Evaluation Methods for User-Centered Child-Robot Interaction

Vicky Charisi[1] Daniel Davison[1] Dennis Reidsma[1] and Vanessa Evers[1]

***Abstract*— This review examines recent methodological approaches for the evaluation of child-robot interaction in learning settings. The main aims are to map existing work from a user-centered perspective, to identify possible trends related to evaluation methods for child-robot interaction, and to discuss potential future directions. We present a systematic review of existing studies, which have been thematically organized based on their research objectives. We then examine the evaluation methods that were used in these studies and we propose a conceptual framework based on the one hand on the themes that emerged, namely the social interaction between the child and the robot, the social acceptance, possible emotional interactions, the learning process and the learning outcome, and on the other hand on the corresponding measures. These methods have been considered in relation with the age ranges of the children, because of the relationship of their cognitive level to the choice of a developmentally appropriate evaluation method. We use this framework to highlight current trends and needs for the field and to contextualize the methodological directions for child-robot interaction. Finally, we discuss the challenges and limitations of the current methodological approaches as well as possible future directions for the evaluation methods of child-robot interaction in learning settings.**

I. Introduction

This review investigates the current approaches of evaluation methods in the field of child-robot interaction (CRI) in learning settings. Our main aim is to identify trends, challenges and needs in existing studies regarding their objectives and the corresponding methods. Our main motivation for this review relates to one of the aims of our current project - Expressive Agents for Symbiotic Education and Learning (EASEL) - which includes the evaluation of robots' behaviour in order to optimize the CRI in learning settings. During a series of studies we tried to evaluate the extent to which the EASEL robot is able to initiate, sustain and support a developmentally appropriate CRI, in which the robot has the role of a learning companion. Despite the well-established methods for evaluation of child development from a psychological perspective e.g. [1] as well the growing maturity of the evaluation methods for child-computer interaction [2], there are several challenges that invite researchers, who work on user-centered approaches of CRI, to think of novel methods. One of the main reasons relates to the synthetic nature of CRI, which seems to be underpinned by different norms than those of child-child and child-computer interaction. The questions that we will attempt to answer with this review are the following: What are the most prevalent objectives of the existing studies on CRI? What are the evaluation methods that are used to address these objectives? Which age-group within childhood is mostly investigated? How does the objectives of the study and the target age group of children affect the choice of the evaluation methods for CRI? Finally, based on the literature, we suggest a conceptual framework for the existing evaluation methods, highlighting the current challenges and possible future directions.

[1]Vicky Charisi, Daniel Davison, Dennis Reidsma and Vanessa Evers are with the Faculty of Electrical Engineering, Mathematics and Computer Science, Human Media Interaction Group, University of Twente, 7500 AE Enschede, The Netherlands {v.charisi, d.p.davison, d.reidsma, v.evers}@utwente.nl

II. Methods

For the reliability of the review process, we used the PRISMA flow diagram for Systematic Review and Meta-analysis [3]. We conducted an electronic database search of studies reported in the English language, covering the fields of robotics, computer science, psychology and education by using the following databases: Scopus, Association for Computing Machinery (ACM), American Psychological Association (APA) and Education Resources Information Center (ERIC). We used these specific databases in order to cover the multiple disciplines that engage in the field of CRI from a user-centered perspective, being aware that these databases represent only a portion of the existing relevant studies. We searched for papers with the terms "child" and "robot" in the body of the title, the abstract and the keywords. We excluded "autism", "ASD" and "disorder", since the focus of this study lies on robot interaction with typically developing children only. In total 376 papers were found. After the removal of 90 duplicates, we had 287 studies for screening. Duplicates refer to both the appearance of the same study in two or more different databases as well as the papers that present the results of a study that has also been reported elsewhere (e.g. as a journal paper) reporting the study in more detail. We decided to include studies for the period from the 1st of January 2000 until the 1st of December 2015, which resulted in the inclusion of 279 studies for further screening according to the criteria that are presented in the following section.

