RESEARCH METHODOLOGY IN CHILD-ROBOT INTERACTION: A LITERATURE STUDY

WELMOED LOOGE (S0208841)

MASTER THESIS

HUMAN FACTORS & ENGINEERING PSYCHOLOGY

UNIVERSITEIT TWENTE

2016

FIRST SUPERVISOR: DR. MARTIN SCHMETTOW

SECOND SUPERVISOR: CRISTINA ZAGA

17 August 2016

ABSTRACT

Due to its relatively young age, the field of Human-Robot Interactions is still looking for a commonly agreed upon set of research methods. Literature on research methods suitable for use in Child-Robot Interaction (CRI) research is even more scarce, despite the fact that children are not yet as cognitively developed as adults and therefore cannot necessarily be researched with the same methods that adults can. This thesis aims to bridge this knowledge gap regarding suitable research methods for CRI. To that end, a systematic literature review was performed to gain insight in which research methods were used in CRI in the past decade to research children between ages 0 and 12. The research methods found were then analyzed in terms of their suitability for use with children of different ages in light of children's cognitive development. Self-report methods (interviews and surveys) were found to be suitable only for children aged 7 and up, whereas other methods such as observations and physiological measures are suitable for children of all ages. This was found to be because the age of 7 is a turning point from which children are able to communicate their thought to others – an essential skill to successfully take part in self-report research. Therefore, even though individual differences exist between children of the same age, it is recommended to only use self-report measures with children aged 7 or older.

SAMENVATTING

Vanwege haar relatief jonge leeftijd is het onderzoeksveld van Human-Robot Interaction (Mens-Robot Interactie, afgekort: HRI) nog steeds op zoek naar een breed geaccepteerde set onderzoeksmethoden. Literatuur over geschikte onderzoeksmethoden voor gebruik in onderzoek naar Child-Robot Interaction (Kind-Robot Interactie, afgekort: CRI) is nog schaarser, ondanks het feit dat kinderen cognitief nog niet zo ver ontwikkeld zijn als volwassenen, waardoor ze niet perse onderzocht kunnen worden met dezelfde methoden als volwassenen. Deze thesis heeft als doel deze kenniskloof over geschikte onderzoeksmethoden voor CRI-onderzoek te dichten. Daarom werd er een systematisch literatuuronderzoek uitgevoerd om inzicht te krijgen in de onderzoeksmethoden die in het afgelopen decennium zijn gebruikt binnen CRI om kinderen tussen 0 en 12 jaar te onderzoeken. De gevonden onderzoeksmethoden werden vervolgens geanalyseerd aan de hand van de cognitieve ontwikkeling van kinderen om zo hun geschiktheid voor gebruik met kinderen van verschillende leeftijden te bepalen. Zelfrapportage-methoden (interview en vragenlijsten) bleken enkel geschikt te zijn voor kinderen van 7 of ouder, terwijl andere methoden, zoals observaties en fysiologische metingen, geschikt zijn voor kinderen van alle leeftijden. Dit bleek

te zijn omdat 7 het kantelpunt is waarop kinderen in staat zijn hun gedachten te communiceren aan anderen – een essentiële vaardigheid om succesvol deel te kunnen nemen aan zelfrapportage-onderzoek. Daarom is het aan te bevelen, ook al bestaan er individuele verschillen tussen kinderen van dezelfde leeftijd, om zelfrapportage-methoden enkel te gebruiken bij kinderen van 7 jaar en ouder.

TABLE OF CONTENTS

1. INTRODUCTION

In 1921, Czech writer Karel Čapek's famous play R.U.R. (Rossum's Universal Robots) introduced the term 'robot' to the world. Loosely translated from Czech, the term means something like 'forced labor'. Even though this translation corresponds to the stand-alone industrial production machines that robots have mainly been for decades, the robots portrayed in Čapek's play were completely different. These robots were some kind of humanoids that consisted of synthetic organic matter and that could think for themselves. They were so much like humans that they could even be mistaken for them. Although this idea of robots that work with humans and that behave in such a human way may have been revolutionary for Čapek's time, it is not that far-fetched now. Due to rapid progress in robotics, increasingly sophisticated robots are being developed, and robots are now capable to be deployed to work with humans instead of isolated from them.

Current robots vary in the degree to which they interact with humans, depending on their function. For instance, personal service robots are being developed with the aim to help elderly people live at home longer. (e.g. Roy et al. (2000); Scopelliti, Giuliani, & Fornara (2005); Broadbent, Stafford, & MacDonald, (2009); Broekens, Heerink, & Rosendal (2009)). Other robots in human lives do not interact with people as much, such as commercially available robotic vacuum cleaners that are capable of navigating and vacuuming spaces autonomously (Forlizzi & Disalvo, 2006). Despite the fact that these robots are not designed with the aim of interacting socially and forming bonds with humans, Sung, Guo, Grinter, and Christensen (2007) found that people often attributed humanlike characteristics to their robotic vacuum cleaners and that they expressed attachment towards them. This is in line with the assertion of Dautenhahn (2007) that humans have a natural tendency to anthropomorphize everything around them, including technology. This goes especially for children, who, according to Salter, Werry and Michaud (2008) "are unlikely to only use a robot as a tool and they will undoubtedly have some sort of interaction that can be considered social." (p. 94).

Several research projects focus specifically on the use of robots with children. One of them is the Aliz-e project, which aims to develop artificial intelligence for small humanoid robots that are intended to interact with children for longer periods of time. In this project, robots are used to educate hospitalized children with diabetes about their disease and how to manage it (Blanson Henkemans et al., 2013; Neerincx, 2010). The Aurora Project researches the use of robots as educational and therapeutical tools for children with Autism Spectrum

Conditions (ASC) (AURORA, 2015). People with ASC are impaired when it comes to social interaction and social communication. Skills for interpreting and predicting human behavior are typically developed in early childhood, but limited in people with ASC. Because of this, human behavior occurs as very unpredictable to them. In contrast, robots react very predictably and their behavior can be repeated tirelessly (Hanson et al., 2012), which creates possibilities for robots to be used as therapeutic tool for people with ASC. The DREAM project aims to take robot-assisted therapy to a next level by developing robot-enhanced therapy. This robot is intended to work more autonomously than current robots by assessing the behavior of the child and to make inferences about their psychological disposition, based upon which it will be able to decide therapeutic actions tailored to individual children's needs (DREAM, 2015).

Because of children's aforementioned tendency to form social bonds with objects such as robots, robots need to be able to interact with humans socially. According to Bartneck and Forlizzi (2004), "a social robot is an autonomous or semi-autonomous robot that interacts and communicates with humans by following the behavioral norms expected by the people with whom the robot is intended to interact" (p. 592). They describe autonomous robots as robots "having the technological capabilities to act on behalf of humans without direct input from humans." (p. 593), which is necessary because most people interacting with robots – especially children – are not trained to operate robots. Therefore, Burghart and Haeussling (2005) state that the interaction between humans and robots should be as intuitive as possible, which "requires the recognition and consideration of the main social parameters of a co-operative task between human and robot" (p. 23). For that reason, research on interaction between children and robots is necessary. A challenge, however, is that the research field of Human-Robot Interaction (HRI) is still young and that commonly agreed upon research methodologies have not yet been established (Dautenhahn, 2007). This goes even more for interaction between children and robots, or Child-Robot Interaction (CRI). Therefore, this thesis aims to investigate what suitable research methods for CRI are.

1.1 METRICS AND METHODOLOGIES FOR ROBOT INTERACTION RESEARCH

Dautenhahn (2007) describes HRI as "a challenging research field at the intersection of psychology, cognitive science, social sciences, artificial intelligence, computer science, robotics, engineering and human-computer interaction." that aims to "investigate 'natural' means by which a human can interact and communicate with a robot" (p. 103). Even though HRI is related to these fields of research, Dautenhahn argues for separate research methods for

HRI because in HRI, “robots and humans need to coordinate their activities in time and space in real-time, often face-to-face” (p. 103), which makes these interactions different from interactions researched in, for example, Human-Computer Interaction (HCI). Due to differences such as this, Dautenhahn states that research methods from related fields need to be adapted to or developed for use in HRI instead of applying them to HRI without change. Defining common metrics for HRI will, according to Steinfeld et al. (2006), lead to better comparability between studies and greater sharing of knowledge as a result of it. Similarly, Dautenhahn stresses the importance of reproducibility of studies, which will improve when a commonly accepted methodology has been established.

In their attempt to identify common metrics for HRI, Steinfeld et al. (2006) stated that “the primary difficulty in defining common metrics is the incredibly diverse range of human-robot applications.” (p. 33). To classify these various forms of human-robot interactions, Yanco and Drury (2002) proposed a taxonomy that categorizes HRI in terms of the “team composition (ratio of people to robots, types of robots), amount of required interaction, decision support provided for the user, and space-time location”. In 2004, they updated this taxonomy to also include the social nature of the interaction between human and robot, the type of task, and robot morphology (anthropomorphic, zoomorphic, or functional). According to Steinfeld and colleagues, it often happens in the early years of new fields of research that researchers use a wide variety of often application specific measures. “Common metrics develop as researchers devote more attention to the core questions of the field” (p. 33). Steinfeld et al. believe that HRI has reached this point, further stressing the need for commonly accepted methods in the field.

In a survey study of 29 papers that concern metrics for HRI, Murphy and Schreckenghost (2013) classified the proposed metrics they found in these papers as either measures of the human, the robot or the system. Within system metrics, they distinguish between metrics for “productivity, efficiency, reliability, safety, and coactivity” (p. 197). Figure 1 shows a taxonomy of all metrics that were found in this study. Murphy and Schreckenghost conclude from their survey that “often these metrics have no functional, or generalizable mechanism for measuring that feature” (p. 197). They found that instead of being directly measured, the human-robot interaction for the system is often inferred from observations and measurements of the robot or the human, which leads to error and noise in the data analysis.

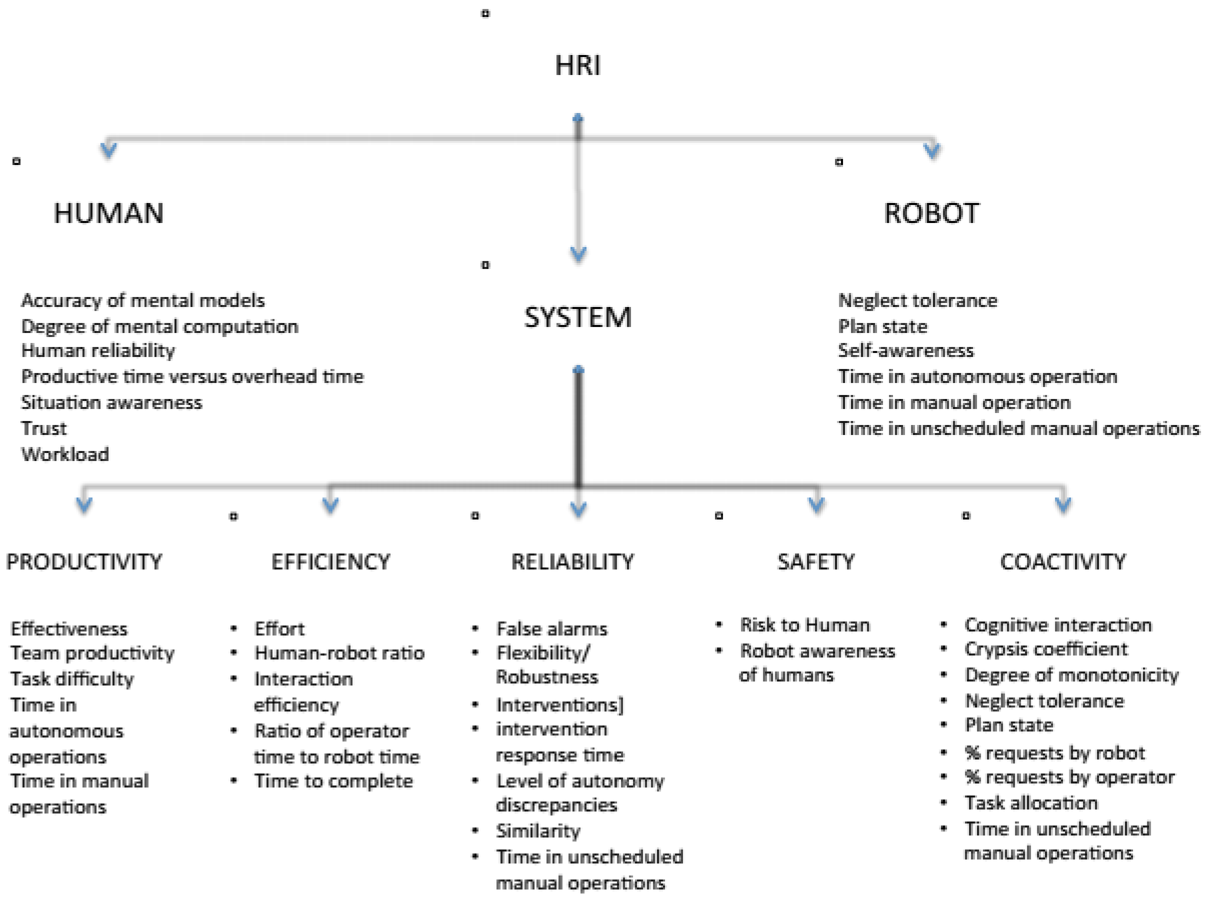

Figure 1. Taxonomy of HRI metrics. Reprinted with permission from "Survey of Metrics for Human-Robot Interaction" by Murphy, R. R., and Schreckenghost, D., 2013, 8th ACM/IEEE International Conference on Human-Robot Interaction (HRI), p. 197. Copyright 20

From all this, it becomes clear that even though the establishment of a commonly accepted set of research methods for HRI is necessary, more research on the topic is needed in order to achieve that. Dautenhahn (2007) warns researchers to avoid methodological battles in their quest to determine suitable research methods for HRI, as she describes that discussions are going on within the field about whether large-scale quantitative research or smaller scale qualitative research such as case studies are best. Dautenhahn argues “that there is no ‘once-and-for-all’ solution applicable across HRI” (p. 103), and that trying to define such solution would damage the development of the young field of HRI. Instead, she poses that it is in HRI’s best interest to recognize that there is more than one way to gain insight into the topics of interest. Despite these efforts to establish a commonly agreed upon research methodology for HRI, literature on research methodologies for CRI is even more scarce.

1.2 RESEARCH WITH CHILDREN

In the previous section, the need for commonly agreed upon research methodologies for HRI was discussed. This section concerns methodologies for child-robot interactions. The central question here is whether these are needed, and if so: why? The same question was asked by Punch (2002), who posed that the way children are perceived by adults is deciding in which methods are used for research. Research with children is often seen as either the same as research with adults, or as completely different. The former view eliminates the need for separate or adapted research methods for children, whereas the latter actually necessitates it. According to Punch, literature that argues research with children to be different from research with adults often offers one of three explanations for this difference: "the position of childhood in adult society, adults' attitudes towards children and the children themselves" (p. 323). Punch suggests that because adults control much of children's lives, children are used to adults having power over them and not used to being taken seriously by adults or expressing their own views. According to Punch, research with children may also differ from research with adults because of adults' attitudes towards children and assumptions about their capabilities. For instance, Punch states that adult researchers often change how they use language in research with children because they assume children to be less articulate. Because children do differ from adults in some aspects, such as their more limited vocabulary and experience in the world, research with children may be different from research with adults as well.

In line with this last assertion, James, Jenks and Prout suggested a third way to see children, namely as being similar to but having different competencies than adults (cited in Punch, 2002, p.322). In addition to that, Punch argues that adults should not assume that all children have the same competencies, such as reading ability or concentration span, and that appropriate research methods should be used. Even though Punch advocates the use of multiple research methods to take these individual differences between children into account, she does not further indicate which research methods are considered appropriate.

A similar view of children to the one proposed by James, Jenks and Prout is the one described by Einarsdóttir (2007), who reported of a study of playschool children's (ages 2-6) experiences with and opinions about the play school they attended. This study "was conducted under the influence of postmodern views of children and childhood, the sociology of childhood, and the children's rights movement" (p. 198). This means "that children, just like adults, hold their own perspectives, have the right to be heard, and are able to speak for themselves if the right methods are used" (p. 199). Einarsdóttir describes that this children's rights movement came into being after the United Nations' 1989 Convention on the Rights of the Child, which states that children have the right to express their own views on matters that concern them ("Convention on the Rights of the Child," 1989). Even though in this quote, Einarsdóttir consideres the use of the right research methods a necessary condition for children to express their opinions, she does not provide an answer as to what the right methods are. Instead, she used multiple research methods in her study (group and individual interviews, analysis of children's drawings and photographs taken by children, and a questionnaire, gathering of artifacts and categorization of pictures). Similar to Punch, Einarsdóttir found that using a variety of research methods is needed to suit the differences between children, and that "different methods can shed light on different aspects and can give a new breadth of understanding" (p. 207).

According to Belpaeme et al. (2013), CRI "is different from interaction between adults and robots in that children have got a different, immature cognitive development" (p. 452), especially with regards to their tendencies to anthropomorphize technology, including robots. This makes the way they interact with robots different to the way adults interact with robots. Furthermore, children are not cognitively as developed as adults. Belpaeme and colleages state, similar to the points of view presented above, that "children are not just small adults" (p. 453), and that children's are not yet linguistically as developed as adults, which poses challenges to the design of robots that are to interact with children. However, if children's more limited linguistic development poses challenges to robot design, it also poses challenges to the way they can be researched.

