Evaluating Research on Social Robots for Individuals with Intellectual Disability

Lonneke Westra
University of Twente
PO Box 217, 7500 AE Enschede
the Netherlands
l.p.m.westra@student.utwente.nl

ABSTRACT

Social robots are becoming more common in today's society. These robots are used in many application areas that require human-robot interaction. There have been various examples of robots being used in activities of children with special needs, such as intellectual disability (ID). Despite the benefits shown of previous research, it still occurs that some robots are rarely used in real-life settings after the relevant study ends. Therefore, it should be studied how adequately current research on social robots for people with ID is performing.

This paper provides a literature review on existing social robotic technologies utilized for intellectually disabled individuals to identify relevant research gaps in existing literature. This research answers the question how current research on social robots tested on intellectually disabled individuals can be improved. In relation to this, this research will unveil strengths and weaknesses of existing technologies in the implementation for intellectual disabled individuals. Firstly, existing literature on social robots reviewed for and tested on people with ID was obtained by performing a search query. Thereafter, a framework was used to specify three evaluation variables ('technical', 'usability' and 'effectiveness of interaction') to evaluate the included papers. In this research, 15 studies were included of which 2 discussed technical evaluation, 10 studies covered at least one of the usability evaluation aspects and 2 studies used a mixed method of quantitative and qualitative research to evaluate the effectiveness of interaction. A few papers performed very well, but in the majority of the papers, there is room for improvement in different aspects. Technical aspects were not thoroughly or not at all evaluated, in many of the included papers, it could not be detected how a robot's aspects influence the outcome of the human-robot interaction and most research solely utilized quantitative or qualitative research instead of both. This paper can be used in a practical matter by robotic corporations or engineers in developing future robot technologies for people with ID. Furthermore, this paper will act as a base for further research on the efficiency of social robots within the area of ID for both technical and social review.

Keywords

Intellectual disability, ID, learning disability, mental retardation, social robot, human-robot interaction, evaluation, literature review

1. INTRODUCTION

Social robots are increasingly common in today's society [3]. These robots help people as capable partners rather than as tools and are believed to be of greatest use for applications in entertainment, education and health care because of their potential to be perceived as trusting, helpful, reliable and engaging [16]. This paper provides a literature review on existing social robotic technologies utilized for individuals with intellectual disability to identify relevant research gaps in existing literature.

Social robots are used in many application areas that require human-robot interaction (e.g. entertainment, rehabilitation, therapy, etc.). For example, one type of assistive robots is social robots used for educational purposes, such as playful learning as is presented in [9]. Within the area of entertainment, several toy robots were reviewed, such as [3] and [8]. Social robots could, furthermore, assist children managing chronic illness through education and encouragement to perform healthy behaviors, help distract children coping with acute medical procedures or provide companionship and comfort [12].

A specific target group for the use of social robots could be children with special needs, such as intellectual disability. Intellectual disability, previously referred to as mental retardation in the U.S or learning disability in the U.K., is a neurodevelopmental disorder characterized by significant limitations in both intellectual functioning and in adaptive behavior, which covers many everyday social and practical skills. This disability originates before the age of 18 [1]. Intellectually disabled people will vary considerably in their level of ability which is usually described in four levels: mild, moderate, severe and profound. In theory, these levels are defined by IQ but practically it is more useful to think of people differing in the level of support they require [24].

People with ID have shown interest in using new technology. Past research has looked at how they use various social media platforms to achieve different goals, use voice-activated interfaces to access information, use portable electronic assistive technology to become more independent and engage individually with social robots [16].

In recent years there have been several examples of robots being used for children with ID in play activities for therapeutic or educational purposes. These robots have shown to be useful in promoting spontaneous play for children with a developmental disorder, such as ID, engaging them in playful interactions [13]. Moreover, it appeared that social robots could have a positive impact on people with intellectual disability such as motivating them to take part in social or group activities [3] or increasing children's communication behavior [24]. Furthermore, social robots have positively influenced children for educational purposes [25,27] or in robot-assistive therapy [8-11, 32].

1.1 Problem Statement

Various available social robotic technologies have been tested for those with learning difficulties and/or intellectual disabilities. Despite the benefits shown of previous research, it still occurs that robots are rarely used in real-life settings after the relevant study ends [30]. Regarding this interesting contradiction, one could question itself: is current research on social robots for people with ID performing adequately? This can be examined by means of a theoretical framework. An existing framework [14] that evaluates the effectiveness of a robotic toy for children with special needs was used for this research. In this framework [14], three evaluation variables are proposed: 'technical', 'usability' and 'effectiveness of interaction'.

Many existing technologies or literature have been reviewed for people with Autism Spectrum Disorder (ASD) or other learning disabilities [6], [12], [15], [23], [29]. Yet, a thorough evaluation of multiple existing technologies within the area of ID has not been performed.

The end purpose of this literature review is to identify relevant research gaps in existing literature on social robots for intellectually disabled individuals, as discussed in the paragraph above.

1.2 Research Question

Considering the problem statement defined above, the following research question with its subcomponents has been formulated.

Main research question: How can current research on social robots tested on intellectually disabled individuals be improved?

i. How does current research on social robots tested on intellectually disabled individuals perform on the technical evaluation aspect?

ii. How does current research on social robots tested on intellectually disabled individuals perform on the usability evaluation aspect?

iii. How does current research on social robots tested on intellectually disabled individuals perform on the effectiveness of interaction evaluation aspect?