A. Screening and Eligibility Criteria

All the titles and abstracts were screened in order for us to identify and remove the studies with the following exclusion criteria. Initially, we excluded studies that were relevant to non-typically developing children or children with some form of disability or health problems. The reason for this exclusion criterion lies in the fact that special features of the target group may need methodological approaches, the examination of which is beyond the scope of this paper. Secondly, we excluded the robotics-centered studies, which were aimed either at technical advances or at the development of robots' cognitive architectures. These type of studies are conducted with the aim of experimentation of the robotic system per se, rather than with a focus on the interaction between the child and the robot. However, it is assumed that any study that investigates the interaction between humans and robots may involve some robotics component. In the case of the studies with a dual aim (child-centered and robot-centered), we decided to include the papers, if they fulfilled the rest of the criteria as discussed in the following section. After the removal of 40 studies according to the exclusion criteria, 239 studies remained to be assessed for eligibility to be coded. The eligibility of the studies was dependent on the following criteria:

- the paper includes at least one user study with children. In this way, we excluded theoretical and review studies;
- the paper clearly identifies the age-range of the participants; and
- the paper reports the results obtained by the study.

After the screening of the papers according to the above-mentioned criteria, we excluded 144 studies, which resulted in the identification of 135 papers as eligible for inclusion in this review. Although that the analysis we present in this review is based on all the 135 papers, due to space limitations, in this paper we cite only some representative papers. The full set of the reviewed papers is available upon request.

B. Variable Definitions

In order for us to organize the review of these papers we annotated them according to 3 variables: (i) the objectives of the study; (ii) the corresponding methods; (iii) the age range of the participant children.

Regarding the objective of the studies, we coded each study in reference to the five umbrella themes, which were derived from the initial bottom up annotation as follows: (i) social interaction between the child and the robot, (ii) the social acceptance, (iii) possible emotional interaction, (iv) the learning process, and (v) the learning outcome. As far as the methodological approaches are concerned, we coded the type of the measurement and the study design in accordance to the following themes: (i) self-assessment methods, (ii) behavioural observations, (iii) psycho-physiological metrics, and (iv) task performance measurements. We coded the research design either as a longitudinal or as single-session study - which was usually called an experimental study. In addition, we annotated the age range of the participant children. In the field of developmental psychology, the most well-established classification of children's developmental stages is the one that was suggested by Jean Piaget [4]; despite some limitations [5], we decided to use this classification as a starting point to code children's age-groups, namely: (i) sensory-motor stage (birth - 2yrs), (ii) pre-operational stage (3yrs - 5yrs), (iii) concrete operational stage (6yrs - 11yrs), and (iv) formal operational stage (12yrs - 16yrs). The combination of these codes was the basis of the conceptual framework for the evaluation methods in CRI that we suggest with this paper. Although we analysed all of the 135 papers, due to space limitations, in this review we report only representative examples of papers for each of the above mentioned codes. Based on these codes, we will describe the development of our conceptual framework for the existing methodological approaches that are used for the evaluation of CRI.

III. Examples from the Literature

As a starting point for the organization of this overview we used the developmental paradigm of socio-cultural theory [6], according to which, children develop better when learning occurs in a social context. During the social interaction, the learner actively contributes to the construction of their knowledge [7]. As a result, the social interactions may influence the learning gains. In this case, children's learning companions include physical humanoid or non-humanoid robots.

Attributing social competence to inanimate entities is a well-investigated phenomenon that appears in young children [8]. The evaluation of children's perspectives of robot's socialness has been mainly investigated with observational methods, semi-structured interviews and questionnaires, which allow for an understanding of children's criteria of social acceptance [9]. A child's perceptions and interpretation of a robot's social behaviour relates to the repertoire of its social acts and its verbal and non-verbal cues. Multimodality creates the opportunities for richer and more effective interactions, however they are more complex to design and evaluate [10]. One of the commonly used method for the evaluation of CRI is the systematic observation of children's behaviours focusing on low-level or high-level behaviours [11], [12].

However, children's *social acceptance* seems to be culture-specific. Intercultural studies that include self-report scores, perception tests and behavioural analyses have shown that children with different cultural backgrounds reacted differently towards robots [13]. Moreover, children appear to be biased by the novelty effect when they interact with a robot for the first time. However, in long-term interactions, social acceptance and effective interactions may be established, when the novelty effect diminishes [14]. Social acceptance in long-term interactions has been evaluated with the use of self-report measurement methods in combination with observational methods as well as with objective methods of automated data collection.