To summarize, even though children are similar to adults in the way that they deserve to be heard, they cannot be researched in the exact same way that adults are researched because children are not yet as cognitively and linguistically developed as adults. This calls for research on suitable research methods for CRI.

1.3 CHILDREN'S COGNITIVE DEVELOPMENT

Because children do differ from adults with regard to their cognitive development, this thesis ascribes to the views that children may be similar to adults but have different capabilities. Children, however, form a heterogeneous group with much developmental differences between individual children. Therefore, in the quest to identify suitable research methods for CRI, it is important to consider children's capabilities at different developmental stages. Borgers, De Leeuw, and Hox (2000) investigated the effects of children's cognitive development on the response quality of surveys. In this research, they classified children according to Piaget's stages of cognitive development. Piaget (1960) recognized four stages of cognitive development: the sensorimotor stage, the preoperational stage, the concrete operational stage and the formal operational stage. The sensorimotor stage lasts from birth to around 18 months of age. During this time, children develop sensorimotor intelligence, which arises from the child's sensory and motor experience with the world and the assimilation of these new experiences into the child's cognitive schemata. The second stage of development, the preoperational stage, lasts from about 18 months through seven or eight years of age. Piaget described two distinct developments of thought within this stage: the development of preconceptual and symbolic thought (until about 4 years of age) and that of intuitive thought (between 4 and 7-8 years). During the development of preconceptual intelligence, children learn to use language and to attach notions, or pre-concepts, to this language. In this stage of development, children are beginning to build cognitive schemata of the world around them, which Piaget described as remaining "midway between the generality of the concept and the individuality of the elements composing it, without arriving either at the one or at the other. (…) It is clear that such a schema, remaining midway between the individual and the general, is not yet a logical concept and is still partly something of a pattern of action and of sensori-motor assimilation" (p. 127-128). Summarizing, this means that children are starting to form the schemata that are at the basis of conceptual and formal thought during this stage. After the development of preconceptual thought, children develop intuitive thought. During the development of intuitive thought, children conceptualize more than in previous stages of development. They move on from "simple half-individual, half-generic figures" (p. 130) to more complex representational structures, forming a rudimentary logic, "but in the form of representative regulations and not yet of operations". Concrete operations are, according to Piaget, developed between ages 7-8 and 11-12. Scott (1997) described that children's thought changes dramatically between the preoperational and concrete operational stage. During the concrete operational stage, "children are not only able to take on the view of others, but they are also capable of logical thinking and deductive reasoning, even if their thought processes are still tied to the concrete operations of their immediate world" (p. 334-335). These changes in children's thinking make that according to Scott (1997) and De Leeuw (2011), children are able to be answer structured constructed questionnaires or interviews from around age seven, depending on their individual development. From around age 11, formal thought starts to develop. This way of thinking is characterized by thinking beyond the present and forming theories about everything, whereas children in previous stages only concern themselves "with action in progress" (Piaget, 1960, p. 148) without forming theories. Children in the formal operational stage are capable of reflective thought and hypothetico-deductive reasoning.

Borgers, de Leeuw and Hox chose to classify children's development according to Piaget's research because it provides "a global classification of developmental stages of children" (p.62). However, they introduce their use of this classification with a few footnotes, because Piaget's theory has received critique in later research. The boundaries between the different developmental stages have been argued to be less distinct than Piaget describes. Instead, there may be overlap between the stages, and one has to take into account that even within these stages there will be considerable differences in cognitive development between individual children (Borgers, De Leeuw, & Hox, 2000). In addition to that, later research has shown that very young children have more reasoning ability than Piaget thought. Machado and Lourenço (1995) encountered the criticism that Piaget underestimated preoperational children's capabilities as one of the most frequent critiques on Piaget's theory, but they argue that these criticisms are based on misinterpretations of Piaget. Regardless of whether or not this is true, Scott (1997) argues that even though younger (preoperational) children may be more capable of reasoning than Piaget thought, their ability to successfully participate in survey research is still limited by their abilities to comprehend language and their verbal memory. Scott furthermore argues that "children's cognitive capacity clearly does increase with age and the rudimentary levels of cognitive development remain relevant for understanding the question and answer process and for highlighting the ways in which children may differ from adult respondents" (p. 334). Therefore, she, as well as Borgers and colleagues, chooses to use Piaget's stages of development classify children according to their cognitive development. In line with these arguments, the research methods found in this thesis will be analyzed in light of children's cognitive development as described in Piaget's theory of cognitive development. Other developmental psychologists, such as Vygotsky, have argued that social development is of more influence on children's cognitive development than Piaget gives credit for, Piaget's classification is used here because contrary to the work of Vygotsky, it provides us with the global classification of children's development at different ages that is necessary to come to global recommendations of suitable research methods for CRI.

1.4 CLASSIFYING RESEARCH METHODS

In order to identify which research methods are suitable for use in the field of child-robot interaction, a classification of research methods is needed. Dooley (2001) categorized measures according to two 2-level dimensions: verbal vs. nonverbal and obtrusive vs. unobtrusive measurement. According to Dooley, "verbal measures apply to written or spoken messages, as in questionnaires. Nonverbal measures apply to physical signs including visual judgments of nonverbal behaviors (e.g. facial expressions) and physiological measures (e.g. blood pressure)" (p. 97). The obtrusiveness of measures is the extent to which participants are aware that they are being measured. During questionnaires and physiological measures, participants are aware of this fact, which may cause them to change their behavior because they know they are being observed. Dooley refers to this change in behavior as reactivity, which is a process that can threaten the validity of the research. Unobtrusive measures can be carried out without the participants knowing, for example by post-hoc analysis of audio or video recordings for certain behaviors or utterances.

Because research methods differ in the extent to which they are verbal and obtrusive, research methods rely on different participant abilities. For example, to analyze verbal utterances, participants need to be capable of expressing themselves verbally, and to conduct a questionnaire, participants need to be able to understand the questions, form their answers, and fill in their answer on the questionnaires. Behavior observation, on the other hand, does not require these capabilities. Therefore, it is important to discuss research methods in light of these dimensions to determine upon which capabilities these methods rely and whether this makes them suitable for use in CRI.

However, classifying research methods only on their obtrusiveness is too broad. Therefore, a more extensive classification based on Lazar, Feng and Hochheiser (2010) is used. In their book on research methods in Human-Computer Interaction, they dedicate chapters to the following research methods: surveys, diaries, case studies, interviews and focus groups, ethnography, usability testing, and measuring the human (eye tracking and physiological measures). It should be noted that interviews are sometimes considered surveys as well, as interviews often are conducted based on a questionnaire. Lazar and colleagues distinguish the two in that they describe surveys as tools that can be self-administered by participants, even when there is no researcher present, whereas interviews have to take place face-to-face. By this definition, surveys can be distributed amongst a large amount of (potential) respondents, whereas interviews are generally carried out with a smaller number of respondents due to the time-intensive nature of this method. When looking at classifications of research methods for psychology, however, Lazar and colleagues' list does not seem to be extensive. In their book on social psychology, Kassin, Fein, and Markus (2008) distinguish between three types of research methods: self-report methods (such as questionnaires and interviews), observations, and technological measurements (such as EEG and physiological measures). Even though the research methods that Lazar and colleagues describe almost all fit one of these three categories, Lazar et al. do not mention observation as a research method. However, because behavior observation is a commonly used research method in psychology and psychology does play a role in CRI, we add behavior observation to the list of research methods that Lazar and colleagues describe.

1.5 RESEARCH QUESTIONS

To summarize, even though there have been some efforts to identify a commonly agreed upon set of research methods for the field of human-robot interaction, such methodology has not been established yet. For child-robot interaction, no consensus about suitable research methods has been reached yet either. Researchers who view children as similar to adults but having different capabilities advocate the use of appropriate methods in research with children. However, the question of what appropriate research methods are remains open. Despite this knowledge gap in suitable research methods for CRI, CRI research is being conducted and theories are being built based upon this research, which in turn form the basis for the development of robots that are to interact with children. This once again stresses the importance of investigating what research methods are suitable for CRI, because these methods form the foundation for the development of robots that are able to interact with children and that are able to do that well. Therefore, the goal of this thesis is to identify which research methods are appropriate for use with children. In line with Dautenhahn's (2007) advice to avoid to dogmatically determine one and only one best research method, this study does not aim to determine which research method is superior, but rather to research which ones are suitable for use in the field of child-robot interaction, given children's cognitive capacities. To that end, a literature study will be conducted to identify which research methods are currently being used in CRI. Then, these research methods will be analyzed in terms of what cognitive capacities they require from participants, and to determine in light of theories on children's cognitive development which are suitable for use in CRI. Finally, this thesis will explore research fields in which commonly agreed upon methodologies for research with children have already been established to determine if methods from these fields can also be applied to CRI.

2. METHODS

To review which methods are currently used in child-robot interaction research, systematic literature review was performed. The method for this systematic review is based on the one used by Riek (2012). Riek perfomed a search with a specific search term and screened the found papers against her inclusion criteria. Then, she analyzed the included papers using certain criteria for analysis in order to gain insight in how the Wizard of Oz-method is used in the field of HRI. In line with this method, this section will describe the search term, inclusion criteria and criteria for analysis that were used in this literature review to gain insight into which research methods are used in CRI-research.

2.1 SEARCH TERM

In November 2015, a search was performed on the Web of Science (http://www.webofknowledge.com) with the following search term:

Child robot interaction AND (research OR experiment OR study OR pilot)

The Web of Science was chosen as search engine for this search because it indexes nearly all journals and conference proceedings listed in rankings of journals on Human Factors and Ergonomics[1] and Human Computer Interaction[2] – fields of research that are most likely to publish articles on CRI. The search on Web of Science yielded a total of 331 results, which will be screened and only included in this study if they fulfill the following criteria. From the publications included in this study, currently used research methods in CRI will be identified and reflected upon. These methods will then be discussed in light of theories on children's social and cognitive development to establish which methods are suitable for children of different ages.

2.2 INCLUSION CRITERIA

In line with the inclusion criteria described by Riek (2012), only papers published in peer-reviewed journals or conferences will be included. Furthermore, only papers published in the last ten years – so between 2005 and 2015 – will be included because older papers may be

---

1 Http://www.scimagojr.com/journalrank.php?area=0&category=3307&country=all&year=2014&order=sjr&min=0&min_type=cd

2 http://www.scimagojr.com/journalrank.php?category=1709&area=0&year=2014&country=&order=sjr&min=0&min_type=cd

outdated due to the fast rate of development in the field of robotics. Only papers published in English will be included.

AGE OF PARTICIPANTS

Because this thesis concerns itself with children's capabilities to participate in research in relation to their cognitive development, the focus of these inclusion criteria is children's cognitive age. If participants' cognitive developmental age is not specifically mentioned, it is assumed that their developmental age matches their chronological age. In this case, developmental age is defined as follows: "age of a person estimated from the degree of anatomic, physiologic, mental, and emotional maturation"[3].

In this literature study, only papers that focus on children up to the age at which they leave primary school will be included. In most countries with a school system of primary and secondary education, the age at which children leave the last grade of primary school is around 12 years. Therefore, only papers focusing on children between ages 0 and 12 will be included in this study. Papers in which the ages of participants range beyond 12 years of biological age will be included only if the developmental age of the participants is younger than 12, or if the papers distinguish different age groups within their sample of participants and present their results for each age group. In that case, only those age groups that contain children up to 12 years of age will be reviewed in this study. For example, if a paper describes a study with participants of ages 9-14, this paper will only be included if it distinguishes different age groups of participants, e.g. 9-11 and 12-14 years. In this case, only the parts of the study concerning the age group of 9-11 years will be included in this study. Papers with participants both younger and older than 12 years of age that do not make this distinction and that only report aggregate findings across all participants will not be included in this study.

In case no developmental or biological age of the participants is reported in the paper, the paper will be screened for indications of participant's age, such as the school grade the participants are in. For instance, papers reporting of pre-school children or 4th graders will be included, because the ages of these participants are within the age limit of this study. Papers using descriptions such as 'students', 'undergraduates', or 'health care specialists' will not be

[3] developmental age. (n.d.) FARLEX PARTNER MEDICAL DICTIONARY. (2012). Retrieved August 17 2016 from http://medical-dictionary.thefreedictionary.com/developmental+age

included because these descriptions imply ages higher than twelve. If the ages of participants cannot be inferred from a paper, this paper will not be included in this study.

2.3 ANALYSIS

The first stage in the analysis of the papers that resulted from the above described search action on the Web of Knowledge is skimming the papers to determine whether these papers fit our inclusion criteria. The papers that do will then be read more thoroughly to extract the following information for analysis from them: the number of participants in the study; participants' ages; and the research methods that were used. To gain insight into the amount of research that has been done on research methods for CRI, we are furthermore interested in whether the research reported in papers is method-oriented, theory-oriented, or design-oriented. Method-oriented papers are papers that focus on the method of doing research within the field of CRI. This category is of special interest to this thesis. Theory-oriented papers focus on building and testing theories about child-robot interaction, for instance regarding what robot behavior is preferred by children or whether interacting with robots can change autistic children's stereotypical autistic behaviors. Lastly, design-oriented papers are characterized by their focus on the design of (parts of) a robot or other system that is to interact with children.

For each paper, the range of participant ages was recorded to analyze for which ages each found method was used. When the specific ages of participants were mentioned in a paper, these ages were used for analysis. When a paper specified an age range, for instance when papers reported that their participants were between 3 and 5 years old, it was assumed that the method(s) used in that paper was used for children aged 3,4, and 5 – unless stated otherwise in the paper. Another example is the study by Feil-Seifer and Matarić (2009), in which four children between the ages of 20 months and 12 years participated. For this paper, the age range of 1-12 years was recorded, even though not all ages within this range were represented. However, because Feil-Seifer and Matarić did not specify otherwise, it was assumed that the same methods that applied to the 20-month-old would also be used with, for instance, 7-year-old or 12-year-old participants. Eight papers were not included in the analysis of participant ages because they only reported their participants' mean age and it's standard deviation. From this information, it could not be inferred what age range the participants had.

3. RESULTS

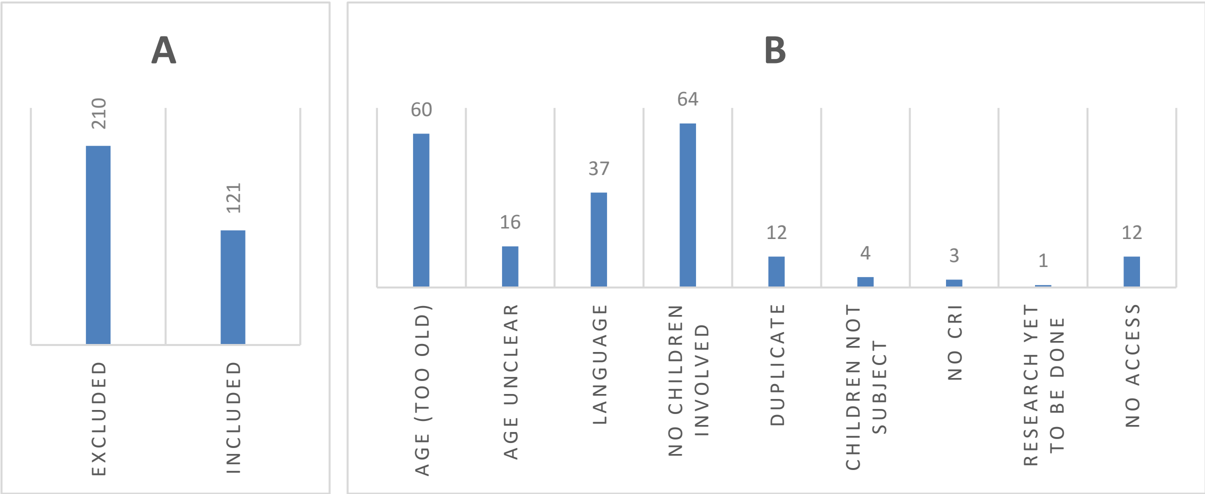

Figure 2. (A) Number of papers included and excluded. (B) Reasons for exclusion.

From the 331 papers that resulted from the search action on the Web of Knowledge, 121 papers fit the inclusion criteria. Figure 2B shows the reasons for exclusion of papers. Of the 210 excluded papers, 60 were excluded because participants were older than our age limit of 12 years, or because the study included older participants but presented the results aggregated over all participants instead of separately. Sixteen papers were excluded because they did not indicate participants' ages and the ages could not be inferred from the papers. 37 papers were excluded because they were not in English (34 were in Korean, two in Turkish and one in French). 64 papers were excluded because they did not describe research that involved children – or any human participants at all, for instance, papers in which robot functionality was tested without human participants. Five papers contained research in which not the children or their behavior was the subject of interest, but the behavior of adults interacting with children. Three papers did not contain research with a robot or computer, which meant that no interaction between children and a robot or computer took place in these experiments ('no interaction' in Figure 2), and one was a research proposal of research that had yet to be done at the time the article was published. These papers were excluded as well. 12 papers were excluded because they were duplicates of other papers. Several of these were papers that were found twice in the search action, and several others were different papers that described the same research. In those cases, the paper published first was included and the duplicates were not. In one case, two different papers describing the same research were presented at the same conference. Out of the two, the paper that described this research in the most detail was included, and the other one was not. Lastly, twelve papers were excluded because the University of Twente did not have access to them and we received no reaction to a request for a copy of the paper from the corresponding authors of these papers. As a result, these papers could not be screened against the inclusion criteria and therefore, they were excluded by default.