To answer the main research question, three sub-questions were formulated based on the framework [14]. Firstly, existing literature on social robots on people with ID was obtained by performing a search query. The included papers were thoroughly analyzed based on the 'technical', 'usability' and 'effectiveness of interaction variables' of the framework [14]. Based on the answers to these sub-questions, the main research question could be answered. A conclusion was thus identified and recommendations for future research were provided.

In the first part of the paper, the method of this review, including the search strategy, eligibility criteria, terminology and definition, data extraction and the used framework are elaborated. The findings and evaluation of the review are presented in a table and discussed more detailed in the results section. The results will be elaborated in the discussion section. Additionally, this paper's strengths and weaknesses are substantiated. Furthermore, recommendations for future research are discussed.

2. RESEARCH METHOD

To answer the research question and its subcomponents, an online search strategy of relevant databases was performed, including cross-checking of reference lists.

2.1 Search Strategy

The literature research was performed on 30 May 2020, by utilizing Google Scholar and Scopus. The search was solely limited to papers published in English. To identify relevant papers for this research, a search query is composed containing keywords of two fields. The "technical" field is included to find the correct robotic technology and the "social" field contains keywords that should match with intellectual disability (Table 1). These keywords were combined with the OR boolean operator between all words and AND boolean operator between the groups.

All papers were reviewed by the author and categorized stepwise in multiple phases. This was performed by screening the titles, abstracts and keywords against the eligibility criteria. Thereafter, the research method and conclusion of selected papers were screened, and for the final papers, full-texts were acquired and screened to extract their data.

Table 1. Search Query Keywords

Technical	Social
Human robot	Intellectual disability
Social robot	Learning disability
Humanoid robot	Mental retardation

2.2 Eligibility Criteria

Articles of all time that are written in English were included in the research. Included publications described or evaluated the design, existence, or use of one or multiple social robots for people with ID. Another criterium was that the publications should be an evaluation or review of a robot that was tested on one or multiple participant(s) with intellectual disability. Articles that solely focused on ethics, approaches, or methodologies on the uptake of social robots were excluded. Additionally, databases or proposed robot systems were not considered as social robots and excluded from this research. Furthermore, literature or systematic reviews on social robots were not included, though their reference lists were cross-checked for papers missed in the original search.

2.3 Terminology and definition

A social robot is an artificial intelligence system that is designed to interact with humans and other robots [21]. A social robot should have at least one feature to interact with humans and give feedback to be considered in this research. All physical types of social robots, humanoid, animaloid, non-humanoid are interpreted as social robot.

An IQ test score of around 70 or as high as 75 indicates a limitation in intellectual functioning [1]. Different levels of ID are defined as the following: mild (IQ 52-69), moderate (IQ 36-51), severe (IQ 20-35) and profound (IQ 19 or below) [28].

2.4 Data Extraction and Framework

Data were extracted by the author using the following variables: name of the robot, type of robot, the purpose of the study, target population, sample size, age of participants and the gender of participants. Furthermore, the type of intervention and type of measurement were extracted from the literature.

Moreover, this research examined how each study measured the performance of a social robot. A framework [14] that evaluates the effectiveness of a robotic toy for children with special needs was used to determine the evaluation variables for the included literature. The authors of the framework [14] state that researchers need to consider the technical, mechanical, and safety aspects, but also need to extend the evaluation to the social role the robot might have and its long term effect on its users [14]. This framework [14] includes the application of social robots on children with ID. Other frameworks were analyzed, nevertheless, these papers focused merely on the implementation of a robot [7] or on human-robot interaction [5] without including the aspect of children with intellectual disability. Considering all of the above, this framework [14] applies best to this research.

From [14], the following evaluation variables were determined: 'technical', 'usability' and 'effectiveness of interaction'. Where technical refers to the technical status of the robot that should be examined by various tests such as safety and compliance tests. Usability refers to the acceptance of the robot as an integral element of the play activity. For an optimal outcome, the usability evaluation needs to be conducted first with secondary users and subsequently with people from the target user groups. Lastly, the effectiveness of interaction represents the effect that the interaction with a robotic toy might have on people with ID. This evaluation should rely on a mixed method of quantitative and qualitative research because in [14], the authors state that by combining the qualitative and quantitative methods, the weakness of every single method can be overcome and a more complete picture of an individual's achievements and possible progress can be obtained. A thorough elaboration of the evaluation variables can be found in Appendix A.