As far as the selection of the methodologies for self-assessment is concerned, one of the most common approaches is the use of questionnaires. However, in the case of young children several challenges arise. A variety of methods, such as pictorial tasks, can be used to address these challenges. Harter and Pike [15], for example, proposed a validated scale of perceived competence and social acceptance for young children. One of the indications of the effectiveness of this approach is that the combination of the pictorial task along with semi-structured interviews help children to use the visual stimulus for their better understanding of the questions. Additionally, while the children tended to polarize their responses trying to give socially desirable answers, by giving a second chance to the child to make a selection of a picture, they avoid polarization (ibid.).

Child-robot *social interaction* often leads to a child's emotional response and eventually to a possible emotional bonding with the robot [16]. In such a case, emotion recognition is associated with the child's ability to distinguish the various emotional expressions in facial, gestural, vocal and verbal displays in robots and to understand their social-contextual meaning and its verbal and vocal expression [17]. Beck et al. [18] measured children's perceptions of robot's emotional states by systematically altering its head positions of six emotional key poses. They used questionnaires to ask children and adults to ascribe the robot's posture to a certain emotional state. For emotional long-term interaction, models of adaptive emotion expression have been developed [19] with the use of video analysis and questionnaires to find that children reacted more expressively to a robot, when it exhibits emotional expressions in an adaptive way.

Long-term repeated studies have used standardized validated existing tools to evaluate the robot's effects on *learning process*. Kanda et al. [20], for example, used continuous automated methods to investigate the process of children's adaptation to robots. Long-term studies have investigated children's learning processes when a robot is used as a tool for children to learn specific skills by using qualitative approaches such as design-based approaches [21]. Other methods for the evaluation of the learning process include observational methods of children's task-related activity over time [20], automatically generated data and measurement of the length of utterances and qualitative categorization of the content of children's utterances in order to investigate the gradual change of children's behaviour [22].

Summative evaluation methods have been widely used to assess children's *learning outcomes*, as a product of the learning process. To give a conventional distinction between the learning process and the learning product, we have conceptualized the code of learning process as the code of how, and the learning product as a code of what. Tests before and after the intervention, as well as task performance during the intervention, reveal the effectiveness of the interactions in the learning task [23], [24], [25], [26]. While summative evaluation methods are objective methods, which allow for generalizations, there are limitations in investigating clear relations and causalities when the intervention happens in a real school setting, where multiple factors may affect the result.

In addition to the above-mentioned features of the studies, the degree of autonomy of the robot, the wildness of the interaction environment and the target group under investigation seem to have a catalytic role on the selection of the evaluation methods. In CRI, target groups may include children within a wide age range, which require developmentally appropriate methods. The robot's social characteristics and its role in the task may also affect the choice of the evaluation metrics. Examples include a tutor, who gives hints or instructions for the child to follow [24]; a peer or co-learner, which adopts empathetic behaviours towards the child [27]; a mediator, who facilitates the social interaction between the children [28]; or a tool for children to acquire specific learning goals such as programming skills [21].

This literature review of representative studies highlighted the interrelation of the objectives that each study aims to address with the evaluation methods that are used.

IV. The Conceptual Framework

A systematic coding of the whole set of eligible studies was conducted which resulted in the following conceptual framework.

A. A Classification of the Objectives in Learning Settings

The examination of the research goals of the 135 eligible studies that we included in this review, led us to their classification into five interconnected thematic constructs, as follows:

Social interaction; describes the short-term and long-term interaction between the child and the robot, such as verbal and non-verbal communication. Despite that every interaction between the child and the robot could be characterized as social, this variable refers to the specific research goal of social features within CRI.

Social acceptance; includes children's attitudes and perspectives about the socialness of the robot and indicates their willingness to accept it as a social agent. This variable builds upon the previously described social interaction between the child and the robot. However, in the case of social acceptance, the child ascribes social characteristics to the robot that trigger the intention for interaction. This variable may encompass novel behaviours exhibited by the child, which do not appear in child-child interaction or child-computer interaction.

Emotional interaction; refers to long-term interaction that may trigger children's emotional engagement and eventual bonding with the robot. It includes the investigation of the prerequisite emotional statements that are needed in order for the emotional engagement and interaction to happen, such as the establishment of trust.

