47n3-5
- *Alimardani, M., Van den Braak, S., Jouen, A.-L., Matsunaka, R., and Hiraki, K. (2021). "Assessment of engagement and learning during child-robot interaction using EEG signals," in *Social robotics. ICSR 2021*. Editors H. Li, S. S. Ge, Y. Wu, A. Wykowska, H. He, and X. Liu (Springer), 671–682. doi:10.1007/978-3-030-90525-5_59
- *Alves-Oliveira, P. (2020). "Experimental evidence of robots as tools for creativity stimulation," in *Boosting children's creativity through creative interactions with social robots*. Editor P. Alves-Oliveira 205–230. Doctoral dissertation, Lisbon: University of Lisbon. Available online at: <https://hdl.handle.net/10071/20620>.
- *Alves-Oliveira, P., Sequeira, P., Melo, F. S., Castellano, G., and Paiva, A. (2019). Empathic robot for group learning: a field study. *ACM Trans. Hum. Robot. Interact.* 8, 1–34. doi:10.1145/3300188
- *Alves-Oliveira, P., Arriaga, P., Cronin, M. A., and Paiva, A. (2020). "Creativity encounters between children and robots," Proceedings of the 2020 ACM/IEEE international conference on human-robot interaction. Cambridge, UK, 379–388. doi:10.1145/3319502.3374817
- *Banaeian, H., and Gilanlioglu, I. (2021). Influence of the NAO robot as a teaching assistant on university students' vocabulary learning and attitudes. *Australas. J. Educ. Technol.* 37, 71–87. doi:10.14742/ajet.6130
- *Baxter, P., Ashurst, E., Read, R., Kennedy, J., and Belpaeme, T. (2017). Robot education peers in a situated primary school study: personalisation promotes child learning. *PLoS One* 12, e0178126. doi:10.1371/journal.pone.0178126
- Belpaeme, T., Kennedy, J., Ramachandran, A., Scassellati, B., and Tanaka, F. (2018). Social robots for education: a review. *Sci. Robotics* 3, eaat5954. doi:10.1126/scirobotics.aat5954
- Benvenuti, M., and Mazzoni, E. (2020). Enhancing wayfinding in pre-school children through robot and socio-cognitive conflict. *Br. J. Educ. Technol.* 51, 436–458. doi:10.1111/bjjet.12848
- *Bhat, A., Chojnacki, A., and Knapp, E. (2016). "The future is nao: teaching mathematics to young schoolchildren using humanoid robots," in Proceedings of the new friends 2016 2nd international conference on social robots in therapy and education. Barcelona, Spain. Available online at: https://www.researchgate.net/publication/310234284_The_Future_is_Nao_Teaching_Mathematics_to_young_Schoolchildren_using_Humanoid_Robots.
- *Blancas-Muñoz, M., Vouloutsi, V., Zucca, R., Mura, A., and Verschure, P. F. (2018). "Hints vs distractions in intelligent tutoring systems: in search of the proper type of help," in Proceedings of the 12th international symposium on intelligent distributed computing. Bilbao, Spain. doi:10.48550/arXiv.1806.07806
- Bobko, P., Roth, P. L., and Buster, M. A. (2007). The usefulness of unit weights in creating composite scores: a literature review, application to content validity, and meta-analysis. *Organ. Res. Methods* 10, 689–709. doi:10.1177/1094428106294734
- Boutron, I., Dutton, S., Ravaut, P., and Altman, D. G. (2010). Reporting and interpretation of randomized controlled trials with statistically nonsignificant results for primary outcomes. *JAMA* 303, 2058–2064. doi:10.1001/jama.2010.651
- *Bravo Perucho, A., and Alimardani, M. (2023). "Social robots in secondary education: can robots assist young adult learners with math learning?," Companion of the 2023 ACM/IEEE international conference on human-robot interaction. Stockholm, Sweden, 355–359. doi:10.1145/3568294.3580105
- Castellano, G., De Carolis, B., D'Errico, F., Macchiarulo, N., and Rossano, V. (2021). Peppercycle: improving children's attitude toward recycling by playing with a social robot. *Int. J. Soc. Robotics* 13, 97–111. doi:10.1007/s12369-021-00754-0
- *Chandra, S., Alves-Oliveira, P., Lemaignan, S., Sequeira, P., Paiva, A., and Dillenbourg, P. (2015). Can a child feel responsible for another in the presence of a robot in a collaborative learning activity? in Proceedings of the 2015 24th IEEE international symposium on robot and human interactive communication (RO-MAN). Japan, Kobe, 167–172. doi:10.1109/ROMAN.2015.7333678
- *Chandra, S., Alves-Oliveira, P., Lemaignan, S., Sequeira, P., Paiva, A., and Dillenbourg, P. (2016). "Children's peer assessment and self-disclosure in the presence of an educational robot," in Proceedings of the 2016 25th IEEE international symposium on robot and human interactive communication (RO-MAN). New York, NY, 539–544. doi:10.1109/ROMAN.2016.7745170
- *Chandra, S., Paradedá, R., Yin, H., Dillenbourg, P., Prada, R., and Paiva, A. (2018). Do children perceive whether a robotic peer is learning or not? Proceedings of the 2018 ACM/IEEE international conference on human-robot interaction. Chicago, IL, 41–49. doi:10.1145/3171221.3171274
- *Chandra, S., Dillenbourg, P., and Paiva, A. (2020). Children teach handwriting to a social robot with different learning competencies. *Int. J. Soc. Robotics* 12, 721–748. doi:10.1007/s12369-019-00589-w
- *Chang, C.-C., and Hwang, G.-J. (2024). A robot-assisted real case-handling approach to improving students' learning performances in vocational training. *Educ. Inf. Technol.* 29, 22521–22544. doi:10.1007/s10639-024-12778-w
- Charpentier, S., Chetouani, M., Truck, I., Cohen, D., and Anzalone, S. M. (2022). "Social robots in learning scenarios: useful tools to improve students' attention or potential sources of distraction?," in *Social robotics. ICSR 2022*. Editors F. Cavallo, J.-J. Cabibihan, L. Fiorini, A. Sorrentino, H. He, and X. Liu (Cham: Springer), 124–134. doi:10.1007/978-3-031-24670-8_12
- *Chen, G.-D., Nurkhamid, Wang, C.-Y., Yang, S.-H., Lu, W.-Y., and Chang, C.-K. (2013). Digital learning playground: supporting authentic learning experiences in the classroom. *Interact. Learn. Environ.* 21, 172–183. doi:10.1080/10494820.2012.705856
- *Chen, H., Park, H. W., and Breazeal, C. (2020a). Teaching and learning with children: impact of reciprocal peer learning with a social robot on children's learning and emotive engagement. *Comput. and Educ.* 150, 103836. doi:10.1016/j.compedu.2020.103836