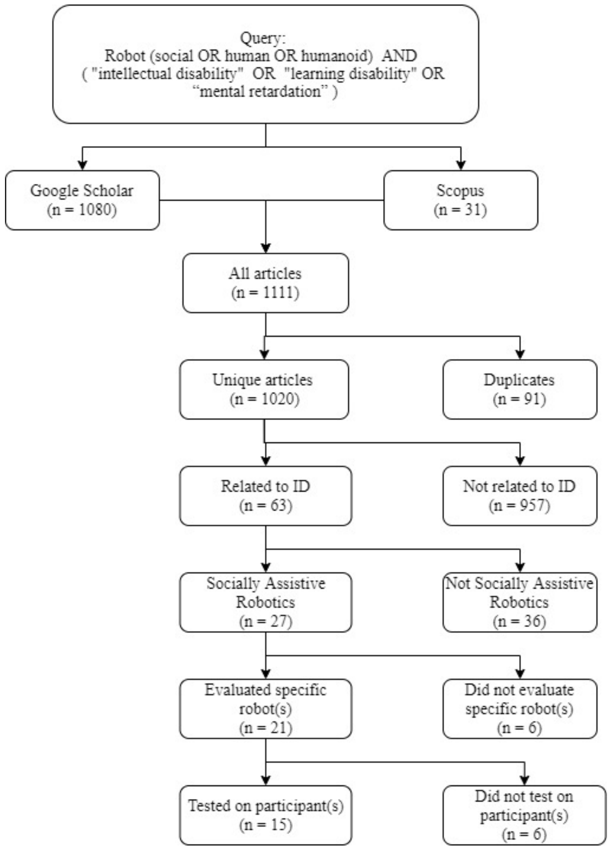

Figure 1. Search Strategy Flow Diagram

3. RESULTS

3.1 Selection Process

The performance of the search query resulted in a total of 1111 articles of which 1080 were obtained in Google Scholar and 31 in Scopus. After the elimination of duplicates, 1020 unique articles were remaining to be screened. The title and abstract of each article were thoroughly screened according to all eligibility criteria which resulted in a selection of 27 papers for which full-texts were acquired. After screening the full-texts, a final selection of 15 papers (Figure 1), in which 18 robots are reviewed, was obtained. All social robots are tested and evaluated on at least one person with intellectual disability. All 15 papers were either published online or part of a conference.

The results (Table 2) are described in sections aligning with the aspects of evaluation. The articles are firstly categorized by the robot's name and secondly by ascending publication date of the paper.

3.2 Research Variables Evaluation

3.2.1 Target population

Children with ID comprised the largest target population group, which applied to 12 papers (n = 12). Furthermore, this group of 12 children contains a subset of pupils or students (n = 4). Another common target group was people with Autism Spectrum Disorder (ASD) and a level of intellectual disability (n = 6) and people with Down Syndrome (n = 4). Less common target groups were adults with ID (n = 2) and people with ID and visual impairments (n = 1). The age of children varied between children/students aged 3-19 years. Besides children, there also was a group of adults aged 26-31 years and a group of elderly aged 59-70 years.

The number of participants varied between large target groups that contained 28 participants (n = 28), smaller target groups (between n = 2 and n = 9) and n =1 studies. The majority reported small sample sizes.

3.2.2 Type and purpose of the robot

Eight different robots were reviewed ranging in stages of development from concept formulation through to commercially available models. NAO was the most common (n = 8), followed by IROMEC (n = 3), followed by Paro (n = 2). The other robots identified were Bee-Bot, Cozmo, iRobiQ, KASPAR and Lego Mindstorms. NAO and KASPAR are humanoid robots (Figure 2), Paro is an animaloid robot (Figure 2) and the others are non-humanoid robots (Figure 3).

Figure 2. Robots NAO (left, right), and PARO (center) [26]

Figure 3. Lego Mindstorms KRAZ 3

The purpose of the robots was most notably to act as an entertaining robot with whom the participants could play, interact and engage. A further purpose was the intervention of robot-assistive therapy (n = 4). The robots were utilized to engage the child during standard treatment to assess playfulness or improve social skills or children's communicative behavior.

Another application was game-based versus conversation-based human-robot interaction aimed at people with ID and visual impairments. Moreover, one paper discussed a user-centered design approach to develop a robot that is able to engage in meaningful interaction with different typologies of disabled children. Furthermore, robots were used educationally (n = 4) to help pupils achieve learning objectives or facilitate the inclusion of children with special needs.

3.2.3 Intervention and Measurement

Most studies (n = 13) studied the short-term effect of robot intervention. The duration of the studies varied between one single session or multiple sessions up to three months. [26] and [27] studied the long-term effect of robot intervention. An evaluation was performed with lengths up to 24 months.

Several studies, [3, 13, 17, 19], assessed different play scenarios. Two studies [19, 20] tested the robot not only on primary users but also on secondary users.

Two studies, [8, 26], evaluated the effectiveness by means of a mixed method of quantitative and qualitative research. Both studies used a questionnaire as a quantitative method and semi-structured interviews as a qualitative method.

One study, [4], solely used quantitative research by conducting various existing standardized tests as pre- and post-assessments on the participants. The remaining studies solely performed qualitative research. 11 studies performed a data analysis on the participants through observations of audio and/or video recordings. Interviews on participants or secondary users were conducted by 9 studies. In two studies, [13, 32] a smiley-scale was used, on top of the interviews, in which the participants could rate their feelings and emotions towards the robot.

3.3 Framework Variables Evaluation

3.3.1 Technical Evaluation

It appeared that only two studies [19, 26] out of 15 discussed technical evaluation aspects. According to the framework [14] adequacy, safety and compliance to national standards of the robot should be studied. One study [19] thoroughly discussed IROMEC's technical components and implemented two main configurations: vertical and horizontal. In the vertical configuration, IROMEC can resemble the shape of the human form. In the horizontal configuration, IROMEC is used as a non-humanoid robot that allows for unconstrained interactions which are suited for children with ASD. The other study that tested NAO and Paro on students with ID and ASD [26], created a questionnaire where safety and robot's durability was evaluated. Paro seems to be more durable and safer than NAO. Safety, besides effectiveness and appearance, seemed to be the most important factor to the participants [26]. None of the other studies mentioned either technical configurations or compliance to international standards or safety in their research paper.

3.3.2 Usability Evaluation

From the 15 studies, 10 studies covered at least one of the usability evaluation aspects that were mentioned in the paper of the framework [14]. Firstly, a study that used a small AI toy robot Cozmo [3], specifically evaluated the quality of the robot's visual feedback, because the eyes of the robot display human emotions. Feelings and emotions were attributed towards the robot when Cozmo displayed emotions with his animated eyes or engaged with the participants by singing or celebrating when a participant won a game. Random activities such as singing, hiccups, snoring or Cozmo referring to participants by name made participants laugh.