Learning process; is the learning trajectory that is shaped by a learner when he or she interacts with the learning companion - in this case the robot - in the context of a specific learning task either in one session or longitudinally.

Learning outcome; is defined as the fulfillment of a specific learning goal, in the form of a cognitive achievement or as a desired skill or competence. Research in developmental psychology has identified developmental stages, which indicate the the cognitive competence of children according to their age or environmental characteristics.

B. A Classification of the Methods

In this section, we take the above-mentioned studies as a starting point to expand the discussion about the evaluation methods that are commonly used in this field. When it is necessary, we provide additional input from the field of child-child interaction. A large body of quantitative methods seek to investigate correlations or causations in child-robot interaction, while qualitative ethnographic-inspired methods seek for a deep understanding of phenomena under investigation in naturalistic settings. Four main categories of methods that are commonly used for the evaluation of CRI are as follows:

Self-assessment methods have been used to investigate participants' perceptions of the phenomenon under investigation. However, children's abilities for reflecting on themselves and their capabilities of accurate verbal expression of their thoughts are under development, which raises several challenges. Attempts to overcome these difficulties include indirect approaches, such as the use of pictorial tasks. Harter and Pike [15] built the pictorial scale of perceived competence and social acceptance for young children, which entails a detailed example for validated construction of developmentally appropriate tools for children's self-assessment. In the field of child-robot interaction, pictorial tasks have been used as a starting point for semi-structured interviews, using stimuli and settings that are meaningful for children [29]. Measures of conversational engagement, the frequency of the use of specific words as well as content analysis of their responses are the main methods for analysis of semi-structured interviews. Additionally, questionnaires that are already validated from human-human interaction have been slightly modified in order to be appropriate for children. Alves-Olivera et al [30] for example, were inspired by a validated scale to build a questionnaire to measure children's expectations and satisfaction for their interaction with robots. Finally, while the use of large-scale surveys seems to inform research in human-robot interaction with adults, it does not seem to be common among the methods of self-assessment for children. Borgers et al. [31] discuss the influence of cognitive development on response quality, which requires appropriate questions for children. Okita and Schwartz [32] addressed this challenge by asking the children only simple questions. However, oversimplification of the questions limits the scope and the goals of the research. Despite the challenges in self-assessment methods with children and the ways that they evaluate their interaction with a robot, it has been shown that children can provide reliable responses if questioned in a manner that they can understand and about events that are meaningful to them [33].

Behavioural observations have been used in order to gain understandings from the child's behavioural cues during her interaction with the robot. Ethnographic approaches, for example, use behavioural observations to discover high-level patterns and regularities in real settings [34]. Exploratory and descriptive methods, such as design-based research, have been used for the identification of newly researched fields [21]. Most of these studies use behavioural observations to assess CRI in real-life settings with the analysis of high-level behaviours; however they have often been critisized by scholars who follow different epistemological paradigms. Behavioural observations have also been used in microgenetic and experimental studies [35]. In these cases, the researchers analyse low-level behaviours, such as touch, gaze, speech, gesture, in order to achieve more objective metrics.

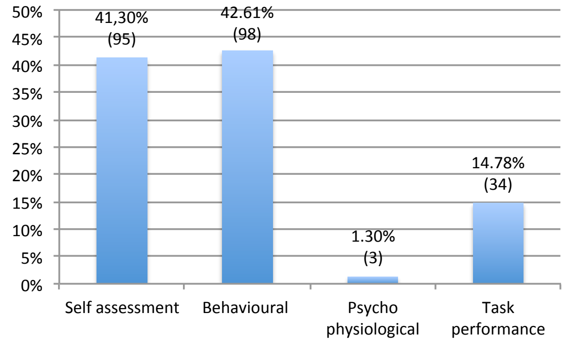

Fig. 1. The division of methods used.

Psycho-physiological metrics for the evaluation of CRI include automated measures that seek to detect or identify a specific human physiological statements or signals. The evaluation of participants' responses to a robot occurs through these signals, which are used for the development, implementation of real-time control and modification of robot behaviours, such as social signal processing (for a review [36]) albeit the challenges that emerge in measuring the quality of the human-robot interaction [37]. Brain-based methods, sociometrics and psychometrics have been used as objective tools to measure psychological statements. Examples include the Emotional Understanding Score (EUS), which measure a participant's ability to correctly recognize and label characters' emotional states [38].