- *Chen, H., Park, H. W., Zhang, X., and Breazeal, C. (2020b). "Impact of interaction context on the student affect-learning relationship in child-robot interaction," Proceedings of the 2020 ACM/IEEE international conference on human-robot interaction. Cambridge, UK, 389–397. doi:10.1145/3319502.3374822
- Chen Hsieh, J. (2021). Digital storytelling outcomes and emotional experience among middle school EFL learners: robot-assisted versus powerpoint-assisted mode. *TESOL Q.* 55, 994–1010. doi:10.1002/tesq.3043
- *Chin, K.-Y., Hong, Z.-W., and Chen, Y.-L. (2014). Impact of using an educational robot-based learning system on students' motivation in elementary education. *IEEE Trans. Learn. Technol.* 7, 333–345. doi:10.1109/TLT.2014.2346756
- *Coninx, A., Baxter, P., Oleari, E., Bellini, S., Bierman, B., Blanson Henkemans, O., et al. (2016). Towards long-term social child-robot interaction: using multi-activity switching to engage young users. *J. Human-Robot Interact.* 5, 32–67. doi:10.5898/JHRI.5.1.Coninx
- *Conti, D., Di Nuovo, A., Cirasa, C., and Di Nuovo, S. (2017). "A comparison of kindergarten storytelling by human and humanoid robot with different social behavior," in Proceedings of the Companion of the 2017 ACM/IEEE international conference on human-robot interaction. Vienna, Austria, 97–98. doi:10.1145/3029798.3038359

- *Conti, D., Cirasa, C., Di Nuovo, S., and Di Nuovo, A. (2020). "Robot, tell me a tale!": a social robot as tool for teachers in kindergarten. *Interact. Stud.* 21, 220–242. doi:10.1075/is.18024.con
- *Darmawansah, D., and Hwang, G.-J. (2024). Effects of robot-based multiple low-stakes assessments on students' oral presentation performance, collective efficacy, and learning attitude. *Educ. Technol. Res. Dev.* 72, 2013–2039. doi:10.1007/s11423-024-10360-2
- Davis, R. O., Park, T., and Vincent, J. (2023). A meta-analytic review on embodied pedagogical agent design and testing formats. *J. Educ. Comput. Res.* 61, 30–67. doi:10.1177/07356331221100556
- *De Haas, M., Vogt, P., and Krahrmer, E. (2020). The effects of feedback on children's engagement and learning outcomes in robot-assisted second language learning. *Front. Robotics AI* 7, 101. doi:10.3389/frobt.2020.00101
- *De Haas, M., Vogt, P., and Krahrmer, E. (2021). When preschoolers interact with an educational robot, does robot feedback influence engagement? *Multimodal Technol. Interact.* 5, 77. doi:10.3390/mti5120077
- De Winter, J. (2024). Can ChatGPT be used to predict citation counts, readership, and social media interaction? An exploration among 2222 scientific abstracts. *Scientometrics* 129, 2469–2487. doi:10.1007/s11192-024-04939-y
- *De Wit, J., Schodde, T., Willemsen, B., Bergmann, K., De Haas, M., Kopp, S., et al. (2018). "The effect of a robot's gestures and adaptive tutoring on children's acquisition of second language vocabularies," in Proceedings of the 2018 ACM/IEEE international conference on human-robot interaction. New York, NY, 50–58. doi:10.1145/3171221.3171277
- *De Wit, J., Brandse, A., Krahrmer, E., and Vogt, P. (2020). "Varied human-like gestures for social robots: investigating the effects on children's engagement and language learning," in Proceedings of the 2020 ACM/IEEE international conference on human-robot interaction. Cambridge, UK, 359–367. doi:10.1145/3319502.3374815
- Derakhshan, A., Teo, T., Saeedy Robot, E., Janebi Enayat, M., and Jahanbakhsh, A. A. (2025). Robot-assisted language learning: a meta-analysis. *Rev. Educ. Res.* 95, 747–774. doi:10.3102/00346543241247227
- *Donnermann, M., Schaper, P., and Lugin, B. (2020). "Integrating a social robot in higher education – a field study," in Proceedings of the 2020 29th IEEE international conference on robot and human interactive communication (RO-MAN). Naples, Italy, 573–579. doi:10.1109/RO-MAN47096.2020.9223602
- *Donnermann, M., Schaper, P., and Lugin, B. (2021). "Towards adaptive robotic tutors in universities: a field study," in *Persuasive technology. PERSUASIVE 2021*. Editors R. Ali, B. Lugin, and F. Charles (Cham: Springer), 33–46. doi:10.1007/978-3-030-79460-6_3
- *Donnermann, M., Schaper, P., and Lugin, B. (2022). Social robots in applied settings: a long-term study on adaptive robotic tutors in higher education. *Front. Robotics AI* 9, 831633. doi:10.3389/frobt.2022.831633
- *Donnermann, M., Heinzmann, P., and Lugin, B. (2024). "Meet or call my robotic tutor? - the effect of a physically vs. virtually present social robot on learning outcomes, engagement and perception," in Companion of the 2024 ACM/IEEE international conference on human-robot interaction. Boulder, CO, 422–426. doi:10.1145/3610978.3640679
- *Draper, T. W., and Clayton, W. W. (1992). Using a personal robot to teach young children. *J. Genet. Psychol.* 153, 269–273. doi:10.1080/00221325.1992.10753723
- *Eimler, S., Von Der Pütten, A., Schächtle, U., Carstens, L., and Krämer, N. (2010). "Following the white rabbit – a robot rabbit as vocabulary trainer for beginners of English," in *HCI in work and learning, life and leisure*. Editors G. Leitner, M. Hitz, and A. Holzinger (Berlin: Springer), 322–339. doi:10.1007/978-3-642-16607-5_22
- *Elgarf, M., Calvo-Barajas, N., Alves-Oliveira, P., Perugia, G., Castellano, G., Peters, C., et al. (2022). "And then what happens?" promoting children's verbal creativity using a robot," in Proceedings of the 2022 ACM/IEEE international conference on human-robot interaction. Hokkaido, Japan, 71–79. doi:10.1109/HRI53351.2022.9889408
- *Esfandbod, A., Rokhi, Z., Meghdari, A. F., Taheri, A., Soleymani, Z., Alemi, M., et al. (2023). Fast mapping in word-learning: a case study on the humanoid social robots' impacts on children's performance. *Int. J. Child-Computer Interact.* 38, 100614. doi:10.1016/j.ijcci.2023.100614
- Feng, Y., Perugia, G., Yu, S., Barakova, E. I., Hu, J., and Rauterberg, G. W. M. (2022). Context-enhanced human-robot interaction: exploring the role of system interactivity and multimodal stimuli on the engagement of people with dementia. *Int. J. Soc. Robotics* 14, 807–826. doi:10.1007/s12369-021-00823-4
- *Gruson, C. (2019). The effect of personalization on engagement and learning gain of children in second-language (L2) tutoring using social robots. MSc thesis. Tilburg: Tilburg University. Available online at: <https://arno.uvt.nl/show.cgi?fid=150279>.