The 121 papers that were included in this study were analyzed for the research methods they reported using. As described earlier, the methods found can be categorized into the following categories: surveys, observation, diaries, case studies, interviews and focus groups, ethnology, usability testing and human measures. Figure 3 shows the frequencies with which these methods were found in the literature. The results will be presented for each of these methods, after which a discussion of the method's suitability for use in CRI will follow.

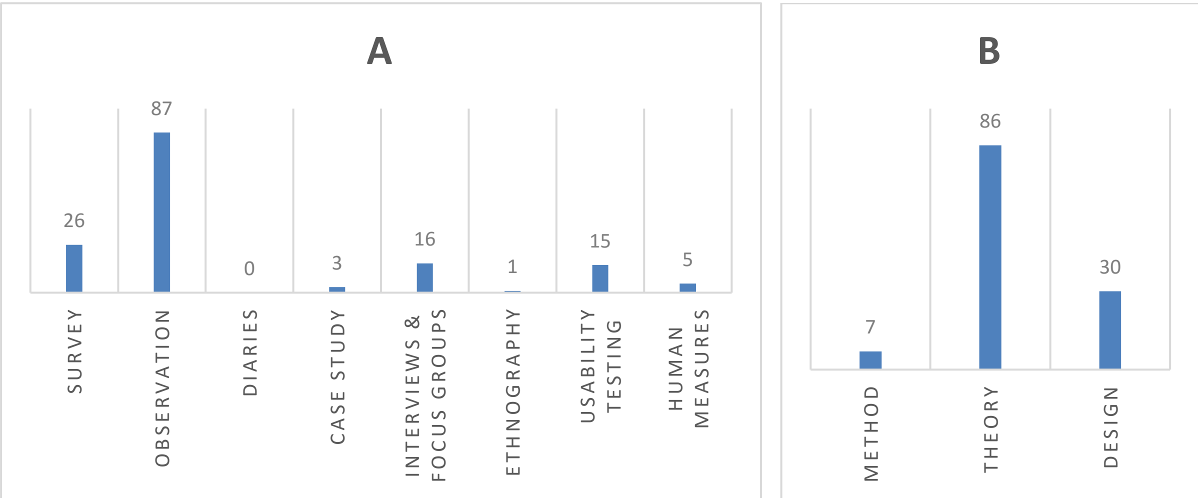

Figure 3.(A) Research methods found in the CRI literature study and their frequencies. (B) Focus of included papers.

3.1 FOCUS

Considering the young age of CRI as a research field and the aforementioned need for research on suitable research methods for this field, it is not surprising that this literature study only found seven papers that focused on developing or validating research methods for CRI (see Figure 3B). At the same time, this actually is surprising because almost 75% of the papers in this study focused on building or testing theories within the field of CRI, despite the knowledge gap in suitable methods for conducting CRI research.

The papers that did focus on the method of doing research in CRI explored different research methods. Leite, Henriques, Martinho and Paiva (2013) researched measuring electrodermal activity (EDA) as a possible method of evaluating child-robot interactions and of recognizing children's affective states. Dickerson, Robins, and Dautenhahn (2013) took a conversation analytic perspective on the interaction of one autistic child with a humanoid robot. With this, they demonstrated the importance of treating every interaction as potentially relevant to not miss important information that might not be noticed when looking for a prespecified set of behaviors. Veenstra and Evers (2011) and Gomes, Sardinha, Segura, Cramer, and Paiva (2014) researched survey methods for CRI. Veenstra and Evers developed and pilot-tested the KidSAR (Children's Social Attitude towards Robots) instrument, and Gomes et al. developed a questionnaire as part of their effort to establish a methodology to evaluate interactions between children and robots that can migrate to virtual entities.

In the next sections, the research methods that were found to be used in literature will be presented and then discussed in terms of their fit with children's cognitive capacities.

3.2 SURVEYS

Surveys were found to be the second-most used method in the papers that qualified for inclusion in this study: 26 papers reporting using questionnaires with children. Four domains of questions became apparent during analysis of the papers: robot-related questions, interaction-related questions, task-related questions, and child-related questions. Table 1 shows an overview of the four domains and examples of themes within these domains that were found to be researched using surveys. The four domains will be discusses in more detail below.

Table 1. Domains of questions researched using surveys and examples of themes within these domains.

	Number of papers	Examples
Robot-related questions	11	
Perceived robot attributes	5	*Robot's perceived intelligence, trustworthiness, and social presence*
Recognition of robot's states	3	*Recognition of robot's emotions*
Acceptance of robot	2	
Perceived robot behavior	2	*Robot's (dis)obedience or performance*
Interaction-related questions	18	
Desire to interact with robot	1	*Desire to communicate and interact physically and emotionally with robot*
Relationship with robot	7	*Intimacy, bonding, social attraction towards robot, robot's role in relationship*
Fun	7	*Fun, satisfaction*
Difficulty of interaction	1	
Engagement	8	*Sensory immersion in play with robot, interest in task, motivation for task, concentration on task*

Task-related questions	3	*Perceived task duration, cognitive demand of task*
Child-related questions	8	*Negative attitude towards robot, experience with robots, experienced pain*

INTERACTION-RELATED QUESTIONS

As can be seen in Table 1, the domain that was researched most often using surveys was the interaction between children and robots. 18 papers reported using surveys that contained interaction-related questions. The most often researched topic with this domain was children's engagement in the interaction with a robot. Whereas some papers did not specify what questionnaire items they used to survey children's engagement (Kose-Bagci, Ferrari, Dautenhahn, Syrdal, & Nehaniv, 2009; Leite, Castellano, Pereira, Martinho, & Paiva, 2014; Leite et al., 2013), Mubin et al. (2010) reported using the Game Experience Questionnaire (GEQ) to measure engagement. Other measures for engagement were children's reported concentration and motivation on the task (Hashimoto, Kobayashi, & Kato, 2011), sensory immersion in the activity with the robot (J. Han, Jo, Hyun, & So, 2015) and interest in the task (Han, Jo, Park, & Kim, 2005; Han, Jo, Jones, & Jo, 2008; Hashimoto et al., 2011).

Children's relationship with the robot was also frequently researched with surveys. Children were asked about the robot's role in the relationship (Oh & Kim, 2010), social attraction towards the robot (Kose-Bagci et al., 2009; Tung, 2011) and their bonding with the robot (Ros, Baroni, & Demiris, 2014; Veenstra & Evers, 2011), as well as about the intimacy children felt with the robot (N. Kim, Han, & Ju, 2014).

Another frequently researched topic was children's experience of fun during the interaction with a robot. Gomes, Sardinha, Segura, Cramer, and Paiva (2014) reported using a Likert-scale to rate children's fun on. Similarly, three papers used a Likert-scale with smiley faces (sometimes referred to as a 'Smileyomter'), intended as visual aids to facilitate children in filling in the questionnaires (Blanson Henkemans et al., 2013; Leite et al., 2014; Shahid, Krahmer, & Swerts, 2014).

ROBOT-RELATED QUESTIONS

Children's perception of robot attributes was the most-researched topic within the domain of robot-related questions. Specifically, the robot's perceived intelligence (Kose-Bagci et al., 2009; Veenstra & Evers, 2011) and social presence and support (Leite et al., 2014, 2013) were researched, in addition to the robot's perceived trustworthiness and care (Veenstra & Evers, 2011), empathy (Han, Jo, Hyun, & So, 2015) and appearance (Kose-Bagci et al., 2009).

Other robot-related questions included children's acceptance of the robot (Hwang & Wu, 2014; Veenstra & Evers, 2011), the robot's likeability (Mubin et al., 2010) and children's regonition of emotions displayed by the robot (Cohen, Looije, & Neerincx, 2014; Goris, Saldien, Vanderniepen, & Lefeber, 2009).

TASK-RELATED QUESTIONS

Task-related questions included the task's cognitive demand (Mubin et al., 2010) and children's perception of task duration (Wood, Dautenhahn, Lehmann, et al., 2013; Wood, Dautenhahn, Rainer, et al., 2013).

CHILD-RELATED QUESTIONS

Eight papers reported asking children child-related questions on surveys. These questions pertained to children's experience with the robot (Mubin et al., 2010; Ros et al., 2014), their negative attitudes towards a robot (Dinet & Vivian (2014), and the amount of pain they experienced during a vaccination they received while being distracted by a robot (Beran, Ramirez-Serrano, Kuzyk, Fior, & Nugent, 2011).

SURVEYS: DISCUSSION

One of the benefits of using surveys is that they allow researchers to obtain a large amount of data relatively easily. Surveys are a form of self-report, and can therefore be completed by participants without a researcher present. In addition to that, the data collected with surveys is easy to analyze. According to Lazar, Feng, and Hochheiser (2010), surveys are especially useful for getting the 'big picture'. However, they state that "an interview question might yield an extensive answer to a question that would generate only a few words in a survey response" (p. 189). Because surveys do not allow researchers the opportunity to ask follow-up questions based on answers received, the depth of information that can be gathered with surveys is limited.

Apart from the limited depth of information that can be gathered, surveying children faces other issues as well, because surveys are not suitable for use with children of all ages. As mentioned earlier, Scott (1997) and De Leeuw (2011) posed that surveys can be used only with children older than seven years of age. In fact, De Leeuw stated that "below the age of 7 children do not have sufficient cognitive skills to be effectively and systematically questioned" and that "below the age of 7, direct questionnaire research of children is not feasible at all" (p. 6). According to Breakwell (as cited in Read and Fine, 2005), "there are four stages in a question-answer process:

1. Understanding and interpreting the question being asked
2. Retrieving the relevant information from memory
3. Integrating this information into a summarized judgement
4. Reporting this judgement by translating it to the format of the represented response scale."

According to Read and Fine, "factors that impact on question answering include developmental effects; language, reading age, and motor abilities, as well as temperamental effects including confidence, self-belief and desire to please". Even though language and reading abilities of children aged 7 to 12 are still developing, De Leeuw (2011) poses that they are sufficiently developed in seven-year-olds to be surveyed, provided that questions consist of short sentences with carefully checked wordings. For instance, children within this age group are not yet able to understand negations and logical operators such as 'or', and in addition to that, special attention should be payed to avoid ambiguous language in the questions as well as answer options because children are even less capable of dealing with ambiguous language than adults (Borgers, De Leeuw, & Hox, 2000; De Leeuw, 2011). And even if these factors are taken into account, surveying young children is challenging because young children are susceptible to suggestibility, which entails that they are easily influenced by questions or researchers conducting the survey. In addition to that, children may succumb to satisficing, which Read and Fine describe, "occurs when a respondent gives more or less superficial responses that generally appear reasonable or acceptable, but without having gone through all the steps in the question-answer process". Factors that influence the level of satisficing are the respondent's motivation, the difficulty of the task and the respondent's cognitive abilities. Considering the aforementioned precautions regarding the construction of surveys that need to be taken in order to succesfully survey children aged 7 and up, the task of answering survey questions is still quite difficult for this age group in relation to their cognitive capacities. This, combined with children's tendency to want to please (Borgers, De Leeuw, & Hox, 2000), explains children's tendency to answer 'yes' regardless of the question (Breakwell, cited in Read & Fine, 2005).

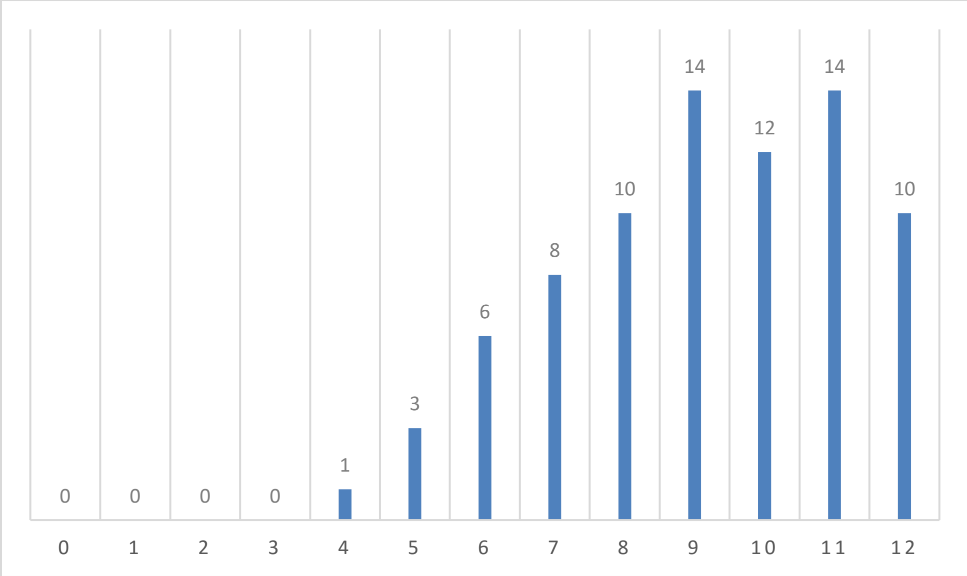

Figure 4. Histogram of participants' ages in studies that reported using surveys.

Despite these difficulties with surveying children older than 7 years of age, as can be seen in Figure 4, there were ten studies that reported using questionnaires with children younger than 7. Beran, Ramirez-Serrano, Vanderkooi, and Kuhn (2013) even reported surveying children as young as four years old. They administered the Faces Pain Scale-Revisited (FPS-R) to children between ages 4 and 9 to measure their pain while they received a vaccination. The FPS-R is a visual analogue scale (VAS) with six faces that express increasing levels of pain. The FPS-R was derived from the Faces Pain Scale (FPS by Bieri, Reeve, Champion, Addicoat, & Ziegler, 1990) by Hicks, Von Baeyer, Spafford, Van Korlaar, & Goodenough (2001). Whereas the FPS consists of seven faces expressing pain, the FPS-R consists of six, allowing for easy scaling of the scores on a 0-5 or 0-10 metric, which are commonly used metrics for self-reports of pain. Even though according to Scott and De Leeuw, children cannot be surveyed below the age of 7, Hicks and colleagues found high correlations between children’s pain ratings on the FPS-R and their ratings on other visual analogue scales even for children as young as 5 or 6 years, which supports the FPS-R’s validity. However, Beran and colleagues studied children as young as 4 years old, and evidence that the FPS-R can be used reliably with children aged 5-6 does not necessarily mean that this holds true for younger children as well.

The evidence towards the reliability of visual analogue scales with children as young as age 5 contradicts findings that suggest that surveys can only be reliably used with children older than 7 years. Therefore, Shields, Palermo, Powers, Grewe, and Smith (2003) researched cognitive and demographic predictors of children’s ability to use a visual analogue scale. They found the best predictor of this ability to be children's age combined with their estimated IQ. These predictors were found to predict with 88% accuracy whether children were able to use a VAS successfully. Shields and colleagues found that "(…) there appears to be a trade-off between age and estimated IQ. The younger the child, the higher the estimated IQ that is needed to complete the VAS successfully. Conversely, the older the child, the less stringent is the estimated IQ requirement for successful completion of the VAS" (p. 287). Figure 5 depicts the probability of a child's successful use of a VAS based on age and IQ. However, Shields et al. noted that measuring children's IQ is a time-consuming task that may only be worth the effort in cases where children need to fill out the VAS regularly (for instance in cases of at-home pain monitoring) instead of only once or twice. On top of that, Shields an colleagues found that only 42% of the children in their study was able to succesfully use the VAS, leading to the conclusion that VAS is not a suitable research method for kindergarten children, and that alternative rating scales are needed for children younger then 7.

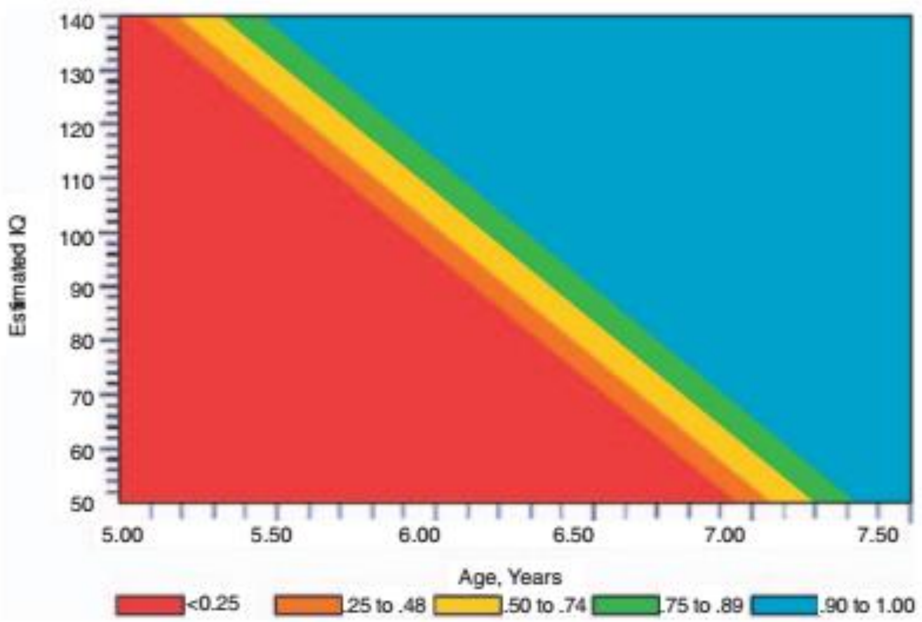

Figure 5. Probability of a child's successful use of a visual analogue scale. Source: "Predictors of a child's ability to use a visual analogue scale" by Shields, Palermo, Powers, Grewe, & Smith (2003).