Lastly, **task performance** is used to evaluate the effectiveness of interventions with robots. More specifically, pretests and posttests as well as standardized tests or time spent on a specific activity are examples of the objective measurements of participants' task performance [34].

V. Results

Based on the above mentioned framework, results from the review of 135 studies reveal that the research design of the 81.5% (110 out of 135) of the studies used one-session study, while 18.5% (25 out of 135) was conducted as a longitudinal study. More than half of the longitudinal studies (16 out of 25) were conducted after the year 2011, with a peak at the year 2014. The objectives of the studies were annotated according to the suggested framework. Some of the studies had more than one objective, which were annotated separately, resulting in 162 cases. Out of these 162 cases, 54 (33.3%) focused on the examination of children's social interaction with the robot, 36 (22.2%) on the ways that children accept (or not) a robot as a social agent, 26 (16.1%) on children's emotional involvement with a robot, 12 (21%) included as one of their objectives the changes in learning outcomes after an intervention with a robot, and 34 out of 162 (7.4%) focused on the investigation of children's learning process.

In total 241 measurements were used in the 135 included papers. From this number of measurements, we excluded 11

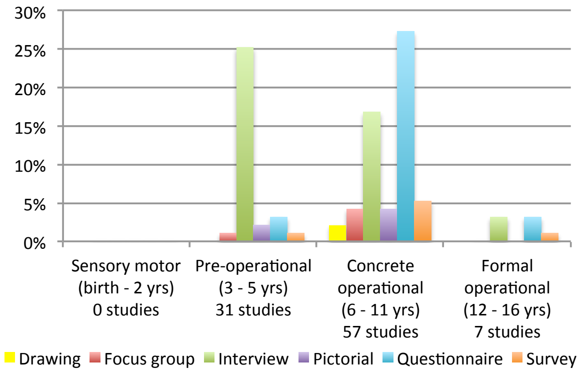

Fig. 2. An overview of self-assessment methods used for various stages of cognitive development.

measurements because they focused on the perceptions of adults (parents, teachers, experts). Out of 230 measurements, 95 (41.30%) were self-assessments, 98 (42.61%) were behavioural observations, 3 (1.30%) studies focused on psychophysiological metrics and 34 (15.78%) on task-performance (fig. 1). It is noticeable that the self-assessments included six different types of measurements (fig. 2), namely 32 questionnaires (33.68%), 43 interviews (45.26%), 5 focus groups (5.26%), 7 surveys (7.37%), 2 drawing tasks (2.11%), and 6 pictorial tasks (6.32%). For the purposes of this review, we made a distinction between the questionnaires and the pictorial tasks and drawings based on the modality that was used to trigger an answer from a child (e.g. written text versus visual representations or child's creations).

The studies that were included in these papers focused on users from a variety of age groups within childhood according to their cognitive stage. In some cases, the same study investigated children from two different age-groups, which were annotated separately, resulting in 151 cases. Out of these 151 cases, 4 (5.88%) studies focused on children of sensory-motor stage, 57(35.29%) studies on the pre-operational stage, 77 studies on the concrete operational stage (50.33%) and 13 (8.50%)studies on the formal operational stage. A further result of interest is related to the choice of the method in relation with the cognitive level of the children. In the case of the use of self-assessment methods, for example, scholars have tried a variety of methods depending on the participant age group (see fig. 2). More specifically, interviews were mainly used with children aged 3-5yrs, while questionnaires with children aged 6-11yrs.

In the following section we discuss the conclusions of this review and suggest some possible future directions.

VI. Conclusion and Future Directions

This paper presents a systematic review of studies on CRI with a focus on the evaluation methods. Because of the idiosyncratic nature of child-robot interaction, it has been previously noted that there is a need for novel methods of evaluation [39]. This review reveals a plurality in methodological approaches for user-centered CRI. However, the discussion is still open on the expansion of these methods taking into account the idiosyncratic nature of CRI.