Furthermore, a humanoid (IROMEC) with a non-humanoid robot (KASPAR) were compared [17] and it appeared that humanoid and child-like characteristics such as embodied facial and corporeal movements imitating the human expression of emotions of KASPAR seemed to stimulate the child much more to interact with the experimenter and the robot. Another study [2] compared a humanoid (NAO) and non-humanoid (Lego Mindstorms) as well but focused more on the variety of functions and control**.** It appeared in this study that by using a tablet, it is easier to control the robot than by using a remote control.

Moreover, a study that tested educational humanoid robot iRobiQ on six students [20] focused more on touching, bodily feedback, sound, head and screen feedback. Touching the robot's arm drew all student's attention and interaction. Also, body movements were effective and attracted attention. Head feedback was the least perceived type of feedback. There was no response to screen feedback. The study [20] concluded that students prefer human interaction involved with emotion and movement by robot.

Focusing on speech and sound effects, in a study that applied game-based and conversation-based human-robot interaction with NAO [13], it appeared that four participants preferred a human voice over a robotic voice. Additionally, another study that researched the long-term impact of NAO [26] concluded that the robot's small size, its simplified human-like form but also its monotonous voice reduced sensory overload in students. The study that tested iRobiQ [20] concluded that the robot's "applaud" effect was observed to be effective. However, it appeared that the use of a realistic human voice was more effective and remarkable compared to audio feedback, thus sound effects together with humanoid robot feedbacks would be most effective.

Lastly, the study that tested a user-centered design approach [19] concluded that IROMEC's non-humanoid appearance and behavior do not seem to evoke an agent with own inner states, which radically reduces the potential of the robot as a mediator of social exchanges.

3.3.3 Effectiveness of Interaction Evaluation

The study that tested Cozmo [3] concluded that social robots can be used to motivate people with ID to take part in social or group activities. It appeared that participants' engagement with each other increased with the introduction of Cozmo. The study with iRobiQ [20] concluded that students with ID prefer human interaction involved with emotion and movement by the robot. Also, the robot's feedbacks drew the participants' attention and increased their motivation in learning.

Furthermore, two studies compared a humanoid with a non-humanoid robot [2, 17]. In one study [17], the results indicated that the participant was more interactive with the experimenter and the robot during the sessions with humanoid KASPAR than with non-humanoid IROMEC. The study [17] furthermore concluded that children with Down are good in non-verbal communication, so more human-like social characteristics means more interaction with the robot. However, in the other study [2] it appeared that three out of four participants had a higher percentage of engagement with the non-humanoid Lego Mindstorms than with humanoid NAO and concluded that Lego Mindstorms can be more or just as engaging as NAO.

Robots were also implemented in assistive therapy for children with ID. In a study that used IROMEC [8], occupational therapists were positive about the robot and appreciation for child and therapist, but less positive about the therapeutic added value for the involved children, to the limited match between participating children's needs and characteristics of the robot as a toy. In a research that studied the intervention of Paro in assistive therapy [31], results indicated that robot-based animal-assisted therapy does not have clear beneficial effects on alertness and mood in adults with moderate to severe ID. However, positive interactions with a robot seal could be of therapeutic value in itself. Conti, et al. 2018 published two papers [10, 11] in which NAO was integrated into standard treatment of therapy. In their preliminary study [10], it appeared that the score of the imitative level significantly increased in 4 out of 6 children with ID. Conti, et al. 2018 concluded that robotic-assistive therapy can be successfully integrated into the standard treatment of autistic children with intellectual disability when the interaction is adapted to the individual level. However, it is needed to find more advanced solutions and approaches for persons with profound ID. Additionally, the results of their other study [11] indicated that all participants benefited from robot-assistive therapy, except for children with profound ID thus Conti, et al., 2018 concluded that lower IQ leads to less engagement with the robot.

Moreover, Shamsuddin, et al., 2012 studied the impact of a robot-based intervention by measuring autistic behavior of children with ID and ASD [24]. Four out of the five children exhibited a decrease of autistic behavior when the robot was executing human-robot interaction modules during the single session of child-robot interaction. The study [24] concluded that NAO was able to attract the children's attention, keep each child engaged with the robot during interaction and hence give a positive impact on the children's communication behavior. Follow-up research was performed with NAO in a classroom setting [25] wherein 5 out of the 6 children responded well during child-robot interaction. Class interaction portrayed a reduced percentage in stereotyped behavior during the child-robot interaction, yet strong proof to link the children's responses with their IQ levels needs to be supported with larger samples of children in the future.

Two studies [3, 26] mentioned difficulties that arose during the research. One study, that tested Cozmo, [3] concluded that speech impairment of the intellectually disabled participants made it difficult for Cozmo to understand the children. Vision impairment made it difficult for the participants to see Cozmo's color and cubes during a game which resulted in not being able to understand the game. Two limitations of NAO were mentioned in another study [26] which were the robot's cost and fragility. Participants also mentioned that a minimum level of cognition and verbal communication is required to maintain interest and motivation when interacting with the NAO robots. This means that the potential benefits for non-verbal students are limited to communication via touch, body movement or flashcards.

Additionally, this study [26] compared a humanoid NAO with an animaloid Paro robot on long-terms and from qualitative evaluation, NAO and Paro were both considered predictable which has a positive effect on children with ID and ASD. The study [26] concluded that NAO and Paro's mistakes and limitations created a safe environment where students felt comfortable to engage in new activities even if they were not proficient. Lastly, another study [27] also studied the long-term effects of NAO and concluded that NAO robots had enriched the learning experiences of students, as well as the robots had become rewarding social partners.