Similar results were found for the Smileyometer – a Likert-type scale with different smiley faces depicting the values on the scale, developed by Read, Macfarlane, and Casey (2002) to measure children's fun. Han, Jo, Hyun, and So (2015) used a Smileyometer-like scale in their study to survey children of 5 and 6 years old. Even though the Smileyometer was originally intended for use with children between 5 and 10 years of age, Read and MacFarlane

(2006) later found that children aged 7-9 showed little variability in their answers, and although they did not test this, they suspected that that there would be even less variability in even younger children's scores. These findings limit the validity of the Smileyometer for young children.

To summarize, even though contrasting findings have been reported regarding the age at which young children are able to successfully complete visual analogue scales, Shields, Palermo, Powers, Grewe, and Smith (2003) found that children aged 7 and up had a higher probability to successfully do so than children younger than 7. Due to individual differences in IQ scores, some children below the age of 7 have been found to be able to use VAS. However, from this analysis it become apparent that the best course of action in cases where estimating individual children's IQs is too time-consuming is to survey only children aged 7 and up.

As described earlier, when surveying children aged 7 and up, special attention has to be paid to the design of the survey in order to accommodate for children's still limited language an thought abilities. In order to check whether the intended respondents are able to understand the wording of the questions and whether the instructions for the survey are clear, pretesting surveys before use is necessary (Scott, 1997; De Leeuw, 2011; Lazar, Feng, & Hochheiser, 2010; Collins, 2003). Lazar, Feng and Hochheiser cite Dillman (2000), who proposed a three-step method for pretesting surveys. The first stage in pretesting a survey is to have it reviewed by knowledgeable colleagues. Once the survey has passed that stage, potential respondents should be interviewed to examine what they think of it. Lastly, a "pilot study of both survey tool and implementation procedures" should be carried out (Lazar, Feng, & Hochheiser, 2010, p. 118). However, Lazar and colleagues also cite Dillman in noting that this process is rarely done thouroughly. The findings of this literature study are in accordance with that notion, because not one study was found reporting a thorough pretest of their survey tool. However, one study was found to make an effort to this extent. Tung (2011) reported that "the wording used in the questionnaires was discussed with teachers and the children to prevent any misunderstanding" (p. 640).

Another noteworthy observation that was made during analysis of the papers reporting the use of surveys was that hardly any of them report measures validity or reliability of the survey tool used. This is understandable in cases where previously validated surveys were used, such as in Beran, Ramirez-Serrano, Vanderkooi, and Kuhn (2013). They used the FPS-R, which was validated by Hicks et al. (2001) and is widely used in the medical field to measure pain. However, several studies reported constructing their own surveys or adapting existing ones, often without reporting measures of the survey's validity or reliability. For example, Leite, Castellano, Pereira, Martinho, and Paiva (2014) reported presenting their respondents with a survey that consisted of parts of other surveys. They used parts of the McGill Friendship Questionnaire that was developed by Mendelson and Aboud (1999). Mendelson and Aboud found these questionnaire subscales to have high internal consistency and validity, so using only these subscales of the entire Friendship Questionnaire can be done reliably. However, Leite et al. also reported using a questionnaire to measure the robot's perceived social presence which they had used in a previous study as well, but neither of these papers reported measures of validity or reliability of this questionnaire. In addition to this questionnaire, Leite and colleagues used a questionnaire to measure children's engagement in interacting with the robot. They report that "the questionnaire items we used for Engagement are based on the questions developed by Sidner et al. to evaluate users' responses towards a robot capable of using several social capabilities to attract the attention of users" (p. 333). Sidner, Kidd, Lee, and Lesh (2003) in turn report that they measured engagement by adapting questions by Lombard et al. (2000) and Lombard and Ditton (1997), whose papers did not report measures of internal consistency, validity or reliability as the questionnaire they described was still in development at the time of writing. Furthermore, despite reporting that they adapted questions from Lombard and colleages, Sidner, Kidd, Lee and Lesh do not report any of these measures for their adaptations either, and yet Leite and colleagues base their questions regarding engagement on these adaptations. This raises the question whether this questionnaire is valid and reliable.

3.3 INTERVIEWS & FOCUS GROUPS

"The ability to 'go deep' is perhaps the strongest argument in favor of interviewing," according to Lazar, Feng, and Hochheiser (2010). With this, they mean that interviews have the potential of gathering more in-depth information than surveys because they pose that it is easier to answer interview questions than questions on a questionnaire. Depending on the amount of structure in the interview, interviews allow researchers to ask follow-up questions where they want more information, or to even abandon all structure and let the conversation flow as it may. At the same time, a drawback to interviewing is that it is more time-intensive to sit down face-to-face with each participant than it is to hand out questionnaires for participants to fill in. In addition to that, processing the data gathered from interviews is a far more time-consuming task than analyzing questionnaires, especially for semi- or unstructured interviews. A solution to the time-intensiveness of interviewing individual people is to interview people in focus groups. That way, data from multiple people can be gathered at the same time, with the added benefit that people can also encourage each other to speak their minds. On the other hand, however, leading successful focus groups is a true skill, according to Lazar, Feng and Hochheiser, because a researcher must be able to deal with less favorable group dynamics as well. For instance, some persons may be so outspoken that they don't give others a chance to speak and that need to be reined in in a tactful manner so as not to offend them, whereas others may tend to keep silent and need to be encouraged to speak.

Interviews were found to be used in sixteen papers in this study. One study reported using a focus group: Mills, Chandra, and Park (2013) used a focus group "to obtain student's reflections on the problem solving learning experiences" (p. 320). As with surveys, interviews were found to contain questions relating to different domains: robot-related, interaction-related, task-related, and child-related questions. Table 2 shows these domains with examples of topics that were researched within them.

Table 2 Domains of questions researched using interviews and examples of themes within these domains.

	Number of papers	**Examples**
Robot-related questions	6	*Mechanical properties, robot's animacy, intentions, mental states, and morality with regard to a robot*
Interaction-related questions	5	*Robot's role in interaction, experienced fun*
Task-related questions	4	*Questions relating to the task that children carried out*
Child-related questions	3	*Experience with robot, thoughts and feelings towards or about the robot*

ROBOT-RELATED INTERVIEWS

Several studies used interviews to understand children's view of robots and their properties. Kahn et al. (2006), Melson et al. (2009), Okita, Schwartz, Shibata, and Tokuda (2005), and Saylor and Levin (2005) interviewed children to discover what animistic and biological qualities they attributed to the robot. Saylor also asked children about the robot's mechanical properties, and Okita and Melson both asked children about the robot's psychological properties, asking about the robot's intentions and mental states, respectively. Melson furthermore researched children's moral view with regards to robots, and whether they viewed robots as companions. Beran et al. (2013) asked children about their acceptance of the robot and Leite et al. (2014) asked about their perception of the robot's behavior.

INTERACTION-RELATED INTERVIEWS

Only four studies reported asking children questions that can be categorized as interaction-related. Bethel, Stevenson, and Scassellati (2011) asked children how much fun they thought interacting with the robot was, and Han, Jo, Park, and Kim (2005) asked children about the robot's role in the interaction. Turkle, Taggart, Kidd, and Dasté (2006) report of having conversations with children in a natural setting to talk about the children's interaction with robots. This insinuates that the conversation had the form of a non-structured interview. Yamazaki et al. (2013) held a group interview with children about their interaction with a robot.

TASK-RELATED INTERVIEWS

Bethel, Stevenson, and Scassellati (2011) told children a secret that they were to keep, and then investigated if they were as inclined to share secrets with a robot as with a human. In a structured interview with questions with levels of prompting ranging from no prompting at all to explicitly asking what the secret was, Bethel assessed children's willingness to share the secret. Hsiao et al. (2012) read toddlers a short story, about which they then asked the children questions to test their understanding of the story.

CHILD-RELATED INTERVIEWS

In the category of child-related interviews, Oh and Kim (2010) asked children questions about their experience with robotics, such as whether children had robotic pets themselves. Ribi et al. (2008), Sipitakiat & Nusen (2012), and Turkle et al. (2006) all researched children's thoughts and feelings about and towards robot, and lastly, Ribi also asked children about their preference of robot.

INTERVIEWS: DISCUSSION

Similar to surveys, successfully answering interview questions involves understanding the question, retrieving the information needed to answer the question, summarizing this information into an answer and reporting the answer to the interviewer. Because of these similarities with surveying, Scott (1997) and De Leeuw (2011) posed that children younger than seven years of age do not possess the cognitive capacities to be systematically surveyed or interviewed. However, research on interviewing young children about adverse past events (such as hospital stays or medical procedures) has shown that children as young as three years old are capable of recalling experiences just as accurately as seven-year-olds, albeit less precisely and less consistent over multiple interviews (Steward et al., 1996). Research in the medical field on this topic has revealed that children's memory capacity is a factor in how much they recall from past events, and that this is influenced by age. For instance, Steward et al. found that six-year-olds were able to remember significantly more details from a stressful medical examination when interviewed immediately after the examination than three-year-olds.

Another issue to consider when interviewing young children is that children rely on mental scripts based on their past experiences to understand and explain real-life situations. According to Docherty and Sandelowski (1999), "script knowledge is obtained by asking "what-happens-when" types of questions" (p. 179). They describe that preschool children were found to be unable to answer questions about an event they had recently experienced, but that they were able to describe a great deal of what usually happens in such an event, thereby also narrating details of what had happened during the event of interest.

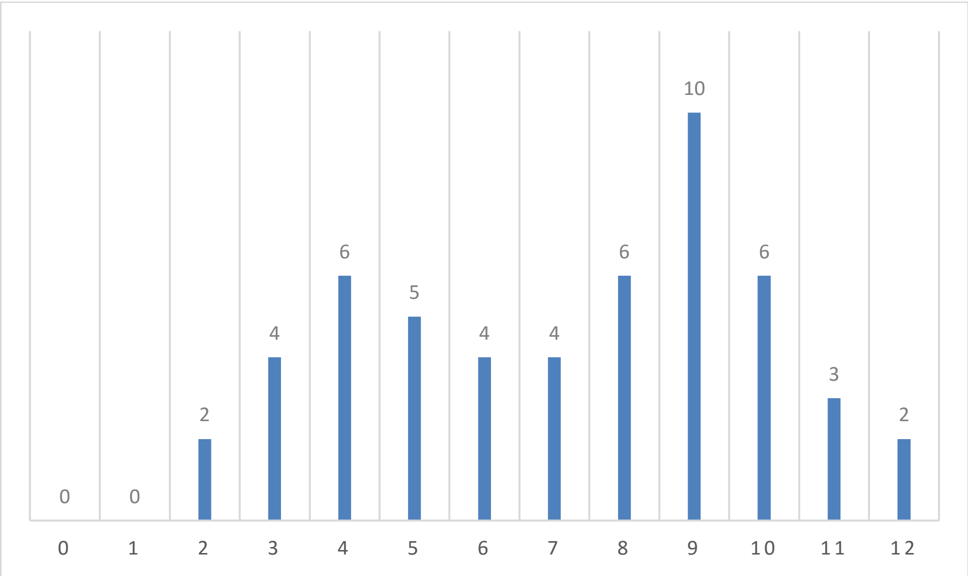

Figure 6. Histogram of ages of participants in studies that reported using interviews.

Even though these are interesting findings, they are all based on interviewing children about negative experiences. In the papers in this literature study, interacting with robot is not intended to be a negative experience. In fact, in the study by Beran, Ramirez-Serrano, Vanderkooi and Kuhn (2013), the robot was even used with the specific aim of making the unpleasant event of receiving a vaccination a less negative experience for children by distracting them from the pain of the injection. This raises the question of whether the same findings will hold true in neutral or even pleasant situations as well. In addition, interacting with a robot is a new experience for many children – an experience for which they may not yet have formed mental scripts they can rely on to answer ‘what-happens-when’-questions. Therefore, it is unclear whether children younger than seven years old are able to successfully parttake in interviews about interacting with robots.

Because of the fact that interviews rely on children’s narrative competences, Steward et al. (1996) found children’s language age scores to be a better predictor of children’s abilities to understand the purpose of an interview than their biological age. They asked children open-ended questions about their pediatric visit and found that children with a lower language age needed a direct question to get them started talking, whereas children with a higher language age started answering the open-ended question. Even though Docherty and Sandelowski (1999) reported that several studies had used brief screening tools to assess children’s developmental age, the same that applied for surveys holds true here: screening children for their developmental age is a time-consuming practice that may not be worth the effort in every context. Therefore, it is advisable to only interview children older than seven years of age in situations where assessing their cognitive capacities is not an option. Nevertheless, as can be seen in Figure 6, children under the age of 7 were interviewed in a number of studies. Two studies even reported interviewing children as young as 2 years old. Kahn, Friedman, Pérez-Granados, and Freier (2006) interviewed children from the age of 34 months about a robotic dog’s biological properties, mental states, social rapport, animacy and mortal standing. Hsiao, Chang, Lin, and Hsu (2012) asked children aged between 2 and 3 years old questions about stories they had read.

3.4 OBSERVATION

As can be seen in Figure 1, behavior analysis is the method most used in CRI literature, with 87 studies reporting of it. Even though the method was found to be used to investigate a wide variety of themes within the field of CRI, one thing that almost all of these studies have in common is that they recorded videos of the interactions for later analysis rather than scoring behavior in real-time. Exceptions to this are Kim et al. (2010) and Beran, Ramirez-Serrano, Vanderkooi, and Kuhn (2013). Kim et al. tested a PDA-application that allowed therapists to measure the frequency and duration of stereotypical autistic behavior in children in real time during interaction with a robot. Beran et al. used a combination of assessing children’s facial expressions of pain on questionnaires during the experiment and coding video recordings of the experiment with the Behavioral Approach-Avoidance Distress Scale (BAADS) to determine whether distraction by a robot during a flu vaccination can reduce children's pain and distress.

As was the case with surveys and interviews, four domains of observed behavior were identified upon analysis of papers in the literature study: robot-related, task-related, child-related and interaction-related behavior. Table 3 shows for each of these domains which topics were researched, with examples of behaviors that were observed. In the next section, each of the four domains of behavior will be discussed in more detail.

CHILD-RELATED BEHAVIOR

Out of the 88 papers that reported observing behavior, 75 observed child-related behavior. This kind of behavior includes displays of emotion through facial expressions (e.g. Blanson Henkemans et al., 2013; Keren, Ben-David, & Fridin, 2012; Okita, Schwartz, Shibata, & Tokuda, 2005; Robins, Dickerson, & Dautenhahn, 2005; Schiavone et al., 2011) or vocal emotional display (Fridin & Belokopytov, 2014a; Fridin, 2014a, 2014b; Keren et al., 2012). Another type of often-researched behavior was behavior that indicates children's attention. This includes eye gaze (e.g. Baxter et al., 2013; Costa, Lehmann, Dautenhahn, Robins, & Soares, 2015; Hsiao, Chang, Lin, & Hsu, 2012), gaze shift (Wainer, Robins, Amirabdollahian, & Dautenhahn, 2014; Wainer, Dautenhahn, Robins, & Amirabdollahian, 2014) and head movement. Other types of movement, such as gestures, pointing, and moving to or from the robot or interaction partner were also frequently observed (for gestures, e.g. Alac, Movellan, & Tanaka, 2011; Andrés, Pardo, Díaz, & Angulo, 2015; Feil-Seifer & Matarić, 2009; for pointing e.g. Boekhorst, Walters, Koay, Dautenhahn, & Nehaniv, 2005; Joshua Wainer, Ferrari, Dautenhahn, & Robins, 2010; for moving to/from robot, see for instance Hyun, Kim, Jang, & Park, 2008; Michaud et al., 2005).

Children's communication was frequently observed as well, for instance children's speech directed at a robot or related to the robot (e.g. Kahn, Friedman, Pérez-Granados, & Freier, 2006; Ros et al., 2014) positive or negative comments (André et al., 2014) or the amount of words in an utterance (Wood, Dautenhahn, Lehmann, et al., 2013; Wood, Dautenhahn, Rainer, et al., 2013).

Table 3. Domains of observed behavior and examples of themes within these domains.

	Number of papers	Examples
Robot-related behavior	5	
Robot action	1	*Moving an object, preparing to move an object*
Robot performance	2	*Speed, accuracy, activity during play*
Robot's use characteristics	1	*Intuitiveness of interface*
Interaction-related behavior	47	
Interaction with robot	20	*Duration of interaction, interaction patterns, initiation of interaction*
Physical interaction with robot	19	*Physical proximity, physical interaction, manipulation of robot (buttons, touching),*
Engagement in interaction	7	*Active engagement with robot*
Interest in interaction(partner)	4	
Interactive behaviors	8	*Reciprocity, mistreatment of robot, playing together with robot or other child, care-giving actions towards robot*
Interaction among children	4	*Interpersonal relationships, group proxemics, collaboration*
Task-related behavior	18	
Child's response	7	*Responding to questions or verbal cues, response time, correct/incorrect response*
Errors	3	*Debugging of a robot*
Performing task	3	
Performance on task	6	*Identifying body parts, storytelling*
Child-related behavior	75	
Emotional expression	27	*Facial expressions, vocal emotional display, affective display*
Understanding the robot	1	
Reaction to robot malfunction	2	*Child's reaction to unexpected malfunctions or failures of social contingency.*
Attention	36	*Attention to robot, eye gaze, gaze shift, eye contact with robot or human, head movement*
Movement	18	*Pointing, moving to or from robot, gestures, following robot*
Imitation	6	*Of robot's facial expressions and gestures*
Social robot-directed behavior	3	*Approach/avoidance of robot, responsiveness to robot*
Communication	27	*Robot-directed or robot-related speech, positive and negative comments, word count of utterances*
ASD stereotypical behavior	6	*Behaviors that are stereotypical to children with Autism Spectrum Disorder, such as rocking or hand flapping*
Other behaviors	9	*Concentration, reading, play behavior, personality traits*

Because several papers in the literature study researched the application of robots in therapy for autistic children, stereotypical autistic behaviors were found to be observed in six of them. These behaviors can include flapping of the hands or rocking with the whole body (e.g. Pioggia et al., 2007; Shamsuddin, Yussof, Mohamed, Hanapiah, & Ismail, 2013; Soares et al., 2013).