More specifically, the objectives across studies focus on children's perceptions of robots' social features and the ways that these social features affected the interaction with the robot as well as the learning process and the learning outcomes. However, it was observed that there is a substantial lack of studies regarding long term CRI and the examination of the learning process with a robotic learning companion. This gap in the literature could possibly relate to: (i) the practical challenges in researching in the classroom for long term, (ii) the technological readiness of the robotic systems, and (iii) to the development of novel methodological approaches that allow the mapping of the process in such an environment. This requires learning tasks that are clear and compact enough for using them in short episodes of learning, but at the same time adaptable enough to allow repeated sessions over a longer period able to keep the interest of the children. Additionally, while there have been successful attempts with novel ways of evaluation, validation of these methods is needed. A number of studies have tried new methodological approaches, which have a lot of potential, such as the pictorial tasks and the human-likeness thermometer. However, their validation and their use across cases would be catalytic for the advance of the field of CRI, since cross-studies examinations would be possible. Furthermore, this review revealed an imbalance between the use of experimental well-controlled studies and naturalistic ones. Despite the increasing number of studies in naturalistic settings, often the setting tend to be artificial. The experimental setup can strongly influence the results of an experiment. The results of this review show that CRI research uses laboratory settings rather than more naturalistic user environments. This requires modification of the tasks in a way that the setup can function more independently as a natural element.

Existing studies have already yielded methodological insights of the interplay between children, robots and learning. The proposed framework was mainly used for the organization of this review; potentially, it can be used as a common framework for the methodological integration of the multi-disciplinary research and possible cross-studies validation. These steps may support the development of more advanced methodological techniques that are developmentally and ethically appropriate for children in different age ranges and offer a basis for future methodological directions.

Acknowledgment

This project has received funding from the European Union Seventh Framework Programme (FP7-ICT-2013-10) as part of EASEL under grant agreement n° 611971.

References

[1] P. H. Mussen and A. L. Baldwin, *Handbook of research methods in child development*. Wiley New York, 1960.

[2] P. Markopoulos, J. Read, J. Hoÿsniemi, and S. MacFarlane, "Child computer interaction: advances in methodological research," *Cognition, Technology & Work*, vol. 10, no. 2, pp. 79–81, 2007.

[3] D. Moher, A. Liberati, J. Tetzlaff, and D. G. Altman, "Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA StatementThe PRISMA Statement," *Annals of Internal Medicine*, vol. 151, no. 4, pp. 264–269, 2009.

[4] J. Piaget, "The theory of stages in cognitive development," in *Measurement and Piaget*, D. R. Green, Ed. New York: McGraw-Hill, 1971, pp. 1–11.

[5] V. Charisi, D. Davison, F. Wijnen, J. van der Meij, D. Reidsma, T. Prescott, W. van Joolingen, and V. Evers, "Towards a child-robot symbiotic co-development: a Theoretical approach," in *Fourth International Symposium on New Frontiers in Human-Robot Interaction*. Canterbury, UK: AISB Convention 2015, 2015.

[6] L. S. Vygotsky, *Mind in society: The development of higher psychological processes*. Harvard university press, 1930-34/1978.

[7] E. Ackermann, "Constructing knowledge and transforming the world," *A learning zone of one's own: Sharing representations and flow in collaborative learning environments*, vol. 1, pp. 15–37, 2004.

[8] J. Piaget, "The child's conception of the worldharcourt & brace," *New York*, 1929.

[9] D. Cameron, S. Fernando, A. Millings, B. R. Moore, A. Sharkey, and T. Prescott, "Children's age influences their perceptions of a humanoid robot as being like a person or machine," in *Living Machines 2015*, S. P. Wilson, P. Verschure, A. Mura, and T. Prescott, Eds. Barcelona, Spain: Springer, 2015, pp. 348–353.

[10] T. Belpaeme, P. Baxter, R. Read, R. Wood, H. Cuaya huitl, B. Kiefer, S. Racioppa, I. Kruijff-Korbayova, G. Athanasopoulos, V. Enescu, R. Looije, M. Neerinex, Y. Demiris, R. Ros-Espinoza, A. Beck, L. Can amero, A. Hiolle, M. Lewis, I. Baroni, M. Nalin, P. Cosi, G. Paci, F. Tesser, G. Sommavilla, and R. Humbert, "Multimodal Child-Robot Interaction: Buiding Social Bonds," *Journal of Human-Robot Interaction*, vol. 1, no. 2, pp. 33–53, 2012.