4. DISCUSSION

4.1 Summary of Evidence

To evaluate how social robots are currently reviewed and tested on people with ID, 15 papers were identified which is, compared to the average of literature reviews, a rather small number. In the first step of the search strategy (Figure 1), only 63 studies remained that were related to ID. The fact that such a large number of studies were not related to ID can be explained by the following. The term 'learning disability' is used in the U.K. for ID but can also represent other learning disabilities such as ADHD, dyslexia or dyscalculia. Moreover, papers were identified that considered Neurodevelopmental Disorders (NDD) or Profound and Multiple Learning Disabilities (PMLD). These terms include several disorders of which intellectual disability is one, however, it could not be guaranteed that one of the study's participants had ID. Lastly, the majority of the papers were related to solely the developmental disorder autism. A few included papers performed very well on the aspects of the framework [14], but in the majority of the papers, there is room for improvement in different aspects.

4.1.1 Technical Evaluation Performance

A user-centered design study with IROMEC [19] concluded that the vertical configuration, where the robot resembles a non-humanoid robot, is more suited for children with ID and ASD than the horizontal configuration, where the robot resembles the shape of a human form because the non-humanoid allowed for unconstrained actions. In this study [19] it was mentioned that children with ASD have difficulties interpreting facial expressions and other social cues in social interaction which could explain the participants' preference. In addition, a study with humanoid NAO in therapy [10] concluded that the robot had a positive effect on the communication behavior of children with ID and ASD, because NAO contains simpler features compared to real humans. Moreover, the long-term effect of NAO and Paro was studied on students with ID and ASD [26] where it was concluded that the robot's predictability had a positive effect on the students. Considering all of the above, it can be concluded that children with ID and autism have more engagement with the robot and improvement of their skills when interacting with a robot that is predictable with simple features.

As mentioned in the results section, in the study with NAO and Paro [26], participants mentioned that safety was one of the three most important aspects of a social robot. This study [26] concluded that the robots' limitations created a safe environment where students felt comfortable engaging in new activities even if they were not proficient. Thus, it can be explained that safety is considered an important factor.

4.1.2 Usability Evaluation Performance

It appeared that human-like characteristics have a positive effect on engagement with the robot [17]. This study tested IROMEC and KASPAR on a child with Down syndrome. It was mentioned that children with Down are good in non-verbal communication. Thus, this research [17] concluded that more human-like social characteristics means more interaction with the robot.

Moreover, one study [20] focused on many usability aspects of the iRobiQ robot and mentioned that head feedback was the least perceived type of feedback. Yet, it is mentioned that according to literature, face feedbacks are very important. The paper [20] justifies this with the fact that the head size was too small and the combined use of face and body feedback caused the physical feedback may be dominant.

Furthermore, a human voice is preferred over a robotic voice in the study wherein game-based and conversation-based human-robot interaction with NAO was applied [13]. On the contrary, another study [26] concluded that NAO's monotonous voice reduced sensory overload in students. This contradiction can be clarified since NAO had a different purpose in both studies. The first study [13] included two sessions with play scenarios and the other study [16] measured the long-term effect with multiple sessions per week for 24 months to enhance learning.

4.1.3 Effectiveness of Interaction Evaluation Performance

One study [8] mentioned that the robot's intervention in therapy did not add value for the children involved. Similarly, another study [31] discussed that robot-based animal-assisted therapy does not have clear beneficial effects on alertness and mood for the participants. Contrarily, two other studies by Conti, et al., 2018 concluded that NAO does add value in robot-assistive therapy for children of all levels of ID, except profound ID. The discrepancy of the studies can be justified by the difference in target population and robot type. One study tested an animaloid Paro on elderly with ID aged 59-70 [31] whereas the others [8, 10-11] tested on children. One study [8] tested on children with ID and the others by Conti, et al., 2018 on children with ID and ASD. As discussed in 4.1.1, autism plays an important factor in the effect of the robot on children with ID and autism, so therefore it is not comparable with the other studies.

Lastly, two papers compared a humanoid with a non-humanoid robot [2, 17]. Humanoid KASPAR believed to have more interaction with the participant than non-humanoid IROMEC [17]. Otherwise, non-humanoid Lego Mindstorms appeared to be just as or even more engaging than humanoid NAO [2]. The first study [17] tested the robots on one participant with Down syndrome while the second study [2] tested the robots on 2 children with Down syndrome and 2 children with another level of ID. Two children with Down and one with ID a significantly higher percentage engagement with the Lego Mindstorms than NAO and the other one with ID had no difference in percentages [2]. So, it could be that Down syndrome plays a factor in the preference between a humanoid and a non-humanoid robot. However, the study with children with ID and Down syndrome [2] tested the robots on pupils at a special needs school as a learning-supportive robot while the other study [17] tested the robot on ten different play scenarios. This difference in purpose justifies the inconsistency between the conclusions of both studies. Lastly, both studies [2, 17] reported very small sample sizes ($n = 1$ and $n = 4$), thus robots should be studied with larger sample sizes before adequate conclusions can be provided.

4.1.4 Strengths and limitations of included papers

Technical aspects were barely included in the reviews nor in tests regarding social robots for people with ID. This is reasonable on one hand since the robots that are being used in these studies are already designed and tested on technical aspects in previous studies. It would not be ideal if the robot's safety still needs to be evaluated while it is already being used on participants. However, it should still be included in some matter since technical components play a part in the evaluation and the uptake of a social robot.