INTERACTION-RELATED BEHAVIOR

After child-related behaviors, interaction-related behaviors were most frequently observed, with 47 papers mentioning them. These behaviors include initiation of interaction with the robot (Huskens, Palmen, Van der Werff, Lourens, & Barakova, 2014; Ribi et al., 2008), interaction patterns (e.g. Wyeth & Wyeth, 2008), and the duration of interactions (e.g. Kanda & Ishiguro, 2006; Simo, Nishida, & Nagashima, 2006).

Physical interaction was also frequently observed. Measures included physical proximity to the robot (e.g. Duquette, Michaud, & Mercier, 2008; Melson et al., 2009; Oh & Kim, 2010), manipulation of the robot by for instant pressing its buttons, such as in Ko, Lee, Nam, Shon, and Ji (2010). Tactile exploration of the robot was also observed (Kahn et al., 2006; Kozima, Nakagawa, & Yasuda, 2005; Melson et al., 2009).

Children's active engagement in the interaction with the robot was observed six times, including by De Greeff et al. (2013). Different interactive behaviors were observed, such as mistreatment of the robot (Kahn et al., 2006), care-giving actions towards the robot (Ghosh & Tanaka, 2011), and absence of sharing between child and robot (Duquette et al., 2008).

ROBOT-RELATED BEHAVIOR

Only five papers reported observing robot-related behavior. These behaviors include robot actions, such as moving an object (Baxter et al., 2013), robot performance measured by speed and accuracy (Boccanfuso & O'Kane, 2011) and the robot's activity during play (Kerepesi, Kubinyi, Jonsson, Magnusson, & Miklósi, 2006), and the intuitiveness of the interface (Kronreif et al., 2005).

TASK-RELATED BEHAVIOR

The last domain of behaviors are task-related behaviors. This domain includes behaviors that pertain specifically to the task that children were presented with. For instance, Sipitakiat and Nusen (2012) tested a way of involving children in the debugging process of robot development. Therefore, they observed behaviors that were related to debugging. Other studies focused on how well children performed a task, for instance in Costa, Lehmann, Dautenhahn, Robins, and Soares (2015), who used a robot to teach autistic children body awareness and appropriate physical interaction. They observed how well their participants were able to identify body parts.

OBSERVATIONS: DISCUSSION

It is not surprising that observations was found to be the most used research method in literature. Other than interviews and surveys, observational research can be carried out unobtrusively and it does not rely on children's reading abilities or abilities to understand and answer questions, which makes this method suitable for use even with very young children. In fact, Fridin (in Fridin (2014a, 2014b) and Fridin and Belokopytov (2014b)) researched three-year-olds and even states that "considering the age of our target population, post-hoc analysis of video footage of interaction sessions is the only feasible method of data collection and analysis". This assertion is in line with the ages of participants researched using observations. Whereas the majority of studies using surveys and interviews reported using them with children aged 7 or older, observations were found to be used frequently across a wider range of ages, as can be seen in Figure 6.

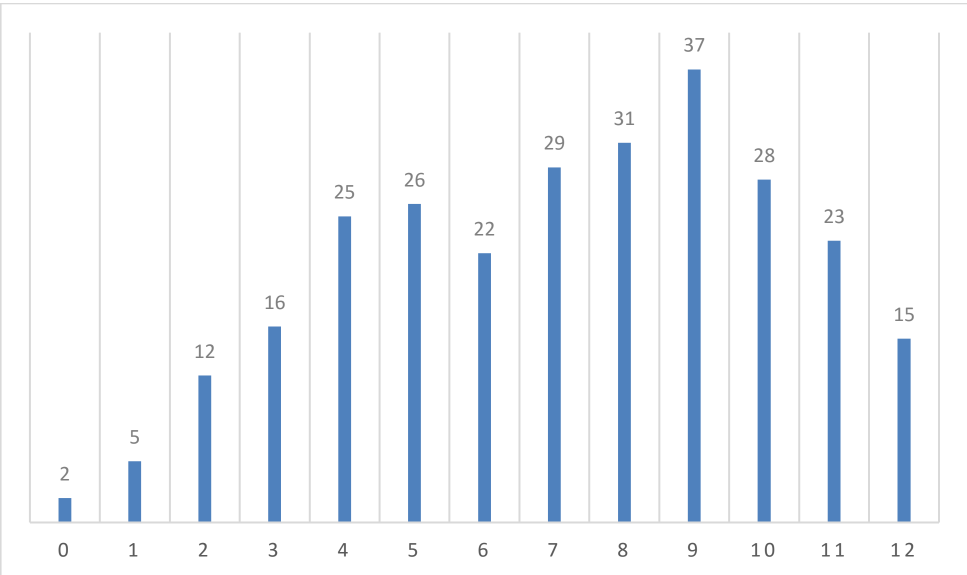

Figure 7. Histogram of participants' ages in studies that reported using observations.

Another advantage of analyzing video recordings of interactions, a method found to be used by nearly all studies reporting using observational research, is that it allows researchers more time to make careful observations or to notice behaviors that might be missed when observing in real time (Basil, 2011). In addition to that, video recordings can be analyzed by multiple researchers, which allows for the calculation of inter-rater reliability and thereby can produce more reliable results than with real-time observations.

Two forms of observatory research were found: exploratory research, aimed at forming theories, and experimental, aimed at testing theories. An example of exploratory observational research is the study by Kennedy, Baxter, and Belpaeme (2013), who explored children's interaction with a touchscreen as interaction mediator. In the analysis of their data, they looked for emerging interaction structures. An example of experimental observational research is the study by Hsiao, Chang, Lin, and Hsu (2012), who researched whether a robot learning companion (RLC) could increase children's learning performance in learning to read. The difference between these two types of research is that experimental observation studies often code behaviors according to predefined coding schemes that contain codes for the behaviors of interest, whereas exploratory observation studies take a more grounded-theory approach. This entails that the data is not coded according to a predefined coding scheme, but that a coding scheme is devised based on emerging patterns of behavior. One thing both approaches to observational research had in common in the literature study was that they were both found to often lack indications of reliability of the data analysis.

3.5 HUMAN MEASURES

Human measures, or physiological measures, have been found to be used five times in this study. For example, Leite, Henriques, Martinho, and Paiva (2013) measured electrodermal activity "to characterize and evaluate the interaction, and to dynamically recognize user's affective states". Pioggia et al. (2007) explored the reactions of autistic children to a highly human-like android face. Because autistic children generally have trouble expressing their emotions, they measured children's heart rate to assess their emotional reactions.

Because human measures do not rely on speech or other particular cognitive capacities, they can be used with children of all ages.

3.6 CASE STUDIES

Benbasat, Goldstein, and Mead (1987) define case studies as follows: "A case study examines a phenomenon in its natural setting, employing multiple methods of data collection to gather information from one or a few entities (people, groups, or organizations)" (p. 370).

What ‘multiple methods of data collection’ are used is dependent of what the researcher wants to investigate, and methods can include interviews, observation, and physical artifacts such as devices, outputs or tools.

In this literature study, three instances of case studies were found. De Greeff et al. (2013) researched whether a robot was able to provide hospitalized children with an environment in which they felt free to indicate their preference on what activities they wanted to play with the robot. Case studies were also found to be used to investigate children’s reactions to robots: Giannopulu and Pradel (2012) studied the interaction between an autistic child and a robot, and similarly, Dickerson, Robins, and Dautenhahn (2013) investigated interactions between autistic children, a co-present adult and a robot.

Because case studies can use different research methods, there is no clear answer as to for which ages case studies are suitable. Instead, this depends on the exact research methods used. For example, De Greeff and colleagues used video footage of children interacting with the robot, as well as game activity logs, semi-structured interviews with participants, and “notes by care professionals on robot related talk from the child”. Because the participants in this study were aged between 7 and 11 years, these research methods are suitable for these participants, considering that children can be successfully interviewed from age 7. However, these same methods would not have been suitable for use with children younger than 7 years old.

3.7 ETHNOGRAPHY

Ethnography is a research method that stems from anthropological research and that aims to understand and explain human behavior through engaged, involved research rather than through “limited interactions and observations” (Lazar, Feng, & Hochheiser, 2010, p. 219). In ethnography, researchers immerse themselves in the group of people they research in order to, as Lazar and colleagues put it, “come as close as possible to achieving the rich perspective that comes from being part of the group being studied” (p. 221). Watson-Gegeo (1988) distinguishes between ethnography as a product and a process. She states that “As product, ethnography is a detailed description and analysis of a social setting and the interaction that goes on within it” (p. 582). Ethnology as a method “includes the techniques of observation, participant-observation (observing while interacting with those under study), informal and formal interviewing of the participants observed in situations, audio- or videotaping of interactions for close analysis, collection of relevant or available documents and other materials from the setting, and other techniques as required to answer research questions posed by a given study" (p. 583). Because the focus of this study is on research methods for CRI research, ethnography as a process is of particular interest here.

One paper was found to use ethnography: Mills, Chandra, and Park (2013) carried out enthographical research in a classroom setting to examine "the convergence of speech and practical activity in children's collaborative problem solving with robotics programming tasks" (p. 315). They included many of the research techniques mentioned above: "micro-genetic analysis of students' speech interactions with tools, peers, and other experts, teacher interviews, and student focus group data" (p. 315) were used to collect data from the 8-9 year old students they researched.

ETHNOGRAPHY: DISCUSSION

Because ethnography as a method is a combination of different research methods, its suitability for use with children depends on the exact research methods used and the ages of participants. In the case of Milss, Chandra and Park, the methods chosen seem suitable for use with their participants because apart from the focus group, all methods they used can be used with children of any age. Furthermore, the participants in their study were, according to their age, cognitively sufficiently developed to take part in focus groups. However, an ethnography consisting of the same set of research methods would not be suitable for children younger than seven years old due to them lacking the cognitive capacities to successfully take part in focus groups.

3.8 USABILITY TESTING

Lazar, Feng and Hochheiser (2010) define usability testing as follows: "Usability testing, in general, involves representative users attempting representative tasks in representative environments, on early prototypes of computer interfaces" (p. 252). According to this definition (when replacing 'computer interfaces' with 'robotic agents'), fifteen studies were found to use usability testing. Unsurprisingly, ten of them focused on robot design. For example, Herberg, Behera, and Saerbeck (2013) researched which robot embodiment was perceived by children as the most effective tutor, and Mazzei et al. (2012) researched whether a robotic face designed for expressing emotions was capable of teaching autistic children to recognize their interlocutor's facial expressions.

The other five studies focused on theory. Kahn, Friedman, Pérez-Granados, and Freier (2006) and Melson et al. (2009) both used card sorting to "assess children's judgments about AIBO's relative similarity to other potentially related artifacts" (Kahn et al., 2006, p. 413). They aimed to get insight to children's understanding of the robotic dog AIBO by presenting them with pictures of pairs of objects such as desktop computer/live dog, live dog/stuffed dog, and robot/live dog. Children were then asked which item of each pair resembled AIBO most. Sapounidis, Demetriadis, and Stamelos (2014) researched the effect of a tangible user interface on children's performance in robot programming. They presented their participants with a programming task, which children in one condition completed using a graphical user interface whereas the other condition used the tangible user interface that consisted of interconnecting blocks. Encarnação et al. (2014) evaluated "the feasibility of using virtual robot mediated play activities to assess cognitive skills". To that end, children performed cognitive tasks with both a robot and virtual representation of a robot to see whether the results for the virtual robot were the same. Lastly, Costa, Lehman, Dautenhahn, Robins, and Soares (2015) researched whether a humanoid robot could be used to teach children with autism appropriate physical interaction.

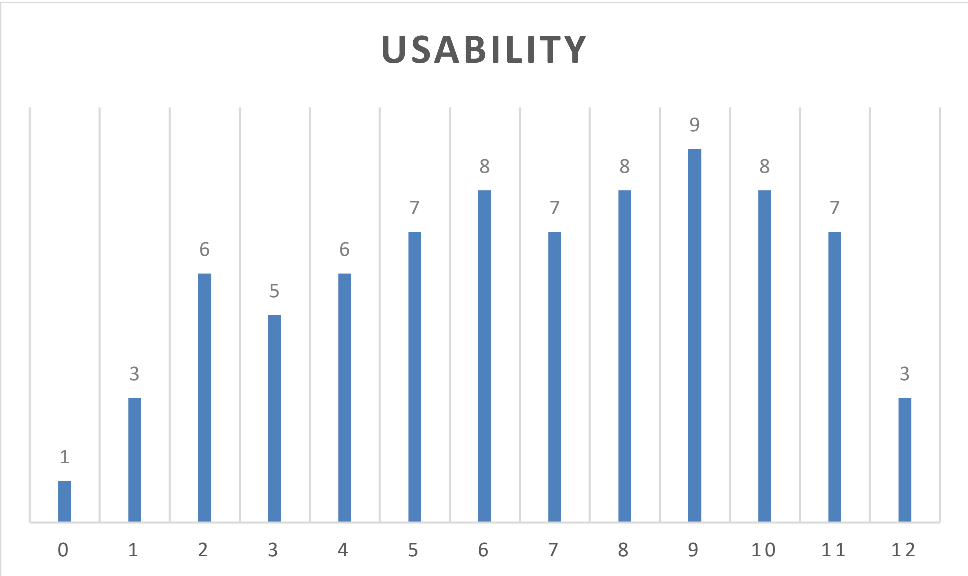

Figure 8. Histogram of participants' ages in studies that reported using usability studies.

In the field of Human-Computer Interaction (HCI), usability testing is a well-researched method. The method often consists of multiple research methods, including observations of posture, facial expressions and verbal utterances, analysis of computer log files, interviews, and questionnaires. Therefore, as is the case with ethnography, it depends on the exact methods used and participants' ages whether or not usability testing is suitable for children.

Hanna, Risden, and Alexander (1997) composed guidelines for usability testing with children. They posed that children younger than 2,5 years generally are not able to engage in usability testing because they lack experience in operating standard computer input devices such as mouse or keyboard. Yet, as can be seen in Table 8, children younger than 3 have been found to participate in usability testing. For instance, a 20-month-old child participated in the study by Feil-Seifer and Matarić (2009), who researched how children with autism reacted to socially assistive robots in a therapeutic setting. The robot used in this experiment had buttons that children could push, instead of the input devices typically used in HCI-research. This could well be the reason that children younger than 2,5 were able to partake in usability testing with a robot, whereas the same group of children may not have been able to do so in the field of HCI. However, Hanna, Risden and Alexander indicated that children aged 2-5 require extensive adaptations for them to be able to participate in usability testing. They pose that children in this age group can best be allowed to explore the computer interface on their own without being guided by a specific task, and that “appeal or engagement” are best measured through observation instead of surveys, because it is hard for such young children to verbally express their likes and dislikes.

Usability testing with children aged 6-10 is, according to Hanna and colleagues, more easy because “their experience in school makes them ready to sit at a task and follow directions from an adult” (p. 10). They also pose that children in this age group are generally able to answer questions, but that six- and seven-year-olds may be more inarticulate in doing so. This statement is in accordance with earlier discussed abilities of children to express themselves verbally on surveys or interviews. However, Donker and Reitsma (2004) found that although 6-year-olds did not say much in a think-aloud task, their remarks were very valuable.

4. DISCUSSION

In the previous section, the results from the literature study were presented and the research methods found were discussed in light of their suitability for use with children of different ages. In this section, these results will be combined to determine which methods can be used to answer predominant research questions in CRI witch children of different ages. Then, a brief discussion of the most important methodological requirements of each research method and the status quo as found in the literature study will be discussed.

4.1 ANSWERING THE MAIN QUESTIONS IN CRI

Table 4. Types of research questions with advised research methods and possible measures for children under and over the age of 7.

	< 7 years		≥ 7 years	
Question	Method	Measures	Method	Measures
Mental models	Card sorting	Most important traits, preference of embodiment	Card sorting Surveys Interviews	Most important traits, preference of embodiment.
Emotional reaction	Observation	Facial expressions, posture, gestures, laughing, positive/negative utterances	Observation	Facial expressions, posture, gestures, laughing, positive/negative utterances
	Human measures	Heart rate, electro-dermal activity	Human measures	Heart rate, electro-dermal activity
			Surveys Interviews	Fun, acceptance, perceived closeness
Influence on...				
Verbal behavior	Observation	Number of utterances, positive/negative utterances	Observation	Number of utterances, positive/negative utterances
Nonverbal behavior	Observation	Gestures, actions, proximity to robot, direction of gaze, engagement (active time)	Observation	Gestures, actions, proximity to robot, direction of gaze, engagement (active time)
Performance	Observation	Verbal or nonverbal behavior, see above	Observation	Verbal or nonverbal behavior, see above
			Surveys	Knowledge test

N.B.: Please note that this table does not present an exhaustive list of possible measures but rather some of the most-found measures in the literature study.