- *Han, J.-H., Jo, M.-H., Jones, V., and Jo, J.-H. (2008). Comparative study on the educational use of home robots for children. *J. Inf. Process. Syst.* 4, 159–168. doi:10.3745/JIPS.2008.4.4.159
- *Hindriks, K. V., and Liebens, S. (2019). "A robot math tutor that gives feedback," in *Social robotics*. Editors M. A. Salichs, S. S. Ge, E. I. Barakova, J.-J. Cabibihan, A. R. Wagner, and A. Castro-González (Cham: Springer), 601–610. ICSR 2019. doi:10.1007/978-3-030-35888-4_56
- *Hong, Z.-W., Huang, Y.-M., Hsu, M., and Shen, W.-W. (2016). Authoring robot-assisted instructional materials for improving learning performance and motivation in EFL classrooms. *J. Educ. Technol. and Soc.* 19, 337–349. Available online at: <https://www.jstor.org/stable/jeductechsoci.19.1.337>.
- *Hong, Z.-W., Tsai, M.-H. M., Ku, C. S., Cheng, W. K., Chen, J.-T., and Lin, J.-M. (2024). Utilizing robot-tutoring approach in oral reading to improve Taiwanese EFL students' English pronunciation. *Cogent Educ.* 11, 2342660. doi:10.1080/2331186X.2024.2342660
- *Hoorn, J. F., Huang, I. S., Konijn, E. A., and Van Buuren, L. (2021). Robot tutoring of multiplication: over one-third learning gain for most, learning loss for some. *Robotics* 10, 16. doi:10.3390/robotics10010016
- *Howley, I., Kanda, T., Hayashi, K., and Rosé, C. (2014). "Effects of social presence and social role on help-seeking and learning," in Proceedings of the 2014 ACM/IEEE international conference on human-robot interaction. Bielefeld, Germany, 415–422. doi:10.1145/2559636.2559667
- *Hsiao, H.-S., Chang, C.-S., Lin, C.-Y., and Hsu, H.-L. (2015). "iRobiQ": the influence of bidirectional interaction on kindergarteners' reading motivation, literacy, and behavior. *Interact. Learn. Environ.* 23, 269–292. doi:10.1080/10494820.2012.745435
- *Hsiao, H.-S., Chen, J.-H., Chang, T.-L., Li, P.-H., and Chung, G.-H. (2024). A study on the effects of using the 6E model and a robot teaching assistant on junior high school students' STEM knowledge, learning motivation, and hands-on performance. *J. Sci. Educ. Technol.* 33, 759–778. doi:10.1007/s10956-024-10119-7
- *Hsieh, M.-C., Pan, H.-C., Hsieh, S.-W., Hsu, M.-J., and Chou, S.-W. (2022). Teaching the concept of computational thinking: a STEM-Based program with tangible robots on project-based learning courses. *Front. Psychol.* 12, 828568. doi:10.3389/fpsyg.2021.828568
- *Hsieh, W.-M., Yeh, H.-C., and Chen, N.-S. (2025). Impact of a robot and tangible object (R&T) integrated learning system on elementary EFL learners' English pronunciation and willingness to communicate. *Comput. Assist. Lang. Learn.* 38, 773–798. doi:10.1080/09588221.2023.2228357
- *Hu, C.-C., Yang, Y.-F., and Chen, N.-S. (2023). Human-robot interface design – the 'Robot with a Tablet' or 'Robot only', which one is better? *Behav. and Inf. Technol.* 42, 1590–1603. doi:10.1080/0144929X.2022.2093271
- *Hyun, E.-J., Kim, S.-Y., Jang, S., and Park, S. (2008). "Comparative study of effects of language instruction program using intelligence robot and multimedia on linguistic ability of young children," in Proceedings of the 17th IEEE international symposium on robot and human interactive communication (RO-MAN). Munich, Germany, 187–192. doi:10.1109/ROMAN.2008.4600664
- *Iio, T., Maeda, R., Ogawa, K., Yoshikawa, Y., Ishiguro, H., Suzuki, K., et al. (2019). Improvement of Japanese adults' English speaking skills via experiences speaking to a robot. *J. Comput. Assisted Learn.* 35, 228–245. doi:10.1111/jcal.12325
- *Iio, T., Yoshikawa, Y., Ogawa, K., and Ishiguro, H. (2024). Comparison of outcomes between robot-assisted language learning system and human tutors: focusing on speaking ability. *Int. J. Soc. Robotics* 16, 743–761. doi:10.1007/s12369-024-01134-0
- *Jao, C.-Y., Yeh, H.-C., and Hung, H.-T. (2024). Exploring the effects of robot-assisted multimodal composition on students' audience awareness for English writing. *Comput. Assist. Lang. Learn.* 38, 1423–1451. doi:10.1080/09588221.2023.2300972
- Jeong, S., Aymerich-Franch, L., Arias, K., Alghowinem, S., Lapedriza, A., Picard, R., et al. (2023). Deploying a robotic positive psychology coach to improve college students' psychological well-being. *User Model. User-Adapted Interact.* 33, 571–615. doi:10.1007/s11257-022-09337-8
- *Jimenez, F., and Kanoh, M. (2013). "Effect of advances in wisdom of robot in collaborative learning," in Proceedings of the 1st international conference on human-agent interaction. Sapporo, Japan. Available online at: <https://hai-conference.net/ihai2013/proceedings/pdf/III-3-2.pdf>.
- *Jimenez, F., Kanoh, M., Yoshikawa, T., and Nakamura, T. (2020). Learning effect of collaborative learning with robots speaking a compliment. *J. Adv. Comput. Intell. Intelligent Inf.* 24, 396–403. doi:10.20965/jacii.2020.p0396
- Johal, W. (2020). Research trends in social robots for learning. *Curr. Robot. Rep.* 1, 75–83. doi:10.1007/s43154-020-00008-3
- *Kagkelidou, Z. (2024). Κοινωνικό ρομπότ και σεξουαλική αγωγή: επιπτώσεις στις στάσεις, αντιλήψεις και μαθησιακά αποτελέσματα των μαθητών δημοτικού [social robot and sexual education: impacts on the attitudes, perceptions, and learning outcomes of primary school students]. MSc thesis. University of Macedonia. Available online at: <http://dspace.lib.uom.gr/handle/2159/30777>.