Moreover, in the studies that are currently available, it appeared that the researchers put their main focus on the participant with ID in terms of measuring attention, behavior and how their characteristics can influence the engagement with the robot. For example, a long-term effect study with NAO and Paro [26] measured mood and alertness of the participant in general and not during the specific actions of the robot. It was solely concluded that robot-assistive therapy can be successfully integrated into the standard treatment of autistic children with ID based on the comparison of the child's behavior, without including a discussion section to elaborate how the robot affected the children.

According to the framework [14], to determine whether a robot is suitable for an individual, the robot's effectiveness in terms of usability aspects should be considered. Nevertheless, in many of the included papers, it cannot be detected how a robot's aspects such as bodily movements, head feedback, facial expressions and sound effects influence the outcome of the human-robot interaction. However, a few studies did include this in their paper. Two studies [3, 20] researched and discussed the usability of the robot thoroughly. Another study [26] conducted a long-term effect study which increases the quality of the effectiveness evaluation. Additionally, in this research [26], technical aspects were included and usability aspects are thoroughly evaluated by primary and secondary users. Furthermore, a mix of qualitative and quantitative methods was used for evaluation and an appropriate comparison was made between a humanoid and non-humanoid.

As discussed in the results section, two studies used a combined method of qualitative and quantitative evaluation research. Apart from one study that employed quantitative research, all studies solely applied qualitative research to evaluate the effect of the robots. Considering the purpose of the studies is to study the effect of human-robot interaction on participants, one would expect a more precise evaluation method. Furthermore, the majority reported small sample sizes which could also decrease the accuracy of the effectiveness of interaction evaluation.

4.2 Strengths and Weaknesses of Research

Since there has not been performed a thorough evaluation of existing social robot technologies for people with ID, this literature review is unique. Many researchers in the technical as well as clinical area could benefit from this research. What strengthens this paper as well is that everything that has been concluded and discussed in the included literature, has been represented in a table, thoroughly elaborated and clearly summarized without repetition. Moreover, existing research gaps are identified, so that research on this topic can be improved in the future.

Nevertheless, there has not been made use of a set method to guarantee that all literature on this topic was considered, so therefore there is an increased probability that the review is biased. For example, grey literature was not included in the research. Other evaluations such as peer-reviewed papers could have been published through other channels and therefore, remaining studies that evaluated and tested social robot technologies could have been neglected. However, in the search query (Figure 1), the author included the terms ‘intellectual disability’, ‘learning disability’ and ‘mental retardation’ which are all the terms that describe intellectual disability. Additionally, the reference lists of included literature were cross-checked. Therefore, the probability that relevant literature was missed out on is limited.

4.2.1 Framework

There are several aspects to which the applied framework [14] can be questioned. Firstly, the framework is applied to children with special needs. This literature review included two papers that focused on adults aged 26-31 [3] and elderly aged 59-71 [31]. However, the included literature mainly reported children as participants. The framework [14] focuses on children with special needs in general. This means that it also applies to children with other learning disabilities such as autism or ADHD. This research would have been more accurate if a framework would have been used that solely focused on children with ID. Additionally, the authors have performed a lot of research on social robots applied to children with ASD. Their experience in this subject improves the accuracy of their research. Notwithstanding, their experience is mainly focused on children with autism and not with ID. Thus, it can be questioned if the authors of [14] have sufficient knowledge about intellectual disability which decreases the accuracy of [14] applied to this literature review.

Furthermore, the framework [14] is used for specifically evaluating a robotic toy. The authors state in [14] that a robotic toy regards robots that are being used in play activities of children with special needs, for therapeutic or educational purposes. These robots have shown to be useful in promoting spontaneous play in children. Many of the papers included in this literature review have applied a robot in robot-assistive therapy or in a classroom setting with a purpose to, for example, act as a social mediator [17] or detect stereotyped behavior [25]. It appeared that many of the papers, that had another purpose than implementing the robot as a robotic toy, did not evaluate the social robot well according to the criteria in [14]. Therefore, it can be discussed whether a robot is still considered a robotic toy if its purpose is not to assess playfulness. However, when users interact with a social robot, the robot gives feedback in terms of talking with the users, moving around or performing activities with the users. Moreover, the robot often has assignments for its users in relation to games or play scenarios. Hence, it could be believed that each robot is a robotic toy which improves the accuracy of the application of [14] in this literature review.

5. CONCLUSION

There are not many reviews on social robots that are tested on people with ID published yet. In the available literature, researchers studied the effect of human-robot interaction on people with ID in different application areas and different types of robots. A framework [14] was used to determine the quality of the existing literature. A few papers performed very well, but in the majority of the papers, there is room for improvement in different aspects.

Three sub-questions were formulated how current research performs on the evaluation aspects of the framework [14], ‘technical’, ‘usability’ and ‘effectiveness of evaluation’. Firstly, it appeared that technical aspects were not at all evaluated, except for two studies. Yet, this was not performed thoroughly according to the framework [14]. Secondly, 10 out of the 15 studies covered at least one of the usability evaluation aspects. However, in the majority of the papers, it could not be detected how a robot’s aspects influence the outcome of the human-robot interaction. Lastly, two studies used a mixed method of quantitative and qualitative research to evaluate the effectiveness of interaction. The remaining studies, except for one, merely employed qualitative evaluation methods. A more detailed elaboration on the performance of current research on social robots for people with ID can be found in 3.3.