[11] M. Heerink, M. Diaz, J. Albo-Canals, C. Angulo, A. Barco, J. Casacuberta, and C. Garriga, "A field study on perception of social presence and interactive behavior with a pet-robot with primary school children," in *Ro-man 2012*, Paris, France, 2012.

[12] F. Wijnen, V. Charisi, D. Davison, J. van der Meij, D. Reidsma, and V. Evers, "Inquiry learning with a social robot: can you explain that to me?" in *New Friends 2015: The 1st international conference on social robots in therapy and education*, M. Heerink and M. de Jong, Eds. Almere, The Netherlands: Windesheim Flevoland, 2015, pp. 24–25.

[13] S. Shahid, E. Krahmer, and M. Swerts, "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*, vol. 40, pp. 86–100, 2014.

[14] F. Tanaka, J. R. Movellan, B. Fortenberry, and K. Aisaka, "Daily HRI evaluation at a classroom environment: reports from dance interaction experiments," in *International Conference of Human-Robot Interaction HRI '06*. Salt Lake City, Utah, USA: ACM Press, 2006.

[15] S. Harter and R. Pike, "The pictorial scale of perceived competence and social acceptance for young children," *Child Development*, vol. 55, no. 6, pp. 1969–1982, 1984.

[16] M. Fior, A. Ramirez-Serrano, T. Beran, S. Nugent, and R. Kuzyk, "Childrens relationships with robots: robot is childs new friend," *Journal of physical agents*, vol. 4, no. 3, pp. 9–17, 2010.

[17] D. Cameron, S. Fernando, E. Collins, A. Millings, B. R. Moore, A. Sharkey, V. Evers, and T. Prescott, "Presence of life-like robot expressions influences children's enjoyment of human-robot interactions in the field," in *4th International Symposium on New Frontiers in Human-Robot Interaction, AISB Convention 2015*, M. Salem, A. Weiss, P. Baxter, and K. Dautenhahn, Eds., Canterbury, UK, 2015.

[18] A. Beck, L. Canamero, G. Damiano, G. Sommavilla, F. Tesser, and P. Cosi, "Children interpretation of emotional body language displayed by a robot," in *International Conference of Social Robotics*. Amsterdam, the Netherlands: 2011, 2011.

[19] M. Tielman, M. Neerincx, J.-J. C. Meyer, and R. Looije, "Adaptive emotional expression in robot-child interaction," in *Conference of Human Robot Interaction, HRI '14*. Bielefeld, Germany: ACM, 2014, pp. 407–4014.

[20] T. Kanda, S. Nishio, H. Ishiguro, and N. Hagita, "Interactive humanoid robots and androids in children's lives," *Children, Youth and Environments*, vol. 19, no. 1, pp. 12–33, 2009.

[21] A. Sullivan and M. U. Umaschi Bers, "Robotics in the early childhood classroom: learning outcomes from an 8-week robotics curriculum in pre-kindergarten through second grade," *International Journal of Technology and Design Education*, 2015.

[22] E. Short, K. Swift-Spong, J. Greczek, A. Ramachandran, A. Litoiu, E. C. Grigore, D. Feil-Seifer, S. Shuster, J. J. Lee, S. Huang, S. Levonisova, S. Litz, J. Li, G. Ragusa, D. Spruijt-Metz, M. Mataric, and B. Scassellati, "How to train your DragonBot: Socially assistive robots for teaching children about nutrition through play," in *Proceedings - IEEE International Workshop on Robot and Human Interactive Communication*, vol. 2014-Octob, 2014, pp. 924–929.

[23] M. Blancas, V. Vouloutsi, K. Grechuta, and P. Verschure, "Effects of the robot's role on human-robot interaction in an educational scenario," in *Living Machines 2015*, S. P. Wilson, P. Verschure, A. Mura, and T. J. Prescott, Eds. Barcelona, Spain: Springer, 2015, pp. 391–402.

[24] V. Vouloutsi, M. B. Munoz, K. Grechuta, S. Lallee, A. Duff, J. Ysard Llobet Puigbo, and P. F. M. J. Verschure, "A new biomimetic approach towards educational robotics: the Distributed Adaptive Control of a Synthetic Tutor Assistant," in *4th International Symposium on New Frontiers in human-Robot Interaction*, 2015.