- *Kanda, T., Hirano, T., Eaton, D., and Ishiguro, H. (2004). Interactive robots as social partners and peer tutors for children: a field trial. *Human-Computer Interact.* 19, 61–84. doi:10.1207/s15327051hci1901and2_4
- Kanda, T., Sato, R., Saiwaki, N., and Ishiguro, H. (2007). A two-month field trial in an elementary school for long-term human-robot interaction. *IEEE Trans. Robotics* 23, 962–971. doi:10.1109/TRO.2007.904904
- *Kanda, T., Shimada, M., and Koizumi, S. (2012). "Children learning with a social robot," Proceedings of the 2012 7th ACM/IEEE international conference on human-robot interaction. Boston, MA, 351–358. doi:10.1145/2157689.2157809

- *Kanero, J., Tunali, E. T., Oranç, C., Gökşun, T., and Küntay, A. C. (2021). When even a robot tutor zooms: a study of embodiment, attitudes, and impressions. *Front. Robotics AI* 8, 679893. doi:10.3389/frobt.2021.679893
- *Kanero, J., Oranç, C., Koşukulu, S., Kumkale, G. T., Gökşun, T., and Küntay, A. C. (2022). Are tutor robots for everyone? The influence of attitudes, anxiety, and personality on robot-led language learning. *Int. J. Soc. Robotics* 14, 297–312. doi:10.1007/s12369-021-00789-3
- *Kennedy, J., Baxter, P., and Belpaeme, T. (2015a). Comparing robot embodiments in a guided discovery learning interaction with children. *Int. J. Soc. Robotics* 7, 293–308. doi:10.1007/s12369-014-0277-4
- *Kennedy, J., Baxter, P., and Belpaeme, T. (2015b). “The robot who tried too hard: social behaviour of a robot tutor can negatively affect child learning,” Proceedings of the 2015 10th ACM/IEEE international conference on human-robot interaction. Portland, OR, 67–74. doi:10.1145/2696454.2696457
- *Kennedy, J., Baxter, P., Senft, E., and Belpaeme, T. (2015c). “Higher nonverbal immediacy leads to greater learning gains in child-robot tutoring interactions,” in *Social robotics*. Editors A. Tapus, E. André, J. C. Martin, F. Ferland, and M. Ammi (Cham: Springer), 327–336. *ICSR* 2015. doi:10.1007/978-3-319-25554-5_33
- *Kennedy, J., Baxter, P., Senft, E., and Belpaeme, T. (2016a). “Heart vs hard drive: children learn more from a human tutor than a social robot,” Proceedings of the 2016 11th ACM/IEEE international conference on human-robot interaction. Christchurch, New Zealand, 451–452. doi:10.1109/HRI.2016.7451801
- *Kennedy, J., Baxter, P., Senft, E., and Belpaeme, T. (2016b). “Social robot tutoring for child second language learning,” Proceedings of the 2016 11th ACM/IEEE international conference on human-robot interaction. Christchurch, New Zealand, 231–238. doi:10.1109/HRI.2016.7451757
- *Kennedy, J., Baxter, P., and Belpaeme, T. (2017). Nonverbal immediacy as a characterisation of social behaviour for human-robot interaction. *Int. J. Soc. Robotics* 9, 109–128. doi:10.1007/s12369-016-0378-3
- *Komatsubara, T., Shiomi, M., Kanda, T., Ishiguro, H., and Hagita, N. (2014). “Can a social robot help children’s understanding of science in classrooms?” in Proceedings of the second international conference on human-agent interaction. Tsukuba, Japan, 83–90. doi:10.1145/2658861.2658881
- *Konijn, E. A., and Hoorn, J. F. (2020). Robot tutor and pupils’ educational ability: teaching the times tables. *Comput. and Educ.* 157, 103970. doi:10.1016/j.compedu.2020.103970
- *Konijn, E. A., Jansen, B., Mondaca Bustos, V., Hobbelenk, V. L. N. F., and Preciado Vanegas, D. (2022). Social robots for (second) language learning in (migrant) primary school children. *Int. J. Soc. Robotics* 14, 827–843. doi:10.1007/s12369-021-00824-3
- *Kory Westlund, J. M., Dickens, L., Jeong, S., Harris, P. L., DeSteno, D., and Breazeal, C. L. (2017). Children use non-verbal cues to learn new words from robots as well as people. *Int. J. Child-Computer Interact.* 13, 1–9. doi:10.1016/j.ijcci.2017.04.001
- *Kory-Westlund, J. M., and Breazeal, C. (2019). A long-term study of young children’s rapport, social emulation, and language learning with a peer-like robot playmate in preschool. *Front. Robotics AI* 6, 81. doi:10.3389/frobt.2019.00081
- Köse, M. F., and Korkmaz, M. (2019). Why are some universities better? An evaluation in terms of organizational culture and academic performance. *High. Educ. Res. and Dev.* 38, 1213–1226. doi:10.1080/07294360.2019.1634679
- *Lanzillotti, R., Piccinno, A., Rossano, V., and Roselli, T. (2021). “Social robot to teach coding in primary school,” in Proceedings of the 2021 international conference on advanced learning technologies. Tartu, Estonia, 102–104. doi:10.1109/ICALT52272.2021.00038
- Lee, H., and Lee, J. H. (2022). The effects of robot-assisted language learning: a meta-analysis. *Educ. Res. Rev.* 35, 100425. doi:10.1016/j.edurev.2021.100425
- Lee, Y., and Specht, M. (2023). “Can we empower attentive e-reading with a social robot? An introductory study with a novel multimodal dataset and deep learning approaches,” in Proceedings of the LAK23: 13th international learning analytics and knowledge conference. Arlington, TX, 520–530. doi:10.1145/3576050.3576122
- Lee, K. M., Jung, Y., Kim, J., and Kim, S. R. (2006). Are physically embodied social agents better than disembodied social agents? the effects of physical embodiment, tactile interaction, and people’s loneliness in human-robot interaction. *Int. J. Human-Computer Stud.* 64, 962–973. doi:10.1016/j.ijhcs.2006.05.002
- *Lee, S., Noh, H., Lee, J., Lee, K., and Lee, G. G. (2010). “POSTECH approaches for dialog-based English conversation tutoring,” in Proceedings of the second APSIPA annual summit and conference. Biopolis, Singapore, 794–803. Available online at: http://www.apsipa.org/proceedings_2010/pdf/APSIPA158.pdf.
- *Levinson, L., Gvirsman, O., Gorodesky, I. M., Perez, A., Gonen, E., and Gordon, G. (2021). Learning in summer camp with social robots: a morphological study. *Int. J. Soc. Robotics* 13, 999–1012. doi:10.1007/s12369-020-00689-y
- Leyzberg, D., Spaulding, S., and Scassellati, B. (2014). “Personalizing robot tutors to individuals’ learning differences,” in Proceedings of the 9th ACM/IEEE international conference on human-robot interaction. Bielefeld, Germany, 423–430. doi:10.1145/2559636.2559671
- *Leyzberg, D., Ramachandran, A., and Scassellati, B. (2018). The effect of personalization in longer-term robot tutoring. *ACM Trans. Human-Robot Interact.* 7, 1–19. doi:10.1145/3283453
- Li, J. (2015). The benefit of being physically present: a survey of experimental works comparing copresent robots, telepresent robots and virtual agents. *Int. J. Human-Computer Stud.* 77, 23–37. doi:10.1016/j.ijhcs.2015.01.001
- *Li, J., Kizilcec, R., Bailenson, J., and Ju, W. (2016). Social robots and virtual agents as lecturers for video instruction. *Comput. Hum. Behav.* 55, 1222–1230. doi:10.1016/j.chb.2015.04.005
- *Liebens, S. (2019). Adaptive feedback by a humanoid robot tutoring math. MSc thesis. Delft University of Technology. Available online at: <https://resolver.tudelft.nl/uuid:6d4c4b0d-3f25-4d40-8802-17ccc0aae694>.