Considering the above, the main research question: “How can current research on social robots tested on intellectually disabled individuals be improved?”, could be answered. In all included studies, the robot had a positive impact on participants after human-robot interaction and in most instances even improved the participants’ behavior. Nevertheless, the influence of the robot has barely been researched in existing literature. This could sound reasonable, since researchers will probably wish to perceive what the robot’s impact on the participants is, rather than how this robot impacts the participants and which of its aspects play a significant part. Notwithstanding, one will never be able to make a direct step towards the improvement of engagement in future research. Therefore, the author recommends researchers to incorporate technical and usability elements of a robot to measure the robot’s effectiveness in future research.

Furthermore, the author recommends researchers to utilize a mixed method of quantitative and qualitative research to be able to discuss the effect of a robot in a more detailed fashion and to ensure completeness. In addition, larger sample sizes should be chosen to obtain the best possible results.

Moreover, the term mental retardation changed to intellectual disability and is therefore obsolete. Remarkably, it is still used in many papers that were obtained during the search strategy. The assumption is that these engineers do not have the same knowledge as people who are educated within the social science area. Therefore, it is recommended to these researchers to study the social part of the topic more in-depth, rather than solely focusing on the technical aspects.

5.1 Further Work

Given this literature review included a relatively small number of papers (n = 15), the search strategy could be reiterated. Moreover, as discussed in 4.2, grey literature was not included in the research. Other evaluations such as peer-reviewed papers could have been published through other channels and therefore, remaining studies that evaluated and tested social robot technologies could have been neglected. This research provides a unique literature review, so it is recommended to expand this literature review in the future.

As elaborated in 4.2.1, there are some limitations in the used framework [14]. Therefore, the framework should be adjusted when expanding this research. The main variable that should be adapted is the 'technical evaluation'. In the framework [14] there are three aspects of this evaluation which are 'safety', 'adequacy' and 'compliance to national standards' (Appendix A). Considering robots are being tested on the last two aspects when they are being designed, it is very unlikely that these factors will play an important role on the effectiveness of the social robot on people with ID and therefore they should not be included in the current framework [14]. Since safety appeared to be an important factor to participants that tested NAO and Paro [26], it should remain included. Additionally, aspects that are currently under 'usability' could be considered as 'technical' such as 'durability'. Moreover, 'technical configuration' could be included since it was also mentioned in a study [19]. Furthermore, framework [14] could possibly be supplemented by variables from other existing frameworks. For instance, frameworks on the implementation of a social robot in general could be studied to determine if some aspects could be applicable to people who have intellectually disability to improve the current framework [14].

Lastly, this research is an expansion of scientific knowledge about the application of existing social robotic technologies in relation to people with ID. This paper can be used in a practical matter by robotic corporations or engineers in developing future robot technologies for people with intellectual disabilities. Furthermore, this paper acts as a base for further research on the efficiency of social robots within the area of intellectual disability for both technical and social review.

6. ACKNOWLEDGMENTS

I gratefully thank my supervisor, Joanneke van der Nagel, who motivated me and supported me during my research. I also wish to acknowledge the support of my track coordinator, Mariët Theune, and to thank the members of the Intelligent Interaction track who provided feedback on my research.

7. REFERENCES

[1] American Association on Intellectual and Developmental Disabilities. (2013). "Definition of Intellectual Disability." http://aaidd.org/intellectual-disability/definition#.VpWG0DY_WCR. Accessed on June 15, 2020.

[2] Aslam, S., Standen, P. J., Shopland, N., Burton, A., & Brown, D. (2016). A comparison of humanoid and non-humanoid robots in supporting the learning of pupils with severe intellectual disabilities. In *2016 International Conference on Interactive Technologies and Games (ITAG)* (pp. 7-12).

[3] Balasuriya, S. S., Sitbon, L., Brereton, M., & Koplick, S. (2019, December). How can social robots spark collaboration and engagement among people with intellectual disability? In *Proceedings of the 31st Australian Conference on Human-Computer-Interaction* (pp. 209-220).

[4] Bargagna, S., et al. (2019). Educational robotics in down syndrome: A feasibility study. *Technology, knowledge and learning*, *24*(2), 315-323.

[5] Bartneck, C., & Forlizzi, J. (2004, September). A design-centred framework for social human-robot interaction. In *RO-MAN 2004. 13th IEEE international workshop on robot and human interactive communication (IEEE Catalog No. 04TH8759)* (pp. 591-594).

[6] Beccaluva, E. et al. (2017). Exploring engagement with robots among persons with neurodevelopmental disorders. In *2017 26th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)* (pp. 903-909).

[7] Belanche, D., Casaló, L. V., Flavián, C., & Schepers, J. (2020). Service robot implementation: a theoretical framework and research agenda. *The Service Industries Journal*, *40*(3-4), 203-225.

[8] Bernd, T., Gelderblom, G. J., Vanstipelen, S., & De Witte, L. (2010). Short term effect evaluation of IROMEC involved therapy for children with intellectual disabilities. In *International Conference on Social Robotics* (pp. 259-264).

[9] Conchinha, C., Osório, P., & de Freitas, J. C. (2015). Playful learning: Educational robotics applied to students with learning disabilities. In *2015 International symposium on computers in education (SIIE)* (pp. 167-171).

[10] Conti, D., Di Nuovo, A., Trubia, G., Buono, S., & Di Nuovo, S. (2018). Adapting robot-assisted therapy of children with autism and different levels of intellectual disability: a preliminary study. In *Companion of the 2018 ACM/IEEE International Conference on Human-Robot Interaction* (pp. 91-92).