In the literature study, papers were found to ask research questions that can be categorized in three main types of questions: questions regarding children's conceptual perception (mental models) of robots, their emotional response to robots, and the robot's influence on children's behavior or performance. In this section, knowledge from the discussions about each research method's suitability for children is applied to these types of questions to come to recommendations on which methods can be used to answer them with children of different ages. An overview of these recommended methods for each type of questions with commonly found measures is given in Table 4. This table will now be explained in more detail.

Questions about children's mental models and conceptual perceptions of robots were found in several studies. For instance, Okita, Schwartz, Shibata, and Tokuda (2005) researched "whether children can apply animistic intuitions to robotic animals, and whether these attributes vary by the child's age and the robot's behaviors and appearance", and Han, Lee, and Cho (2005) researched what role (friend, secretary or machine) robots were perceived as fulfilling during interaction. To answer questions such as these, the most obvious choice is to ask children themselves. Therefore, questionnaires and interviews are a good choice to gain insight in children's perceptions of robots. As discussed earlier, interviews and questionnaires are most reliably used with children older than 7. However, that does not mean that the perceptions of children younger than 7 cannot be researched. To measure younger children's perceptions, a visual method could be used. For example, Kahn, Friedman, Pérez-Granados, and Freier (2006) studied children's reasoning about robotic pets with regards to robots' animacy. Kahn and colleagues presented their participants, who were aged between 2 and 6 years, with cards with pictures of pairs of two artifacts (stuffed dog, robot, desktop computer or real dog). Children were then asked to indicate whether a robotic dog was more similar to one or the other artifact pictured. This way, Kahn and colleagues gained insight in the children's mental models of the robotic dog. To design a tutor robot, Herberg, Behera, and Saerbeck (2013) used a combination of a similar method and a card sort task to gain insight into what traits children find most important for tutors (card sort task) and which robotic embodiment they perceived as best fulfilling these traits. Even though the participants in this study were aged between 8 and 11, card sorting may also be a useful method for younger children. This possibility will be discussed more in depth later.

Children's emotional reaction to robots was found to be frequently researched, with fun being researched the most. As was the case for mental models, the self-report methods of surveys and interviews seem to be an obvious first choice here. Tools such as the Smileyometer and non-visual Likert-scales are frequently used to measure fun, as well as other affects. For younger children, observations can be used to study facial expressions and body posture as indications of children's emotions. In addition to that, verbal indications of emotions, such as children saying that something is fun, can also be used. Lastly, psychophysiological measures can be used to measure emotions. In the literature study, heart rate and electro-dermal activity were found to be used, but only on a few occasions. Leite, Henriques, Martinho, and Paiva (2013) researched the possibility of measuring electro-dermal activity (EDA) to measure children's affective states. They found evidence that engagement, motivation and attention could be measured using EDA. Because measuring EDA and other human measures such as heart rate does not rely on children's capabilities to fulfill self-report measures such as surveys, human measures can be used with children of all ages.

The third type of research question concerned the robot's influence on children. More specifically, the robot can influence children's verbal and nonverbal behavior, and their performance. When behavior is the topic of research, observation is the method that first comes to mind. Because it can be carried out unobtrusively and does not rely of chidren's verbal and mental capacities, observation is a suitable method for children of all ages. Measures of verbal behavior that were found in the literature study include the number of utterances made and whether utterances were positive or negative. Nonverbal behavior includes gestures, proximity to the robot, direction of gaze, and time spent actively interacting with the robot. For children's performance, the way performance is operationalized defines the way it can be measured. If the performance in question is nonverbal, then measures for nonverbal behavior can be used. Similarly, if verbal performance is measured, this can be done by the same measures as for verbal behavior. However, performance cannot always be measured through observation of behavior. For instance, Han, Jo, Park and Kim (2005) and Hyun, Kim, Jang and Park (2008) researched whether a robot was as good in instructing children who were learning English as books or computer instruction. They measured students' improvement in English with a paper and pencil test. Therefore, surveys can also be used as pretest and posttest to measure performance that is not behavioral.

4.2 THE MAGICAL THRESHOLD OF 7

In line with literature that was presented earlier on children's capabilities to take part in self-report research, Table 1 shows a clear divide between children under and over 7 years of age. This divide is due to the fact that only children aged 7 or older are capable of successfully taking part in self-report research. In this section, we'll take a closer look at children's cognitive development around age seven to determine what it is that makes this age so magical that it allows children to be surveyed and interviewed whereas this is not possible with younger children.

The answer to this question is explained by Piaget (2002). In short, the age of 7 marks the end of the development of intuitive thought and the beginning of the operational stage. In children younger than 7, specifically children between 3 and 7, egocentric thought dominates. This is thought that children cannot communicate to others. Piaget illustrates this type of thought with the example of understanding something but being unable to explain to others what it is you understand. Characterized by its intuitive nature, meaning that children often cannot explain how they know what they know, egocentric thought is not deductive and not logical yet. According to Piaget, children older than 7, egocentric thought has been diminished and largely replaced by communicated intelligence. In contrast to egocentric thought, communicated intelligence is more logical and does rely on deduction and proof instead of on intuition. On top of that, this kind of thought is communicable to others. It is this shift from non-communicable egocentric thought to communicated intelligence that allows children to be surveyed and interviewed from age 7.

As mentioned earlier, the age of 7 is not a strict limit. Even though each child moves through each of Piaget’s stages of cognitive development in the same order, not all children do so at the same pace. Some children younger than 7 may be developed enough to successfully take part in self-report research, and there may be children older than 7 that are not. However, as was found by Shields, Palermo, Powers, Grewe, and Smith (2003) with regard to visual analogue scales (VAS), the chance that children with average IQ were able to successfully complete a VAS was greatest when children were older than 7 than younger. This thesis aims to give general advice on which methods are suitable for use with children of different ages, and therefore it is advised to only use surveys and interviews with children aged 7 or older, unless the researcher deems it worth the time and effort to pretest children to determine whether they possess sufficient cognitive capacities to successfully take part in self-report research.

4.3 STATUS QUO OF METHODOLOGICAL REQUIREMENTS

During analysis of the papers in the literature study, it became apparent that quite a few studies seemed to neglect the most important methodological requirements of the research methods they used. Each research method has certain requirements that need to be met in order for the method to result in valid and reliable data. In this section, the research methods that were found to be lacking in this regard will be discussed in terms of these requirements and their status quo as revealed by the literature study.

Being one of the most-used research methods, it was quite surprising to often find insufficient reports of validation of questionnaires. Often, researchers reported constructing their own questionnaires – either from scratch or by combining (parts of) existing questionnaires. However, many times, no measure of the resulting questionnaire's validity was reported. Without these measures, one cannot be sure that the questionnaire measures what it is intended to measure. Even when parts are taken from questionnaires that have been found valid, the parts are not necessarily valid by themselves. Therefore, it is important that researchers that take the time to construct new questionnaires also take the time to validate them. Of course, this can be a time-consuming practice, but there may be another solution. A plethora of questionnaires already exists, including many questionnaires that have been found to be valid. There is no need to reinvent the wheel by developing new questionnaires when the constructs of interest can also be measured with existing ones.

Another method that was often found failing to meet critical methodological requirements was observations. In observational research, reliability of codings is most important to assure valid results. However, many observational studies failed to report measures such as inter-rater reliability, which indicates that the results found are not just one person's interpretations of the data but sufficiently consistent across different coders. Of the studies lacking such indications of reliability, the majority of the studies was exploratory. This entails that the observations are made without a pre-existing coding scheme to see what behavioral patterns emerge. Often, a coding scheme is the result of such studies instead of the starting point. Still, reliability is important in this type of research too. To be able to attach value to the findings of observational studies, researchers ought to have their data coded or interpreted by more than one coder.

For usability studies, an important requirement is the number of participants. In papers included in the literature study, usability studies were reported with sample sizes ranging from 4 to 109 participants. As with any research method, large sample sizes generally produce more reliable results. However, as Schmettow (2012) describes, there has been a long-standing debate in the field of HCI about what a sufficient sample size is. Some people claim that as few as five participants is enough to discover 85% of all usability problems, whereas other maintain that the golden rule is to use 10±2 participants. In a reaction to those claims, Schmettow shows that a magic number for sample size in usability studies does not exist. These rules are calculated from the probability that a usability problem is detected and based on the assumption that this probability is the same for every study. In 2008, Schmettow presented a way to determine whether this is the case and found that this probability often differed across usability studies. Instead of basing estimates of sample size on formulas that assume this probability to be the same every time, Schmettow (2012) proposes a way to determine sample size by continuously monitoring a usability study. This is done by calculating, based on the number of usability problems that have been found in a study, the amount of usability problems that have yet to be discovered. The proportion of yet undiscovered usability problems then informs whether more testing (and thus more participants) is needed, until the desired proportion, for instance 80%, of usability problems has been detected.

In the papers in the literature study, none of the studies that reported usability testing detailed how they had arrived at their sample sizes. The objective of these studies may not have been to detect as many usability problems as possible but to test whether or not robots had the desired effect on children or whether their functionality worked as intended. However, in order for readers as well as the researchers themselves to base trust in the findings of their usability studies, it would have been best if they had shown consideration of their sample sizes.

In short, it can be concluded from this that the research methods that are found to be used in CRI research are not always used well. Failing to take into account the fundamental methodological requirements and considerations of research methods can result in invalid and unreliable data, as well as low replicability of the results. The latter has been much discussed recently in the field of psychology, where a replicability crisis has been reported (Open Science Collaboration et al., 2015).

4.4 MULTIPLE AND MIXED METHODS RESEARCH

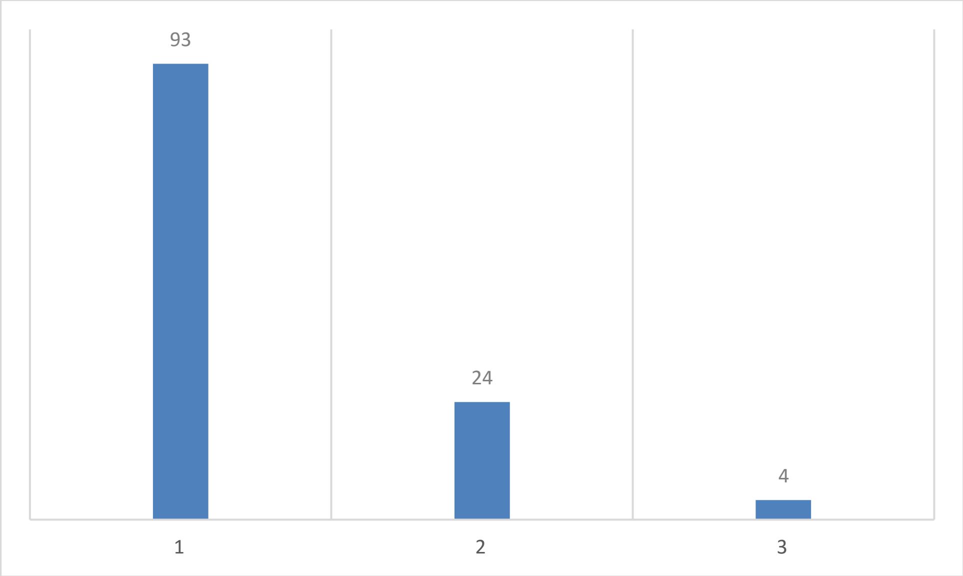

Figure 9. Histogram of the number of research methods used in papers included in the literature study.

Another curious finding from analysis of papers included in the literature study is that the vast majority of the papers (77%) reported using only one research method, despite the wide attention that multiple and mixed method research has received in literature.

Mixed methods research uses a combination of quantitative and qualitative research methods. It aims to use the strengths of both of these approaches and to minimize the weaknesses, resulting in more superior research than when only one research method is used (Johnson & Onwuegbuzie, 2004). Greene, Caracelli, and Graham (1989) described five purposes of mixed methods research. The first, triangulation, entails that data gathered from one research method can be used to corroborate or verify the results from another research method. Secondly, different research methods can be used to complement each other. Furthermore, mixed methods research can be used for "discovering paradoxes and contradictions that lead to a re-framing of the research question" (Johnson and Onwuegbuzie, 2004, p. 22). Greene and colleagues called this 'initiation'. A fourth purpose of mixed method is development, which entails that one research method is informed by findings from another. Lastly, mixed methods research can be used for expansion, or, as Johnson and Onwuegbuzie paraphrase: "seeking to expand the breadth and range of research by using different methods for different inquiry components". In research with children, especially with younger children who are not yet capable of participating in self-report research such as surveys and interviews, combining the research methods that are suitable for them can help to yield as rich information from them as possible. Combining methods – complementing them – can help to gain a more complete picture of children's experiences with robots. Considering these benefits, it is surprising that not more studies were found to use more than one research method. Perhaps this is due to the knowledge gap in what suitable research methods for children are, perhaps researchers don't know what other methods they could use. Of course, the choice of research methods is informed by the research question, so not all studies may benefit from using more than one method. However, considering the above described lack of consideration of fundamental methodological issues, using mixed methods may help with the reliability of results by triangulating them with results from another research method.

4.5 LIMITATIONS

As explained earlier, the choice to use Piaget's classification of children's cognitive development was used because a global classification of children's development at different ages was necessary to come to a global recommendation of suitable research methods for children of different ages. However, due to this choice, the focus of this thesis was purely on children's cognitive development. Other developmental researchers, for instance Vygotsky, have stressed the importance of social development for the development of cognitive capacities. This is outside the scope of this thesis, but nonetheless an important aspect to consider when choosing research methods for children. For instance, a child with autism may be cognitively developed enough to successfully fill in a survey. At this stage, it is to be expected that a typically developing child can also be interviewed. However, due to the social impairments that are characteristic of Autism Spectrum Disorders, an autistic child that is capable of being surveyed may not be able to participate in an interview due to the social demands of that research method.

Another limitation to this study is that the list of research methods in Table 4 is not an exhaustive list of research methods that are suitable for use with children. Instead, Table 4 contains the research methods that were found to be used in CRI literature. However, in the recent past, literature in other fields of research, such as sociology, has devoted attention to new participatory research methods that can be used to gain insight in children's experiences in new ways. Examples of such methods are analysis of photographs taken by children and analysis of drawings (e.g. Fargas-Malet, McSherry, Larkin, & Robinson, 2010; Barker & Weller, 2003; Einarsdóttir, 2007; Punch, 2002). Because these methods do not rely on children's linguistic development, they bear the potential of being suitable for use with children that are not yet able to be surveyed or interviewed. To our knowledge, however, no research has been done yet on the suitability of such methods for use in the field of CRI. Future research is needed to determine whether these methods can be used in CRI research, and for which age group they are suitable.

Another method that bears potential with regards to its suitability to be used with young children but that lacks research to back up this theory is card sorting. Even though several studies have used card sorting or card sorting-like tasks to research children's mental models of robots, little research on this topic is known. This is surprising, considering the possibilities of card sorting. First, card sorting tasks are traditionally done with written items on cards, for instance the titles of items in a website's navigational menu. However, card sorting can also be done with images instead of words on the cards. This way, even preliterate children can take part in card sorting tasks. On top of that, card sorting requires participants to organize cards according to their own logic. Even though verbal explanation of completed card sorts are sometimes used for clarification purposes, card sorting does not necessarily require this. Therefore, the method is also suitable for children younger than age 7, who, as discussed earlier, are not yet capable of communicating their own thought.

Applications of card sorting in CRI include gathering design requirements for robots (as Herberg, Behera, and Saerbeck (2013) did when researching which personality traits children deemed most important for a tutor robot), and gathering insight in children's preference of robotic embodiment, or what task they found most fun to do with a robot.

A variation on traditional card sorting is the Q-method, in which participants are asked to sort cards on a grid that resembles a normal distribution curve. On this grid, participants place each card with the value they feel fits each card best according to the assignment they received for the sort. For instance, Taylor and Delprato (1994) carried out a Q-sort study with young children (aged 3-4) where children were asked to sort pictures of children out of magazines on a Q-sort grid depending on different questions. An example of a question they were given was to sort these pictures according to which they felt resembled them most and least. In a way, this method resembles a Likert-scale because each card is given a value on a continuum. However, children tend to show little variation in their answers when presented with Likert-type scales. The Q-method combats this by forcing variation by assigning how many cards can be placed at each value. The neutral value receives the most cards, whereas values towards either of the extremes of the continuum receive less cards, with the most extreme values allowing for the least amount of cards each.

Future research is needed to determine whether card sorting is indeed suitable for use with young children in the context of CRI. A way to research that is to validate card sort data about, for instance, children's experienced fun during interaction with different robotic embodiments against data obtained with other methods, such as observations of children's facial expressions and utterances.