- *Liles, K. R. (2019). “Ms. An (meeting students’ academic needs): engaging students in math education,” in *Adaptive instructional systems. HCII 2019*. Editors R. Sottilare, and J. Schwarz (Springer), 645–661. doi:10.1007/978-3-030-22341-0_50
- *Liles, K. R., and Beer, J. M. (2015). Rural minority students’ perceptions of ms. An, the robot teaching assistant, as a social teaching tool. *Proc. Hum. Factors Ergonomics Soc. Annu. Meet.* 59, 372–376. doi:10.1177/1541931215591077
- *Lin, V., Yeh, H.-C., Huang, H.-H., and Chen, N.-S. (2022). Enhancing EFL vocabulary learning with multimodal cues supported by an educational robot and an IoT-based 3D book. *System* 104, 102691. doi:10.1016/j.system.2021.102691
- *Liu, X., Ma, J., and Wang, Q. (2024). A social robot as your reading companion: exploring the relationships between gaze patterns and knowledge gains. *J. Multimodal User Interfaces* 18, 21–41. doi:10.1007/s12193-023-00418-5
- *Lu, Y., Chen, C., Chen, P., and Yu, S. (2023). Designing social robot for adults using self-determination theory and AI technologies. *IEEE Trans. Learn. Technol.* 16, 206–218. doi:10.1109/TLT.2023.3250465
- *Lubbers, E. (2020). Personalised second language robot tutoring: the use of an adaptive system with prior knowledge MSc thesis. Tilburg University. Available online at: <https://arno.uvt.nl/show.cgi?fid=151022>.
- *Lubold, N. (2018a). “Effects on rapport and learning with Emma,” in *Producing acoustic-prosodic entrainment in a robotic learning companion to build learner rapport*. Editor N. Lubold (Doctoral dissertation, Arizona State University), 175–199. Available online at: <https://hdl.handle.net/2286/R.I.51667>.
- *Lubold, N. (2018b). “Effects on rapport and learning with Nico,” in *Producing acoustic-prosodic entrainment in a robotic learning companion to build learner rapport*. Editor N. Lubold (Doctoral dissertation, Arizona State University), 133–159. Available online at: <https://hdl.handle.net/2286/R.I.51667>.
- *Lubold, N., Walker, E., Pon-Barry, H., and Ogan, A. (2019). “Comfort with robots influences rapport with a social, entraining teachable robot,” in *Artificial intelligence in education*. Editors S. Isotani, E. Millán, A. Ogan, P. Hastings, B. McLaren, and R. Luckin (Cham: Springer), 231–243. *AIED* 2019. doi:10.1007/978-3-030-23204-7_20
- *Manikutty, G., Potapragada, S. A., Pasupuleti, D., Unnithan, M. S., Venugopal, A., Prabha, P., et al. (2024). “Exploring child-robot interaction in individual and group settings in India,” in Proceedings of the 2024 8th international conference on robotics and automation sciences. Tokyo, Japan, 37–42. doi:10.1109/ICRAS62427.2024.10654467
- *Mazzoni, E., and Benvenuti, M. (2015). A robot-partner for preschool children learning English using socio-cognitive conflict. *J. Educ. Technol. and Soc.* 18, 474–485. Available online at: <https://www.jstor.org/stable/jeductechsoci.18.4.474>.
- *McIntosh, C., Elvin, A., Smyth, W., Birks, M., and Nagle, C. (2022). Health promotion, health literacy and vaccine hesitancy: the role of humanoid robots. *Inq. J. Health Care Organ. Provis. Financing* 59, 469580221078515. doi:10.1177/00469580221078515
- *Michaelis, J. E., and Mutlu, B. (2019). “Supporting interest in science learning with a social robot,” Proceedings of the 18th ACM international conference on interaction design and children. Boise, 71–82. doi:10.1145/3311927.3323154
- *Miyachi, K., Jimenez, F., Yoshikawa, T., Furuhashi, T., and Kanoh, M. (2020). Learning effects of robots teaching based on cognitive apprenticeship theory. *J. Adv. Comput. Intell. Intelligent Inf.* 24, 101–112. doi:10.20965/jaciii.2020.p0101
- *Molenaar, B., Soliño Fernández, B., Polimeno, A., Barakova, E., and Chen, A. (2021). Pitch it right: using prosodic entrainment to improve robot-assisted foreign language learning in school-aged children. *Multimodal Technol. Interact.* 5, 76. doi:10.3390/mti5120076
- *Moorlag, F. (2021). The effects of a social robot’s gestures on learning outcomes MSc thesis. Delft University of Technology. Available online at: <https://resolver.tudelft.nl/uuid:ae1edd23-ae6b-4d36-a7b7-8ec4524bf56d>.
- *Movellan, J., Eckhardt, M., Virnes, M., and Rodriguez, A. (2009). “Sociable robot improves toddler vocabulary skills,” in Proceedings of the 4th ACM/IEEE international conference on human-robot interaction. La Jolla, CA, 307–308. doi:10.1145/1514095.1514189
- *Mubin, O., Alhashmi, M., Baroud, R., and Alnajjar, F. S. (2019). “Humanoid robots as teaching assistants in an Arab school,” in Proceedings of the 31st Australian conference on human-computer-interaction. Fremantle, Australia, 462–466. doi:10.1145/3369457.3369517

- Neumann, M. M. (2020). Social robots and young children's early language and literacy learning. *Early Child. Educ. J.* 48, 157–170. doi:10.1007/s10643-019-00997-7
- *Norman, U., Chin, A., Bruno, B., and Dillenbourg, P. (2022). "Efficacy of a 'misconceiving' robot to improve computational thinking in a collaborative problem solving activity: a pilot study," in Proceedings of the 2022 31st IEEE international conference on robot and human interactive communication (RO-MAN), Naples, Italy. doi:10.1109/RO-MAN53752.2022.9900775
- *Okawa, K., Jimenez, F., Akizuki, S., and Yoshikawa, T. (2024). Proposal of learning support model for teacher-type robot supporting learning according to learner's perplexed facial expressions. *J. Robotics Mechatronics* 36, 168–180. doi:10.20965/jrm.2024.p0168
- *Oralbayeva, N., Amirova, A., Telisheva, Z., Zhanatkyzy, A., Aimysheva, A., and Sandygulova, A. (2023). "Montessori-based design of long-term child-robot interaction for alphabet learning," Companion of the 2023 ACM/IEEE international conference on human-robot interaction. Stockholm, Sweden, 691–695. doi:10.1145/3568294.3580175
- Osburn, H. G., and Callender, J. (1992). A note on the sampling variance of the mean uncorrected correlation in meta-analysis and validity generalization. *J. Appl. Psychol.* 77, 115–122. doi:10.1037/0021-9010.77.2.115
- Ouyang, F., and Xu, W. (2024). The effects of educational robotics in STEM education: a multilevel meta-analysis. *Int. J. STEM Educ.* 11, 7. doi:10.1186/s40594-024-00469-4
- Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., et al. (2021). The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ* 372, n71. doi:10.1136/bmj.n71