[11] Conti, D., Trubia, G., Buono, S., Di Nuovo, S., & Di Nuovo, A. (2018). Evaluation of a robot-assisted therapy for children with autism and intellectual disability. In *Annual Conference Towards Autonomous Robotic Systems* (pp. 405-415).

[12] Dawe, J., Sutherland, C., Barco, A., & Broadbent, E. (2019). Can social robots help children in healthcare contexts? A scoping review. *BMJ paediatrics open*, *3*(1).

[13] De Groot, J. J., Barakova, E., Lourens, T., van Wingerden, E., & Sterkenburg, P. (2019). Game-based human-robot interaction promotes self-disclosure in people with visual impairments and intellectual disabilities. In *International Work-Conference on the Interplay Between Natural and Artificial Computation* (pp. 262-272).

[14] Ferrari, E., Robins, B., & Dautenhahn, K. 2010. "Does it work?" A framework to evaluate the effectiveness of a robotic toy for children with special needs. In *19th International Symposium in Robot and Human Interactive Communication* (pp. 100-106).

[15] Gelsomini, M., et al. (2017). Puffy-an inflatable mobile interactive companion for children with Neurodevelopmental Disorders. In *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems* (pp. 2599-2606).

[16] Kidd, C. D., & Breazeal, C. (2004). Effect of a robot on user perceptions. In *2004 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)(IEEE Cat. No. 04CH37566)* (Vol. 4, pp. 3559-3564).

[17] Lehmann, H., Iacono, I., Dautenhahn, K., Marti, P., & Robins, B. (2014). Robot companions for children with down syndrome: a case study. *Interaction Studies*, *15*(1), 99-112.

[18] Logan, S. W., et al. (2017). Toy-based technologies for children with disabilities simultaneously supporting self-directed mobility, participation, and function: a tech report. *Frontiers in Robotics and AI*, *4*, 7.

[19] Marti, P., & Giusti, L. (2010). A robot companion for inclusive games: A user-centred design perspective. In *2010 IEEE International Conference on Robotics and Automation* (pp. 4348-4353).

[20] Özdemir, D., & Karaman, S. (2017). Investigating interactions between students with mild mental retardation and humanoid robot in terms of feedback types. *Egitim ve Bilim*, *42*(191).

[21] Rabbitt, S. M., Kazdin, A. E., & Scassellati, B. (2015). Integrating socially assistive robotics into mental healthcare interventions: Applications and recommendations for expanded use. *Clinical psychology review*, *35*, 35-46.

[22] Rouse, M. (2019).” Social robot” SearchEnterpriseAI. https://searchenterpriseai.techtarget.com/definition/social-robot. Accessed on June 21, 2020.

[23] Scoglio, A. A., Reilly, E. D., Gorman, J. A., & Drebing, C. E. (2019). Use of social robots in mental health and well-being research: systematic review. *Journal of medical Internet research*, *21*(7), e13322.

[24] Shamsuddin, S., et al. (2012). Humanoid robot NAO interacting with autistic children of moderately impaired intelligence to augment communication skills. *Procedia Engineering*, *41*, 1533-1538.

[25] Shamsuddin, S., Yussof, H., Mohamed, S., Hanapiah, F. A., & Ismail, L. I. (2013). Stereotyped behavior of autistic children with lower IQ level in HRI with a humanoid robot. In *2013 IEEE Workshop on Advanced Robotics and its Social Impacts* (pp. 175-180).

[26] Silvera-Tawil, D., & Roberts-Yates, C. (2018). Socially-assistive robots to enhance learning for secondary students with intellectual disabilities and autism. In *2018 27th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)* (pp. 838-843).

[27] Silvera-Tawil, D., Bradford, D., & Roberts-Yates, C. (2018). Talk to me: The role of human-robot interaction in improving verbal communication skills in students with autism or intellectual disability. In *2018 27th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)* (pp. 1-6).

[28] Sulkes, S.B. (2020). “Intellectual Disability” https://www.msdmanuals.com/home/children-s-health-issues/learning-and-developmental-disorders/intellectual-disability. Accessed on June 15, 2020.

[29] Syriopoulou-Delli, C. K., & Gkiolnta, E. (2020). Review of assistive technology in the training of children with autism spectrum disorders. *International Journal of Developmental Disabilities*, 1-13.

[30] Trigo, M.J.G., Standen, P.J., Cobb, S.V.G. (2019). Robots in special eduction: reasons for low uptake. Journal of Enabling Technologies. 2398-6263.

[31] Wagemaker, E., Dekkers, T. J., Agelink van Rentergem, J. A., Volkers, K. M., & Huizenga, H. M. (2017). Advances in mental health care: five n= 1 studies on the effects of the robot seal paro in adults with severe intellectual disabilities. *Journal of Mental Health Research in Intellectual Disabilities*, *10*(4), 309-320.

APPENDIX A

Technical evaluation	**Usability evaluation**	**Effectiveness of human-robot interaction**
Adequacy	Position, color and size of robot’s elements: buttons; icons; head; eyes	Quantitative research: questionnaire; survey; adult who is familiar with the person with ID as respondent
Safety	Overall appearance of robot	Qualitative research: interviews; observation
Compliance to international standards	Sound features: audio feedback, general sound of robot’s motors	
	Battery issues: position; weight; time to recharge	
	Quality of the robot’s visual feedback and movements	
	The screen’s visibility	
	GUI analysis	
	Assessment of play scenarios: considering difficulties that may arise during robot’s performance	