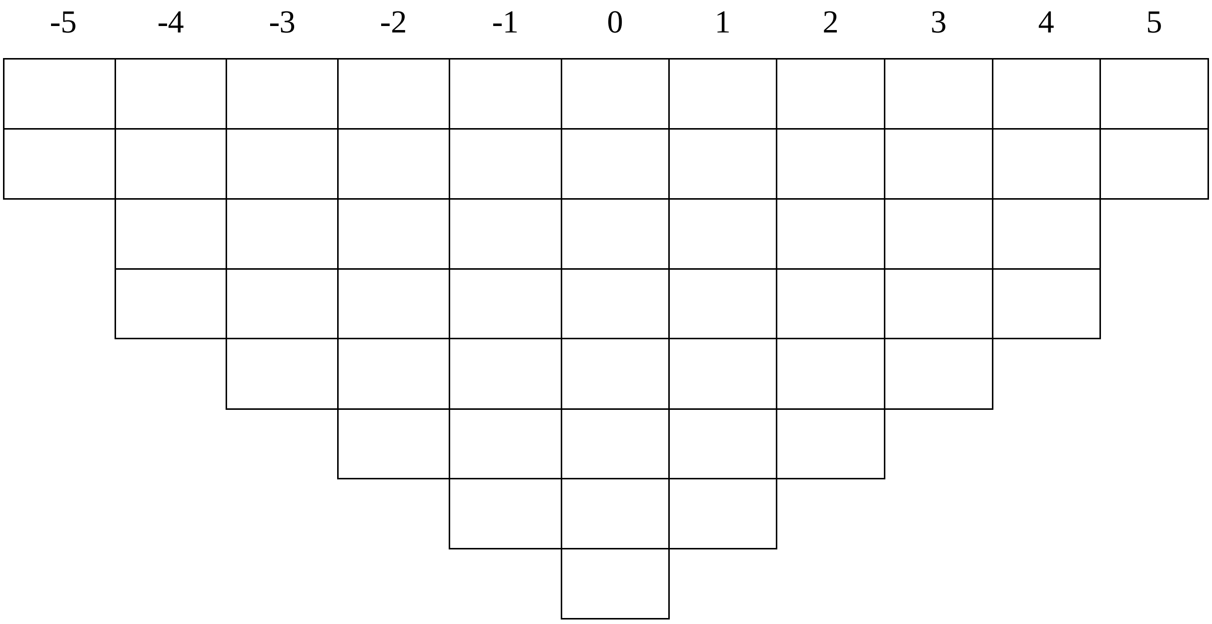

Figure 10. An example of a Q-sort grid.

5. CONCLUSION

This thesis aimed to discover which research methods are suitable for use in Child-Robot Interaction (CRI) research. To that end, a systematic literature study was performed to gain insight into which research methods were used in CRI research in the past decade. Then, these research methods were discussed in terms of children's cognitive development as described by Piaget to determine for which ages these different research methods are suitable. In short, self-report methods (interviews and surveys) proved to be suitable only for children over seven years of age, whereas other methods that did not rely on self-report, such as physiological measures and observations, are suitable for children of all ages. The age of seven is, according to Piaget, a turning point for children's ability to take part in self-report research because it marks the stage at which children become able to communicate their thought with others – a skill necessary to successfully partake in self-report research. Although Piaget's classification assumes individual differences with children of the same age and developmental stage, the short and crude recommendation of this research is to only use self-report methods with children over seven years of age, unless the time and effort of pretesting children for their cognitive development are warranted by the aim of the study (e.g. long-term studies).

REFERENCES

Alac, M., Movellan, J., & Tanaka, F. (2011). *When a robot is social: Spatial arrangements and multimodal semiotic engagement in the practice of social robotics*. *Social Studies of Science* (Vol. 41). http://doi.org/10.1177/0306312711420565

André, V., Jost, C., Hausberger, M., Le Pévédic, B., Jubin, R., Duhaut, D., & Lemasson, a. (2014). Ethorobotics applied to human behaviour: can animated objects influence children's behaviour in cognitive tasks? *Animal Behaviour*, *96*, 69–77. http://doi.org/10.1016/j.anbehav.2014.07.020

Andrés, A., Pardo, D. E., Díaz, M., & Angulo, C. (2015). New instrumentation for human robot interaction assessment based on observational methods. *Journal of Ambient Intelligence and Smart Environments*, *7*(4), 397–413. http://doi.org/10.3233/AIS-150331

AURORA. (2015). The Aurora project. Retrieved from http://aurora-project.com/

Bartneck, C., & Forlizzi, J. (2004). A design-centred framework for social human-robot interaction. *RO-MAN 2004. 13th IEEE International Workshop on Robot and Human Interactive Communication (IEEE Catalog No.04TH8759)*, 591–594. http://doi.org/10.1109/ROMAN.2004.1374827

Basil, M. (2011). Use of photography and video in observational research. *Qualitative Market Research: An International Journal*, *14*(3), 246–257. http://doi.org/10.1108/13522751111137488

Baxter, P., Kennedy, J., Belpaeme, T., Wood, R., Baroni, I., & Nalin, M. (2013). Emergence of turn-taking in unstructured child-robot social interactions. *ACM/IEEE International Conference on Human-Robot Interaction*, (2), 77–78. http://doi.org/10.1109/HRI.2013.6483509

Belpaeme, T., Baxter, P., Greeff, J. De, Kennedy, J., Read, R., Looije, R., … Zelati, M. C. (2013). Child-Robot Interaction : Perspectives and Challenges. *Social Robotics*, 452–459.

Benbasat, I., Goldstein, D. K., & Mead, M. (1987). The Case Research Strategy in Studies of Information Systems. *MIS Quarterly*, *11*(3), 369–386.

Beran, T. N., Ramirez-Serrano, A., Kuzyk, R., Fior, M., & Nugent, S. (2011). Understanding how children understand robots: Perceived animism in child–robot interaction. *International Journal of Human-Computer Studies*, *69*(7-8), 539–550. http://doi.org/10.1016/j.ijhcs.2011.04.003

Beran, T. N., Ramirez-Serrano, A., Vanderkooi, O. G., & Kuhn, S. (2013). Reducing children's pain and distress towards flu vaccinations: a novel and effective application of humanoid robotics. *Vaccine*, *31*(25), 2772–7. http://doi.org/10.1016/j.vaccine.2013.03.056

Bethel, C. L., Stevenson, M. R., & Scassellati, B. (2011). Secret-sharing: Interactions between a child, robot, and adult. *2011 IEEE International Conference on Systems, Man, and Cybernetics*, 2489–2494. http://doi.org/10.1109/ICSMC.2011.6084051

Bieri, D., Reeve, R. A., Champion, G. D., Addicoat, L., & Ziegler, J. B. (1990). The faces pain scale for the self-assessment of the severity of pain experienced by children: Development, initial validation, and preliminary investigation for ratio scale properties. *Pain*, *41*(2), 139–150. http://doi.org/10.1016/0304-3959(90)90018-9

Blanson Henkemans, O. a, Bierman, B. P. B., Janssen, J., Neerincx, M. a, Looije, R., van der

Bosch, H., & van der Giessen, J. a M. (2013). Using a robot to personalise health education for children with diabetes type 1: a pilot study. *Patient Education and Counseling*, *92*(2), 174–81. http://doi.org/10.1016/j.pec.2013.04.012

Boccanfuso, L., & O'Kane, J. M. (2011). CHARLIE : An Adaptive Robot Design with Hand and Face Tracking for Use in Autism Therapy. *International Journal of Social Robotics*, *3*(4), 337–347. http://doi.org/10.1007/s12369-011-0110-2

Boekhorst, R., Walters, M. L., Koay, K. L. K. K. L., Dautenhahn, K., & Nehaniv, C. L. (2005). A study of a single robot interacting with groups of children in a rotation game scenario. *2005 International Symposium on Computational Intelligence in Robotics and Automation*, 35–40. http://doi.org/10.1109/CIRA.2005.1554251

Borgers, N., De Leeuw, E., & Hox, J. (2000). Children as respondents in survey research: Cognitive development and response quality. *Bulletin de Méthodologie Sociologique*, *66*, 60–75.

Broadbent, E., Stafford, R., & MacDonald, B. (2009). Acceptance of healthcare robots for the older population: Review and future directions. *International Journal of Social Robotics*, *1*(4), 319–330. http://doi.org/10.1007/s12369-009-0030-6

Burghart, C., & Haeussling, R. (2005). Evaluation criteria for human robot interaction. *... : Hard Problems and Open Challenges in ...*. Retrieved from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.102.5958&rep=rep1&type=pdf#page=36

Cohen, I., Looije, R., & Neerincx, M. a. (2014). Child's Perception of Robot's Emotions: Effects of Platform, Context and Experience. *International Journal of Social Robotics*, *6*(4), 507–518. http://doi.org/10.1007/s12369-014-0230-6

Collins, D. (2003). Pretesting survey instruments: An overview of cognitive methods. *Quality of Life Research*, *12*(3), 229–238. http://doi.org/10.1023/A:1023254226592

Convention on the Rights of the Child. (1989). Retrieved February 16, 2016, from http://www.ohchr.org/EN/ProfessionalInterest/Pages/CRC.aspx

Costa, S., Lehmann, H., Dautenhahn, K., Robins, B., & Soares, F. (2015). Using a Humanoid Robot to Elicit Body Awareness and Appropriate Physical Interaction in Children with Autism. *International Journal of Social Robotics*, 1–14. http://doi.org/10.1007/s12369-014-0250-2

Dautenhahn, K. (2007). Methodology & Themes of Human-Robot Interaction: A Growing Research Field. *International Journal of Advanced Robotic Systems*, *4*(1), 1. http://doi.org/10.5772/5702

De Greeff, J., Janssen, J. B., Looije, R., Mioch, T., Alpay, L., Neerincx, M. A., Baxter, P., Belpaeme, T. (2013). Activity switching in child-robot interaction: a hospital case study. *Lecture Notes in Computer Science*, *8239*, 585–586.

De Leeuw, E. (2011). *Improving Data Quality when Surveying Children and Adolescents : Cognitive and Social Development and its Role in Questionnaire Construction and*.

Dickerson, P., Robins, B., & Dautenhahn, K. (2013). Where the action is: A conversation analytic perspective on interaction between a humanoid robot, a co-present adult and a child with an ASD. *Interaction Studies: Social Behaviour and Communication in Biological and Artificial Systems*, *14*(2), 297–316. http://doi.org/10.1075/is.14.2.07dic

Dinet & Vivian (2014). (n.d.).

Docherty, S., & Sandelowski, M. (1999). Focus on qualitative methods: Interviewing children. *Research in Nursing and Health*, *22*(2), 177–185.

Donker, A., & Reitsma, P. (2004). Usability Testing With Young Children. *Interaction Design and Children*, 43–48. http://doi.org/10.1145/1017833.1017839

Dooley, D. (2001). *Social Research Methods* (4th ed.). Englewood Cliffs, New Jersey: Prentice-Hall.

DREAM. (2015). DREAM: Development of Robot-Enhanced therapy for children with AutisM spectrum disorders. Retrieved December 18, 2015, from https://www.utwente.nl/ctit/research/research_projects/international/fp7-streps/dream-gw/

Duquette, A., Michaud, F., & Mercier, H. (2008). Exploring the use of a mobile robot as an imitation agent with children with low-functioning autism. *Autonomous Robots*, *24*, 147–157. http://doi.org/10.1007/s10514-007-9056-5

Einarsdóttir, J. (2007). Research with children : methodological and ethical challenges. *European Early Childhood Education Research Journal*, *15*(2), 197–211. http://doi.org/10.1080/13502930701321477

Encarnação, P., Alvarez, L., Rios, A., Maya, C., Adams, K., & Cook, A. (2014). Using virtual robot-mediated play activities to assess cognitive skills. *Disability and Rehabilitation. Assistive Technology*, *9*(3), 231–41. http://doi.org/10.3109/17483107.2013.782577

Feil-Seifer, D., & Matarić, M. J. (2009). Toward Socially Assistive Robotics for Augmenting Interventions for Children with Autism Spectrum Disorders. *Springer Tracts in Advanced Robotics*, *54*, 201–210. http://doi.org/10.1007/978-3-642-00196-3_24

Forlizzi, J., & Disalvo, C. (2006). Service Robots in the Domestic Environment: A Study of the Roomba Vacuum in the Home. *Design*, *2006*, 258–265. http://doi.org/10.1145/1121241.1121286

Fridin, M. (2014a). Kindergarten social assistive robot: First meeting and ethical issues. *Computers in Human Behavior*, *30*, 262–272. http://doi.org/10.1016/j.chb.2013.09.005

Fridin, M. (2014b). Storytelling by a kindergarten social assistive robot: A tool for constructive learning in preschool education. *Computers & Education*, *70*, 53–64. http://doi.org/10.1016/j.compedu.2013.07.043

Fridin, M., & Belokopytov, M. (2014a). Acceptance of socially assistive humanoid robot by preschool and elementary school teachers. *Computers in Human Behavior*, *33*, 23–31. http://doi.org/10.1016/j.chb.2013.12.016

Fridin, M., & Belokopytov, M. (2014b). Embodied Robot versus Virtual Agent: Involvement of Preschool Children in Motor Task Performance. *International Journal of Human-Computer Interaction*, *30*(November), 459–469. http://doi.org/10.1080/10447318.2014.888500

Ghosh, M., & Tanaka, F. (2011). The impact of different competence levels of care-receiving robot on children. *IEEE International Conference on Intelligent Robots and Systems*, 2409–2415. http://doi.org/10.1109/IROS.2011.6048743

Giannopulu, I., & Pradel, G. (2012). From child-robot interaction to child-robot-therapist interaction: A case study in autism. *Applied Bionics and Biomechanics*, *9*(2), 173–179. http://doi.org/10.3233/JAD-2011-0042

Gomes, P. F., Sardinha, A., Segura, E. M., Cramer, H., & Paiva, A. (2014). Migration Between Two Embodiments of an Artificial Pet. *International Journal of Humanoid*

Robotics, *11*(1), 1450001. http://doi.org/10.1142/S0219843614500017

Goris, K., Saldien, J., Vanderniepen, I., & Lefeber, D. (2009). The {Huggable} {Robot} {Probo}, a {Multi}-disciplinary {Research} {Platform}. *Research and {Education} in {Robotics} — {EUROBOT} 2008*, (33), 29–41. http://doi.org/10.1007/978-3-642-03558-6_4

Han, J., Jo, M., Hyun, E., & So, H. (2015). Examining young children's perception toward augmented reality-infused dramatic play. *Educational Technology Research and Development*, *63*(3), 455–474. http://doi.org/10.1007/s11423-015-9374-9

Han, J., Jo, M., Park, S., & Kim, S. (2005a). The educational use of home robots for children. In *ROMAN 2005. IEEE International Workshop on Robot and Human Interactive Communication, 2005.* (pp. 378–383). IEEE. http://doi.org/10.1109/ROMAN.2005.1513808

Han, J., Jo, M., Park, S., & Kim, S. (2005b). The educational use of home robots for children. In *ROMAN 2005. IEEE International Workshop on Robot and Human Interactive Communication, 2005.* (pp. 378–383). IEEE. Retrieved from http://ieeexplore.ieee.org/articleDetails.jsp?arnumber=1513808

Han, J.-H., Jo, M.-H., Jones, V., & Jo, J.-H. (2008). Comparative Study on the Educational Use of Home Robots for Children. *Journal of Information Processing Systems*, *4*(4), 159–168. http://doi.org/10.3745/JIPS.2008.4.4.159

Hanna, L., Risden, K., & Alexander, K. J. (1997). Methods & Tools. *Interactions*, (september +october), 9–14.

Hanson, D., Mazzei, D., Garver, C., Ahluwalia, A., De Rossi, D., Stevenson, M., & Reynolds, K. (2012). ASD, Social Training, and Research; Shown to Appeal to Youths with ASD, Cause Physiological Arousal, and Increase Human-to-Human Social Engagement. *Conference on Pervasive Technologies Related to Assistive*, 1–7. Retrieved from http://www.faceteam.it/wp-content/uploads/2012/07/PETRA-2012_Realistic-Humanlike-Robots-for-ASD.pdf

Hashimoto, T., Kobayashi, H., & Kato, N. (2011). Educational system with the android robot SAYA and field trial. *IEEE International Conference on Fuzzy Systems*, *8*(3), 766–771. http://doi.org/10.1109/FUZZY.2011.6007430

Herberg, J. S., Behera, D. C., & Saerbeck, M. (2013). Eliciting ideal tutor trait perception in robots Pinpointing effective robot design space elements for smooth tutor interactions. *2013 8th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, 137–138. http://doi.org/10.1109/HRI.2013.6483539

Hicks, C. L., Von Baeyer, C. L., Spafford, P. A., Van Korlaar, I., & Goodenough, B. (2001). The Faces Pain Scale – Revised: Toward a common metric in pediatric pain measurement. *Pain*, *93*, 173–183.

Hsiao, H.-S., Chang, C.-S., Lin, C.-Y., & Hsu, H.-L. (2012). iRobiQ": the influence of bidirectional interaction on kindergarteners' reading motivation, literacy, and behavior. *Interactive Learning Environments*, *4820*(February 2013), 1–24. http://doi.org/10.1080/10494820.2012.745435

Huskens, B., Palmen, A., Van der Werff, M., Lourens, T., & Barakova, E. (2014). Improving Collaborative Play Between Children with Autism Spectrum Disorders and Their Siblings: The Effectiveness of a Robot-Mediated Intervention Based on Lego® Therapy. *Journal of Autism and Developmental Disorders*, *45*(11), 3746–3755. http://doi.org/10.1007/s10803-014-2326-0

Hwang, W.-Y., & Wu, S.-Y. (2014). A case study of collaboration with multi-robots and its effect on children's interaction. *Interactive Learning Environments*, *22*(4), 429–443. http://doi.org/10.1080/10494820.2012.680968

Kahn, Jr., P. H., Friedman, B., Pérez-Granados, D. R., & Freier, N. G. (2006). Robotic pets in the lives of preschool children. *Interaction Studies*, *7*, 405–436. http://doi.org/10.1075/is.7.3.13kah

Kanda, T., & Ishiguro, H. (2006). An approach for a social robot to understand human relationships: Friendship estimation through interaction with robots. *Interaction Studies*, *7*(3), 369–403. http://doi.org/10.1075/is.7.3.12kan

Kassin, S., Fein, S., & Markus, H. R. (2008). *Social Psychology* (7th editio). Boston: Houghton Mifflin Company.