- Papadopoulos, I., Lazzarino, R., Miah, S., Weaver, T., Thomas, B., and Koulouglioti, C. (2020). A systematic review of the literature regarding socially assistive robots in pre-tertiary education. *Comput. and Educ.* 155, 103924. doi:10.1016/j.compedu.2020.103924
- *Park, S. J., Han, J. H., Kang, B. H., and Shin, K. C. (2011). "Teaching assistant robot, ROBOSEM, in English class and practical issues for its diffusion," in Proceedings of the IEEE workshop on advanced robotics and its social impacts. Menlo Park, CA, 8–11. doi:10.1109/ARSO.2011.6301971
- *Park, H. W., Rosenberg-Kima, R., Rosenberg, M., Gordon, G., and Breazeal, C. (2017). "Growing growth mindset with a social robot peer," in Proceedings of the 2017 ACM/IEEE international conference on human-robot interaction, Vienna, Austria, 137–145. doi:10.1145/2909824.3020213
- *Park, H. W., Grover, I., Spaulding, S., Gomez, L., and Breazeal, C. (2019). A model-free affective reinforcement learning approach to personalization of an autonomous social robot companion for early literacy education. *Proc. AAAI Conf. Artif. Intell.* 33, 687–694. doi:10.1609/aaai.v33i01.3301687
- *Pasilidou, C., Fachantidis, N., and Koivu, E. (2023). "Using augmented reality and a social robot to teach geography in primary school," in Proceedings of the international conference on human-computer interaction. Cham: Springer Nature, 371–385. doi:10.1007/978-3-031-34550-0_27
- Pérez, J., Azuaje, M., León, C., and Pedroza, O. (2021). Effects of social robotics on episodic memory in children with intellectual disabilities. *IEEE Rev. Iberoam. Tecnol. del Aprendiz.* 16, 393–399. doi:10.1109/RITA.2021.3125899
- *Peura, L., Mutta, M., and Johansson, M. (2023). Playing with pronunciation: a study on robot-assisted French pronunciation in a learning game. *Nordic J. Digital Lit.* 18, 100–115. doi:10.18261/njdl.18.2.3
- *Ponce, P., Molina, A., Lopez Caudana, E. O., Reyes, G. B., and Parra, N. M. (2019). Improving education in developing countries using robotic platforms. *Int. J. Interact. Des. Manuf.* 13, 1401–1422. doi:10.1007/s12008-019-00576-5
- *Ramachandran, A., Huang, C.-M., and Scassellati, B. (2017). "Give me a break! personalized timing strategies to promote learning in robot-child tutoring," Proceedings of the 2017 ACM/IEEE international conference on human-robot interaction. Vienna, Austria, 146–155. doi:10.1145/2909824.3020209
- *Ramachandran, A., Huang, C.-M., Gartland, E., and Scassellati, B. (2018). "Thinking aloud with a tutoring robot to enhance learning," Proceedings of the 2018 ACM/IEEE international conference on human-robot interaction. Chicago, IL, 59–68. doi:10.1145/3171221.3171250
- *Ramachandran, A., Huang, C.-M., and Scassellati, B. (2019a). Toward effective robot-child tutoring: internal motivation, behavioral intervention, and learning outcomes. *ACM Trans. Interact. Intell. Syst.* 9, 1, 23. doi:10.1145/3213768
- *Ramachandran, A., Sebo, S. S., and Scassellati, B. (2019b). Personalized robot tutoring using the Assistive Tutor POMDP (AT-POMDP). *Proc. AAAI Conf. Artif. Intell.* 33, 8050–8057. doi:10.1609/aaai.v33i01.33018050
- Randall, N. (2019). A survey of robot-assisted language learning (RALL). *ACM Trans. Human-Robot Interact.* 9, 1–36. doi:10.1145/3345506
- Reardon, C., Zhang, H., Wright, R., and Parker, L. E. (2019). Robots can teach students with intellectual disabilities: educational benefits of using robotic and augmented reality applications. *IEEE Robotics and Automation Mag.* 26, 79–93. doi:10.1109/MRA.2018.2868865
- *Resing, W. C. M., Bakker, M., Elliott, J. G., and Vogelaar, B. (2019). Dynamic testing: can a robot as tutor be of help in assessing children's potential for learning? *J. Comput. Assisted Learn.* 35, 540–554. doi:10.1111/jcal.12358
- *Resing, W. C. M., Vogelaar, B., and Elliott, J. G. (2020). Children's solving of "Tower of Hanoi" tasks: dynamic testing with the help of a robot. *Educ. Psychol.* 40, 1136–1163. doi:10.1080/01443410.2019.1684450
- *Riedmann, A., Schaper, P., and Lugin, B. (2024). Integration of a social robot and gamification in adult learning and effects on motivation, engagement and performance. *AI and Soc.* 39, 369–388. doi:10.1007/s00146-022-01514-y
- *Rintjema, E., Van den Berghe, R., Kessels, A., De Wit, J., and Vogt, P. (2018). "A robot teaching young children a second language: the effect of multiple interactions on engagement and performance," Companion of the 2018 ACM/IEEE international conference on human-robot interaction. Chicago, IL, 219–220. doi:10.1145/3173386.3177059
- *Rosenthal-von der Pütten, A. M., Straßmann, C., and Krämer, N. C. (2016). "Robots or agents – neither helps you more or less during second language acquisition: experimental study on the effects of embodiment and type of speech output on evaluation and alignment," in *Intelligent Virtual Agents. IVA 2016*. Editors D. Traum, W. Swartout, P. Khooshabeh, S. Kopp, S. Scherer, and A. Leuski (Cham: Springer), 256–268. doi:10.1007/978-3-319-47665-0_23
- *Rosi, A., Dall'Asta, M., Brighenti, F., Del Rio, D., Volta, E., Baroni, I., et al. (2016). The use of new technologies for nutritional education in primary schools: a pilot study. *Public Health* 140, 50–55. doi:10.1016/j.puhe.2016.08.021
- *Salomons, N., Pineda, K. T., Adéjare, A., and Scassellati, B. (2022). "We make a great team!": adults with low prior domain knowledge learn more from a peer robot than a tutor robot," Proceedings of the 2022 ACM/IEEE international conference on human-robot interaction. Sapporo, Japan, 176–184. doi:10.1109/HRI53351.2022.9889441
- *Sandygulova, A., Johal, W., Zhexenova, Z., Tleubayev, B., Zhanatkyzy, A., Turarova, A., et al. (2020). "CoWriting Kazakh: learning a new script with a robot," in Proceedings of the 2020 ACM/IEEE international conference on human-robot interaction. Cambridge, UK, 113–120. doi:10.1145/3319502.3374813
- *Sarmonov, S., Shakerimov, A., Aimysheva, A., Amirova, A., Oralbayeva, N., Zhanatkyzy, A., et al. (2023). "Robot-assisted first language learning in a new Latin alphabet: the reinforcement learning-based QWriter system," Companion of the 2023 ACM/IEEE international conference on human-robot interaction. Stockholm, Sweden, 677–681. doi:10.1145/3568294.3580172