[25] J. Holmes, "Designing agents to support learning by explaining," *Computers and Education*, vol. 48, pp. 523–547, 2007.

[26] T. Kanda, T. Hirano, D. Eaton, and H. Ishiguro, "Interactive robots as social partners and peer tutors for children: a field trial," *Human-Computer Interaction*, vol. 19, no. 1, pp. 61–84, 2004.

[27] I. Leite, G. Castellano, A. Pereira, C. Martinho, and A. Paiva, "Empathic robots for long-term interaction: evaluating social presence, engagement and perceived support in children," *International Journal of Social Robotics*, vol. 6, pp. 329–341, 2014.

[28] T. Mochizuki, Y. Mitate, Y. Tateno, T. Wakimoto, Y. Miyata, J. Nakahara, and N. Miyake, "Robot as a Learning Partner for Promoting Proactive Discussion in peer Groups: a Case Study for Career Development," in *International Conference on Computers in Education*, L.-H. Wong, Ed. Asia-Pacific Society for Computers in Education, 2013.

[29] T. N. Beran, A. Ramirez-Serrano, R. Kuzyk, M. Fior, and S. Nugent, "Understanding how children understand robots: Perceived animism in childrobot interaction," *International Journal of Human-Computer Studies*, vol. 69, no. 78, pp. 539–550, 2011.

[30] P. Alves-Olivera, T. Ribeiro, S. Petisca, E. di Tullio, F. S. Melo, and A. Paiva, "An empathic robotic tutor for school classrooms: cosidering expection and satisfaction of children as end-users," in *International Conference of Social Robotics*. Paris, France: Spinger, 2015, pp. 21–30.

[31] N. Borgers, E. de Leeuw, and J. Hox, "Children as respondents in survey research: cognitive development and response quality," *Bulletin de Methodologie Sociologique*, vol. 66, pp. 60–75, 2000.

[32] S. Y. Okita and D. L. Schwartz, "Young children's understanding of animacy and entertainment robots," *International Journal of Humanoid Robotics*, vol. 3, no. 3, pp. 393–412, 2006.

[33] M. Kellet and S. Ding, "Middle childhood," in *Doing research with children and young people*, S. Fraser, V. Lewis, S. Ding, M. Kellett, and C. Robinson, Eds. London: Sage, 2004, pp. 161–174.

[34] A. Coninx, P. Baxter, E. Oleari, S. Bellini, B. Bierman, B. Henkemans, L. Can amero, P. Cosi, V. Enescu, R. R. Espinoza, A. Hiolle, R. Humbert, B. Kiefer, I. Kruijff-Korbayova, R. Looije, M. Mosconi, M. Neerincx, G. Paci, G. Patsis, C. Pozzi, F. Scchitelli, H. Sahli, A. Sanna, G. Sommavilla, F. Tesser, Y. Demiris, and T. Belpaeme, "Towards long-term social child-robot interaction: using multi-activity switching to engage young users," *Journal of Human Robot Interaction*, vol. in press, 2015.

[35] L. Takayama, D. Dooley, and W. Ju, "Expressing thought: improving robot readability with animation principles," in *Human-Robot Interaction, HRI*. Lausanne, CH: ACM, 2011, pp. 69–76.

[36] C. L. Bethel, K. Salomon, R. R. Murphy, and J. L. Burke, "Survey of psyhophysiology measurements applied to human- robot interaction," in *RO-MAN*, 2007.

[37] S. M. Anzalone, S. Boucenna, S. Ivaldi, and M. Chetouani, "Evaluating the engagement with social robots," *International Journal of Social Robotics*, vol. 7, no. 4, pp. 465–478, 2015.

[38] I. Leite, R. Henriques, C. Martinho, and A. Paiva, "Sensors in the wild: Exploring electrodermal activity in child-robot interaction," in *Proceedings of the 8th ACM/IEEE international conference on Human-robot interaction*. IEEE Press, 2013, pp. 41–48.

[39] K. Dautenhahn, "Methodology and themes of human-robot interaction: a growing research field," *International Journal of Advanced Robotic Systems*, vol. 4, no. 1, pp. 103–108, 2007.