Kennedy, J., Baxter, P., & Belpaeme, T. (2013). Constraining content in mediated unstructured social interactions: Studies in the wild. *Proceedings - 2013 Humaine Association Conference on Affective Computing and Intelligent Interaction, ACII 2013*, (SEPTEMBER), 728–733. http://doi.org/10.1109/ACII.2013.135

Keren, G., Ben-David, a., & Fridin, M. (2012). Kindergarten assistive robotics (KAR) as a tool for spatial cognition development in pre-school education. *IEEE International Conference on Intelligent Robots and Systems*, 1084–1089. http://doi.org/10.1109/IROS.2012.6385645

Kerepesi, a., Kubinyi, E., Jonsson, G. K., Magnusson, M. S., & Miklósi, Á. (2006). Behavioural comparison of human–animal (dog) and human–robot (AIBO) interactions. *Behavioural Processes*, *73*(1), 92–99. http://doi.org/10.1016/j.beproc.2006.04.001

Kim, N., Han, J., & Ju, W. (2014). Is a robot better than video for initiating remote social connections among children? *Proceedings of the 2014 ACM/IEEE International Conference on Human-Robot Interaction - HRI '14*, 208–209. http://doi.org/10.1145/2559636.2563692

Kim, Y.-D., Hong, J.-W., Kang, W.-S., Baek, S.-S., Lee, H.-S., & An, J. (2010). Design of Robot Assisted Observation System for Therapy and Education of Children with Autism. *Social Robotics SE - 23*, *6414*, 222–231. http://doi.org/10.1007/978-3-642-17248-9_23

Ko, W. H., Lee, S. M., Nam, K. T., Shon, W. H., & Ji, S. H. (2010). Design of a personalized r-learning system for children. *2010 IEEE/RSJ International Conference on Intelligent Robots and Systems*, 3893–3898. http://doi.org/10.1109/IROS.2010.5649668

Kose-Bagci, H., Ferrari, E., Dautenhahn, K., Syrdal, D. S., & Nehaniv, C. L. (2009). Effects of Embodiment and Gestures on Social Interaction in Drumming Games with a Humanoid Robot. *Advanced Robotics*, *23*(14), 1951–1996. http://doi.org/10.1163/016918609X12518783330360

Kozima, H., Nakagawa, C., & Yasuda, Y. (2005). Designing and observing human-robot interactions for the study of social development and its disorders. *2005 International Symposium on Computational Intelligence in Robotics and Automation*, 41–46. http://doi.org/10.1109/CIRA.2005.1554252

Kronreif, G., Prazak, B., Mina, S., Kornfeld, M., Meindl, M., & Fürst, M. (2005). PlayROB - Robot-assisted playing for children with severe physical disabilities. *Proceedings of the 2005 IEEE 9th International Conference on Rehabilitation Robotics*, *2005*, 193–196. http://doi.org/10.1109/ICORR.2005.1501082

Lazar, J., Feng, J. H., & Hochheiser, H. (2010). *Research Methods in Human-Computer Interaction*. John Wiley & Sons. Retrieved from https://books.google.com/books?id=H_r6prUFpc4C&pgis=1

Leite, I., Castellano, G., Pereira, A., Martinho, C., & Paiva, A. (2014). Empathic Robots for Long-term Interaction. *International Journal of Social Robotics*, *6*(3), 329–341. http://doi.org/10.1007/s12369-014-0227-1

Leite, I., Henriques, R., Martinho, C., & Paiva, A. (2013). Sensors in the wild: Exploring electrodermal activity in child-robot interaction. *ACM/IEEE International Conference on Human-Robot Interaction*, 41–48. http://doi.org/10.1109/HRI.2013.6483500

Lombard, M., & Ditton, T. (1997). At the Heart of It All: The Concept of Presence. *Journal of Computer-Mediated Communication*, *3*(2). http://doi.org/10.1111/j.1083-6101.1997.tb00072.x

Lombard, M., Ditton, T. B., Crane, D., Davis, B., Gil-Egui, G., Horvath, K., & Rossman, J. (2000). Measuring presence: a literature-based approach to the development of a standardized paper-and-pencil instrument. Retrieved April 20, 2016, from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.132.4737&rep=rep1&type=pdf

Machado, A., & Lourenço, O. (1995). In Defense of Piaget ' s Theory : A Reply to 10 Common Criticisms, (DECEMBER). http://doi.org/10.1037/0033-295X.103.1.143

Mazzei, D., Lazzeri, N., Hanson, D., & De-Rossi, D. (2012). HEFES: An Hybrid Engine for Facial Expressions Synthesis to control human-like androids and avatars. *Biomedical Robotics and Biomechatronics (BioRob), 2012 4th IEEE RAS & EMBS International Conference on*, 195–200. http://doi.org/10.1109/BioRob.2012.6290687

Melson, G. F., Kahn, P. H., Beck, A., Friedman, B., Roberts, T., Garrett, E., & Gill, B. T. (2009). Children's behavior toward and understanding of robotic and living dogs. *Journal of Applied Developmental Psychology*, *30*(2), 92–102. http://doi.org/10.1016/j.appdev.2008.10.011

Mendelson, M. J., & Aboud, F. E. (1999). Measuring friendship quality in late adolescents and young adults: McGill friendship questionnaires. *Canadian Journal of Behavioural Science*, *31*(2), 130–132. http://doi.org/10.1037/h0087080

Mills, K. a, Chandra, V., & Park, J. Y. (2013). The architecture of children ' s use of language and tools when problem solving collaboratively with robotics, 315–337. http://doi.org/10.1007/s13384-013-0094-z

Mills, K. a, Chandra, V., & Yong, J. (2013). The architecture of children ' s use of language and tools when problem solving collaboratively with robotics, 315–337. http://doi.org/10.1007/s13384-013-0094-z

Mubin, O., Shahid, S., Van De Sande, E., Krahmer, E., Swerts, M., Bartneck, C., & Feijs, L. (2010). Using child-robot interaction to investigate the user acceptance of constrained and artificial languages. *Ro-Man, 2010 Ieee*, 588–593. http://doi.org/10.1109/ROMAN.2010.5598731

Murphy, R. R., & Schreckenghost, D. (2013). Survey of metrics for human-robot interaction. In *2013 8th ACM/IEEE International Conference on Human-Robot Interaction (HRI)* (pp. 197–198). IEEE. http://doi.org/10.1109/HRI.2013.6483569

Neerincx, M. A. (2010). Social Robots for Self-management of Health-promoting Activities. *CyberTherapy & Rehabilitation*, *3*(3), 38–39.

Oh, K., & Kim, M. (2010). Social attributes of robotic products: Observations of child-robot interactions in a school environment. *International Journal of Design*, *4*(1), 45–55. http://doi.org/G300-cX1077991.vn0p106

Okita, S. Y., Schwartz, D. L., Shibata, T., & Tokuda, H. (2005). Exploring young children's attributions through entertainment robots. *Proceedings - IEEE International Workshop on Robot and Human Interactive Communication*, *2005*, 390–395. http://doi.org/10.1109/ROMAN.2005.1513810

Piaget, J. (1960). *Psychology of Intelligence*. Littlefield, Adams & Co.

Pioggia, G., Sica, M. L., Ferro, M., Igliozzi, R., Muratori, F., Ahluwalia, a., & De Rossi, D. (2007). Human-robot interaction in autism: FACE, an android-based social therapy. *Proceedings - IEEE International Workshop on Robot and Human Interactive Communication*, 605–612. http://doi.org/10.1109/ROMAN.2007.4415156

Punch, S. (2002). Research with children: The same or different from research with adults? *Childhood*, *9*(3), 321–341. http://doi.org/10.1177/0907568202009003045

Read, J. C., & Fine, K. (2005). Using Survey Methods for Design and Evaluation in Child Computer Interaction. *Workshop on Child Computer Interaction: Methodological Research at Interact 2005*. Retrieved from https://www.mendeley.com/research/using-survey-methods-design-evaluation-child-computer-interaction-3/?utm_source=desktop&utm_medium=1.12.3&utm_campaign=open_catalog&userDocumentId=%7b1f6fab6c-6d1b-4035-a063-9364a428fd0f%7d\nhttp://citeseerx.ist.psu.edu/v

Read, J. C., & MacFarlane, S. (2006). Using the fun toolkit and other survey methods to gather opinions in child computer interaction. *Proceeding of the 2006 Conference on Interaction Design and Children IDC 06*, 81. http://doi.org/10.1145/1139073.1139096

Read, J., Macfarlane, S., & Casey, C. (2002). Endurability , Engagement and Expectations : Measuring Children ' s Fun. *Interaction Design and Children*, *2*, 1–23. http://doi.org/10.1.1.100.9319

Ribi, F. N., Yokoyama, A., & Turner, D. C. (2008). Comparison of Children ' s Behavior toward Sony ' s Robotic Dog AIBO and a Real Dog : A Pilot Study. *Anthrozoös*, *21*(3), 245–256. http://doi.org/10.2752/175303708X332053

Riek, L. (2012). Wizard of Oz Studies in HRI: A Systematic Review and New Reporting Guidelines. *Journal of Human-Robot Interaction*, *1*(1), 119–136. http://doi.org/10.5898/JHRI.1.1.Riek

Robins, B., Dickerson, P., & Dautenhahn, K. (2005). ROMAN 2005. IEEE International Workshop on Robot and Human Interactive Communication, 2005. *ROMAN 2005. IEEE International Workshop on Robot and Human Interactive Communication, 2005.*, 54–59. http://doi.org/10.1109/ROMAN.2005.1513756

Ros, R., Baroni, I., & Demiris, Y. (2014). Adaptive human–robot interaction in sensorimotor task instruction: From human to robot dance tutors. *Robotics and Autonomous Systems*, *62*(6), 707–720. http://doi.org/10.1016/j.robot.2014.03.005

Roy, N., Baltus, G., Fox, D., Gemperle, F., Goetz, J., Hirsch, T., … Thrun, S. (2000). Towards Personal Service Robots for the Elderly. *Workshop on Interactive Robots and Entertainment (WIRE 2000) (2000)*, *25*(2000), 184. http://doi.org/10.1007/s12369-014-0232-4

Salter, T., Werry, I., & Michaud, F. (2008). Going into the wild in child–robot interaction studies: issues in social robotic development. *Intelligent Service Robotics*, *1*(2), 93–108. Retrieved from http://link.springer.com/10.1007/s11370-007-0009-9

Sapounidis, T., Demetriadis, S., & Stamelos, I. (2014). Evaluating children performance with graphical and tangible robot programming tools. *Personal and Ubiquitous Computing*, *19*(1), 225–237. http://doi.org/10.1007/s00779-014-0774-3

Saylor, M. M., & Levin, D. T. (2005). Thinking and seeing in intentional and mechanical systems. *Robot and Human Interactive Communication, 2005. ROMAN 2005. IEEE International Workshop on*, 710–715. http://doi.org/10.1109/ROMAN.2005.1513863

Schiavone, G., Formica, D., Taffoni, F., Campolo, D., Guglielmelli, E., & Keller, F. (2011). Multimodal Ecological Technology: From Child's Social Behavior Assessment to Child-Robot Interaction Improvement. *International Journal of Social Robotics*, *3*(1), 69–81. http://doi.org/10.1007/s12369-010-0080-9

Scopelliti, M., Giuliani, M. V., & Fornara, F. (2005). Robots in a domestic setting: A psychological approach. *Universal Access in the Information Society*, *4*(2), 146–155. http://doi.org/10.1007/s10209-005-0118-1

Scott, J. (1997). Children as Respondents: Methods for Improving Data Quality. In L. Lyberg, P. Biemer, M. Collins, E. De Leeuw, C. Dippo, N. Schwarz, & D. Trewin (Eds.), *Survey Measurement and Process Quality* (pp. 331–350). John Wiley and Sons Ltd.

Shahid, S., Krahmer, E., & Swerts, M. (2014). Child–robot interaction across cultures: How does playing a game with a social robot compare to playing a game alone or with a friend? *Computers in Human Behavior*, *40*, 86–100. http://doi.org/10.1016/j.chb.2014.07.043

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. *Proceedings of IEEE Workshop on Advanced Robotics and Its Social Impacts, ARSO*, 175–180. http://doi.org/10.1109/ARSO.2013.6705525

Shields, B. J., Palermo, T. M., Powers, J. D., Grewe, S. D., & Smith, G. a. (2003). Predictors of a child ' s ability to use a visual analogue scale, 281–290.

Sidner, C. L., Kidd, C. D., Lee, C., & Lesh, N. (2003). Where to Look : A Study of Human-Robot Engagement. *Proceedings of Intelligent User Interfaces*, 78–84. http://doi.org/10.1145/964442.964458

Simo, A., Nishida, Y., & Nagashima, K. (2006). A Humanoid Robot to Prevent Children Accidents. *Interactive Technologies and Sociotechnical Systems*, 476–485.

Sipitakiat, A., & Nusen, N. (2012). Robo-Blocks: Designing Debugging Abilities in a Tangible Programming System for Early Primary School Children. *Proceedings of the 11th International Conference on Interaction Design and Children (IDC 2012)*, 98–105. http://doi.org/10.1145/2307096.2307108

Soares, F., Costa, S., Silva, S., Goncalves, N., Rodrigues, J., Santos, C., … Moreira, M. F. (2013). Rob??tica-Autismo project: Technology for autistic children. *3rd Portuguese Bioengineering Meeting, ENBENG 2013 - Book of Proceedings*. http://doi.org/10.1109/ENBENG.2013.6518415

Steinfeld, A., Fong, T., Kaber, D., Lewis, M., Scholtz, J., Schultz, A., & Goodrich, M. (2006). Common metrics for human-robot interaction. In *Proceeding of the 1st ACM SIGCHI/SIGART conference on Human-robot interaction - HRI '06* (p. 33). New York, New York, USA: ACM Press. http://doi.org/10.1145/1121241.1121249

Steward, M. S., Steward, D. S., Farquhar, L., Myers, J. E. B., Reinhart, M., Welker, J., … Morgan, J. (1996). Interviewing young children about body touch and handling. *Monographs of the Society for Research in Child Development*, *61*(4-5). Retrieved from http://discovery.ucl.ac.uk/134227/

Sung, J., Guo, L., Grinter, R. E., & Christensen, H. I. (2007). “My Roomba Is Rambo”: Intimate Home Appliances. *Ubiquitous Computing*, 145–162. http://doi.org/10.1007/978-3-540-74853-3_9

Tung, F. (2011). Influence of Gender and Age on the Attitudes of Children, 637–646.

Turkle, S., Taggart, W., Kidd, C. D., & Dasté, O. (2006). Relational artifacts with children and elders: the complexities of cybercompanionship. *Connection Science*. http://doi.org/10.1080/09540090600868912

Veenstra, D. N., & Evers, V. (2011). The Development of an Online Research Tool to Investigate Children ’ s Social Bonds with Robots. *Engineering*, 19–26.

Wainer, J., Dautenhahn, K., Robins, B., & Amirabdollahian, F. (2014). A Pilot Study with a Novel Setup for Collaborative Play of the Humanoid Robot KASPAR with Children with Autism. *International Journal of Social Robotics*, *6*(1), 45–65. http://doi.org/10.1007/s12369-013-0195-x

Wainer, J., Ferrari, E., Dautenhahn, K., & Robins, B. (2010). The effectiveness of using a robotics class to foster collaboration among groups of children with autism in an exploratory study. *Personal and Ubiquitous Computing*, *14*(5), 445–455. http://doi.org/10.1007/s00779-009-0266-z

Wainer, J., Robins, B., Amirabdollahian, F., & Dautenhahn, K. (2014). Using the Humanoid Robot KASPAR to Autonomously Play Triadic Games and Facilitate Collaborative Play Among Children With Autism. *IEEE Transactions on Autonomous Mental Development*, *6*(3), 183–199. http://doi.org/10.1109/TAMD.2014.2303116

Watson-Gegeo, K. A. (1988). Ethnography in ESL: Defining the Essentials. *TESOL Quarterly*, *22*(4), 5755–92. http://doi.org/10.2307/3587257

Wood, L. J., Dautenhahn, K., Lehmann, H., Robins, B., Rainer, A., & Syrdal, D. S. (2013). Robot-mediated interviews: Do robots possess advantages over human interviewers when talking to children with special needs? *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *8239 LNAI*, 54–63. http://doi.org/10.1007/978-3-319-02675-6_6

Wood, L. J., Dautenhahn, K., Rainer, A., Robins, B., Lehmann, H., & Syrdal, D. S. (2013). Robot-mediated interviews--how effective is a humanoid robot as a tool for interviewing young children? *PloS One*, *8*(3), e59448. http://doi.org/10.1371/journal.pone.0059448

Wyeth, P. & Wyeth, G. (2008). Robot building for preschoolers. *RoboCup 2007: Robot Soccer World Cup XI*, 124–135. http://doi.org/10.1007/978-3-540-68847-1_11

Yamazaki, R., Nishio, S., Ogawa, K., Matsumura, K., Minato, T., Ishiguro, H., … Nishikawa, M. (2013). Promoting Socialization of Schoolchildren Using a Teleoperated Android: an Interaction Study. *International Journal of Humanoid Robotics*, *10*(01), 1350007. http://doi.org/10.1142/S0219843613500072

Yanco, H. A., & Drury, J. L. (2002). A Taxonomy for Human-Robot Interaction. *AAAI Technical Report FS-02-03*.

Yanco, H. A., & Dury, J. (2004). Classifying Human-Robot Interaction: An Updated

Taxonomy.