- *Schodde, T., Hoffmann, L., Stange, S., and Kopp, S. (2019). Adapt, explain, engage—A study on how social robots can scaffold second-language learning of children. *ACM Trans. Human-Robot Interact.* 9, 1–27. doi:10.1145/3366422
- Serholt, S., Ekström, S., Küster, D., Ljungblad, S., and Pareto, L. (2022). Comparing a robot tutor to a human tutor in a learning-by-teaching scenario with children. *Front. Robotics AI* 9, 836462. doi:10.3389/frobt.2022.836462
- *Shakerimov, A., Sarmonov, S., Amirova, A., Oralbayeva, N., Zhanatkyzy, A., Telisheva, Z., et al. (2023). "QWriter: technology-Enhanced alphabet acquisition based on reinforcement learning," Extended Abstracts of the 2023 CHI conference on human factors in computing systems. Hamburg, Germany, 1–7. doi:10.1145/3544549.3585611
- Shuster, J. J. (2010). Empirical vs natural weighting in random effects meta-analysis. *Statistics Med.* 29, 1259–1265. doi:10.1002/sim.3607
- Singh, D. K., Kumar, M., Fosch-Villaronga, E., Singh, D., and Shukla, J. (2023). Ethical considerations from child-robot interactions in under-resourced communities. *Int. J. Soc. Robotics* 15, 2055–2071. doi:10.1007/s12369-022-00882-1
- So, W.-C., Wong, M. K.-Y., Lam, C. K.-Y., Lam, W.-Y., Chui, A. T.-F., Lee, T.-L., et al. (2018). Using a social robot to teach gestural recognition and production in children with autism spectrum disorders. *Disabil. Rehabil. Assistive Technol.* 13, 527–539. doi:10.1080/17483107.2017.1344886
- *Steinhausser, S. C., Riedmann, A., Schaper, P., Guthmann, E., Pfister, J., Schmitt, K., et al. (2022). "Second language learning through storytelling with a social robot – an online case study," in Proceedings of the 2022 31st IEEE international conference on robot and human interactive communication (RO-MAN). Naples, Italy, 67–74. doi:10.1109/RO-MAN53752.2022.9900628
- Sukardi, Setyawan, H., Risfendra, R., Yanto, P., and Tri, D. (2024). Effectiveness of robotic technology in vocational education: a meta-analysis. *Int. J. Inf. Educ. Technol.* 14, 521–532. doi:10.18178/ijiet.2024.14.4.2073
- *Suzuki, K., and Kanoh, M. (2015). "Effectiveness of a robot for supporting expression education," in Proceedings of the 2015 conference on technologies and applications of artificial intelligence. Tainan, Taiwan, 498–501. doi:10.1109/TAAL.2015.7407119
- *Suzuki, K., and Kanoh, M. (2017). Investigating effectiveness of an expression education support robot that nods and gives hints. *J. Adv. Comput. Intell. Intelligent Inf.* 21, 483–495. doi:10.20965/jaciii.2017.p0483
- Tabone, W., and De Winter, J. (2023). Using ChatGPT for human-computer interaction research: a primer. *R. Soc. Open Sci.* 10, 231053. doi:10.1098/rsos.231053
- Taheri, A., Meghdari, A., Alemi, M., Pouretamad, H., Poorgoldooz, P., and Roohbakhsh, M. (2016). "Social robots and teaching music to autistic children: myth or reality?," in *Social robotics*. Editors A. Agah, J. J. Cabibihan, A. Howard, M. Salichs, and H. He (Cham: Springer), 541–550. *ICSR 2016*. doi:10.1007/978-3-319-47437-3_53
- *Tanaka, F., and Matsuzoe, S. (2012). Children teach a care-receiving robot to promote their learning: field experiments in a classroom for vocabulary learning. *J. Human-Robot Interact.* 1, 78–95. doi:10.5898/JHRI.1.1.Tanaka

- *Tang, H. K., Smakman, M., De Haas, M., and Van den Berghe, R. (2023). "L2 vocabulary learning through lexical inferencing stories with a social robot." Companion of the 2023 ACM/IEEE International Conference on Human-Robot Interaction. Stockholm, Sweden, 526–530. doi:10.1145/3568294.3580140
- Tang, R., Zhang, X., Ma, X., Lin, J., and Ture, F. (2024). "Found in the middle: permutation self-consistency improves listwise ranking in large language models," in *Proceedings of the 2024 conference of the north American chapter of the Association for computational linguistics: human language technologies (volume 1: long papers)*. Editors K. Duh, H. Gomez, and S. Bethard (Mexico City: Association for Computational Linguistics), 2327–2340. doi:10.18653/v1/2024.naacl-long.129
- Tozadore, D. C., Cezayirlioglu, M., Wang, C., Bruno, B., and Dillenbourg, P. (2023). "Immediate effects of short-duration wellbeing practices on children's handwriting and posture guided by a social robot," in *Proceedings of the 2023 32nd IEEE international conference on robot and human interactive communication (RO-MAN)*, Busan, Republic of Korea, 1475–1480. doi:10.1109/RO-MAN57019.2023.10309605
- *Tsai, C.-C. (2019). Adapting an anthropomorphized robot for enhancing EFL learning motivation and performance at an elementary school in Taiwan. *New Educ. Rev.* 58, 22–35. doi:10.15804/ner.19.58.4.02
- Valentí Soler, M., Agüera-Ortiz, L., Olazarán Rodríguez, J., Mendoza Rebolledo, C., Pérez Muñoz, A., Rodríguez Pérez, L., et al. (2015). Social robots in advanced dementia. *Front. Aging Neurosci.* 7, 133. doi:10.3389/fnagi.2015.00133
- *Van den Berghe, R. (2019). "Comparing children's L2 word learning between a vocabulary training without a peer, with a child peer, or with a humanoid robot," in *Social robots as second-language tutors for young children: challenges and opportunities*. Editor R. van den Berghe, 99–131. (Doctoral dissertation, Utrecht University) Available online at: <https://dspace.library.uu.nl/handle/1874/381140>.
- *Van Minkelen, P., Gruson, C., Van Hees, P., Willems, M., De Wit, J., Aarts, R., et al. (2020). "Using self-determination theory in social robots to increase motivation in L2 word learning," *Proceedings of the 2020 ACM/IEEE international conference on human-robot interaction*. Cambridge, UK, 369–377. doi:10.1145/3319502.3374828
- *Velentza, A.-M. (2024a). "Experiment IV: human robot co-tutoring," Human-robot interaction: The role of humanoid robots in learning process and cognitive tasks. Editor A.-M. Velentza, 71–92. (Doctoral dissertation, University of Macedonia). Available online at: <http://dspace.lib.uom.gr/handle/2159/30991>.
- *Velentza, A.-M. (2024b). "Experiment VI: kindergarten application: traffic safety," in *Human-robot interaction: the role of humanoid robots in learning process and cognitive tasks*. Editor A.-M. Velentza, 110–124. (Doctoral dissertation, University of Macedonia). Available online at: <http://dspace.lib.uom.gr/handle/2159/30991>.
- *Velentza, A.-M., Fachantidis, N., and Lefkos, I. (2021). Learn with surprise from a robot professor. *Comput. and Educ.* 173, 104272. doi:10.1016/j.compedu.2021.104272
- *Velentza, A.-M., Kefalouka, E., and Fachantidis, N. (2025). Socially assistive robot in sexual health: group and individual student-robot interaction activities promoting disclosure, learning and positive attitudes. *Int. J. Soc. Robotics* 17, 1161–1176. doi:10.1007/s12369-025-01284-9
- *Verhelst, E. (2023). Adaptive second language tutoring through generative AI and social robots MSc thesis. Ghent University. Available online at: https://libcatalog.ugent.be/permalink/32RUG_INST/ton6o8/alma990031501940409161.
- *Verhelst, E., Janssens, R., Demeester, T., and Belpaeme, T. (2024). "Adaptive second language tutoring using generative AI and a social robot," in *Companion of the 2024 ACM/IEEE international conference on human-robot interaction*. Boulder, CO, 1080–1084. doi:10.1145/3610978.3640559
- Vickers, A., Goyal, N., Harland, R., and Rees, R. (1998). Do certain countries produce only positive results? A systematic review of controlled trials. *Control. Clin. Trials* 19, 159–166. doi:10.1016/S0197-2456(97)00150-5
- *Vogt, P., Van den Berghe, R., De Haas, M., Hoffman, L., Kanero, J., Mamus, E., et al. (2019). Second language tutoring using social robots: a large-scale study. *Proceedings of the 2019 14th ACM/IEEE International Conference on Human-Robot Interaction*, Daegu, South Korea, 497–505. doi:10.1109/HRI.2019.8673077
- *Walker, E., Giroto, V., Kim, Y., and Muldner, K. (2016). "The effects of physical form and embodied action in a teachable robot for geometry learning," in *Proceedings of the 2016 IEEE 16th international conference on advanced learning technologies*. Austin, TX, 381–385. doi:10.1109/ICALT.2016.129
- *Wang, Y. H., Young, S. S. C., and Jang, J.-S. R. (2013). Using tangible companions for enhancing learning English conversation. *J. Educ. Technol. and Soc.* 16, 296–309. Available online at: <https://www.jstor.org/stable/jeductechsoci.16.2.296>.
- Wang, K., Sang, G.-Y., Huang, L.-Z., Li, S.-H., and Guo, J.-W. (2023). The effectiveness of educational robots in improving learning outcomes: a meta-analysis. *Sustainability* 15, 4637. doi:10.3390/su15054637
- Wang, X., Wei, J., Schuurmans, D., Le, Q., Chi, E., Narang, S., et al. (2023). "Self-consistency improves chain of thought reasoning in language models," in *Proceedings of the international conference on learning representations*. Kigali, Rwanda. Available online at: <https://openreview.net/pdf?id=IPL1NIMMrw>.
- Wedenborn, A. (2015). A physical robot's effect on vocabulary learning MSc thesis. KTH Royal Institute of Technology. Available online at: <https://kth.diva-portal.org/smash/record.jsf?pid=diva2%3A886159&dsid=6994>.
- *Wei, C.-W., Kao, H.-Y., Wu, W.-H., Chen, C.-Y., and Fu, H.-P. (2021). The influence of robot-assisted learning system on health literacy and learning perception. *Int. J. Environ. Res. Public Health* 18, 11053. doi:10.3390/ijerph182111053
- Woo, H., LeTendre, G. K., Pham-Shouse, T., and Xiong, Y. (2021). The use of social robots in classrooms: a review of field-based studies. *Educ. Res. Rev.* 33, 100388. doi:10.1016/j.edurev.2021.100388
- Wu, X., and Li, R. (2024). Effects of robot-assisted language learning on English-as-a-foreign-language skill development. *J. Educ. Comput. Res.* 62, 790–814. doi:10.1177/07356331231226171
- *Wu, W.-C. V., Wang, R.-J., and Chen, N.-S. (2015). Instructional design using an in-house built teaching assistant robot to enhance elementary school English-as-a-foreign-language learning. *Interact. Learn. Environ.* 23, 696–714. doi:10.1080/10494820.2013.792844
- Yang, J., Vindolet, C., Guadarrama Olvera, J. R., and Cheng, G. (2024). On the impact of robot personalization on human-robot interaction: a review. *arXiv*. doi:10.48550/arXiv.2401.11776
- *Yin, J., Guo, W., Zheng, W., Ren, M., Wang, S., and Jiang, Y. (2024). The influence of robot social behaviors on second language learning in preschoolers. *Int. J. Human-Computer Interact.* 40, 1600–1608. doi:10.1080/10447318.2022.2144828
- *Yoshizawa, R., Jimenez, F., and Murakami, K. (2020). Proposal of a behavioral model for robots supporting learning according to learners' learning performance. *J. Robotics Mechatronics* 32, 769–779. doi:10.20965/jrm.2020.p0769
- *Yueh, H.-P., Lin, W., Wang, S.-C., and Fu, L.-C. (2020). Reading with robot and human companions in library literacy activities: a comparison study. *Br. J. Educ. Technol.* 51, 1884–1900. doi:10.1111/bjet.13016
- *Yuliani, S., Rizqiani, D., and Linarta, A. (2024). Words and bots: an empirical experiment vocabulary mastery with robotics in young children. *J. Lang. Lang. Teach.* 12, 392–403. doi:10.33394/jollt.v12i1.9372
- *Zachrisson, E. (2021). *Effects of the social robot Furhat on vocabulary learning Summer school report*. KTH Royal Institute of Technology. Available online at: https://static1.squarespace.com/static/56b6357e01dbaea0266fe701/t/61141feffed1f76a0722ba9c/1744538904776/Elsa_Zachrisson.pdf.
- *Zhanatkyzy, A., Telisheva, Z., Amirova, A., Oralbayeva, N., Aimysheva, A., and Sandygulova, A. (2022). "Moveable Älpi: the Montessori method for robot-assisted alphabet learning," in *Social robotics*. Editors F. Cavallo, J.-J. Cabibihan, L. Fiorini, A. Sorrentino, H. He, X. Liu, et al. (Cham: Springer), 114–123. ICSR 2022. doi:10.1007/978-3-031-24670-8_11
- Zhang, D., Freemantle, N., and Cheng, K. K. (2011). Are randomized trials conducted in China or India biased? A comparative empirical analysis. *J. Clin. Epidemiol.* 64, 90–95. doi:10.1016/j.jclinepi.2010.02.010
- *Zhang, X., Breazeal, C., and Park, H. W. (2023). "A social robot reading partner for explorative guidance," in *Proceedings of the 2023 ACM/IEEE international conference on human-robot interaction*. Stockholm, Sweden, 341–349. doi:10.1145/3568162.3576968
- *Zhexenova, Z., Amirova, A., Abdikarimova, M., Kudaibergenov, K., Baimakhan, N., Tleubayev, B., et al. (2020). A comparison of social robot to tablet and teacher in a new script learning context. *Front. Robotics AI* 7, 99. doi:10.3389/frobt.2020.00099