Curiosity Notebook: A Platform for Learning by Teaching Conversational Agents

Edith Law
Parastoo Baghaei Ravari
Nalin Chhibber
Dana Kulic[†]
Stephanie Lin
Kevin D. Pantasdo
Jessy Ceha
Sangho Suh
Nicole Dillen
University of Waterloo, Canada
Monash University, Australia[†]
edith.law@uwaterloo.ca
parastoo.baghaei.ravari@uwaterloo.ca
nalin.chhibber@uwaterloo.ca
dana.kulic@uwaterloo.ca
stephanielin78@gmail.com
kevin.pantasdo@edu.uwaterloo.ca
jceha@uwaterloo.ca
shsuh@uwaterloo.ca
nicole.dillen@uwaterloo.ca

CHI '20 Extended Abstracts, April 25–30, 2020, Honolulu, HI, USA.

ACM ISBN 978-1-4503-6819-3/20/04.
http://dx.doi.org/10.1145/3334480.3382783

Abstract

Learning by teaching is an established pedagogical technique; however, the exact process through which learning happens remains difficult to assess, in part due to the variability in the tutor-tutee pairing and interaction. Prior research proposed the use of teachable agents acting as students, in order to facilitate more controlled studies of the learning by teaching phenomenon. In this work, we introduce a learning by teaching platform, Curiosity Notebook, which allows students to work individually or in groups to teach a conversational agent a classification task in a variety of subject topics. We conducted a 4-week exploratory study with 12 fourth and fifth grade elementary school children, who taught a conversational robot how to classify animals, rocks/minerals and paintings. This paper outlines the architecture of our system, describes the lessons learned from the study, and contributes design considerations on how to design conversational agents and applications for learning by teaching scenarios.

Author Keywords

conversational agents; learning by teaching

CCS Concepts

•Human-centered computing → User studies; •Applied computing → Interactive learning environments; Collaborative learning;

Introduction

Learning by teaching is a popular and well studied pedagogical technique. Research has shown that this technique produces the *protégé effect* [7]—students who are asked to teach others learn through the generation of explanations and questions, which requires a form of knowledge building that is particularly conducive to learning—students synthesize and structure materials, become aware of their own learning process, and expend more effort to learn.

Despite extensive research, our understanding of the exact conditions that make learning by teaching effective is limited, mostly because there can be high variability in the tutor and tutee behaviour. One recommendation, put forth by Roscoe in a large survey on tutor learning [19], is to "develop teachable agents to test hypotheses about specific tutor behaviours" by systematically manipulating the teachable agent's question asking behavior (e.g., shallow vs deep questions), accuracy (e.g., few vs frequent mistakes), and level of prior knowledge (e.g., high vs low).

Figure 1: Learning by Teaching a Conversational Robot

In this work, we introduce a learning by teaching web application called the Curiosity Notebook that supports learning by teaching through dialogue. We conducted a 4-week exploratory study with 12 fourth and fifth grade elementary school students teaching a humanoid robot to classify objects, to understand how to design a platform that can support research on learning by teaching. In summary, our work contributes:

- a configurable learning by teaching platform that enables students to teach a virtual, voice-only, or physical robot agent, individually or in groups, and on different topics,
- an outline of design goals that are important for the development of group-based learning by teaching platforms,
- insights from an exploratory study on potential refinements of the design goals.

Background

In education research, learning by teaching is closely related or synonymous to other terms, such as peer tutoring (PT), cooperative learning (CL), and peer-assisted learning (PAL). It is hypothesized that many of the activities demanded by teaching—e.g., explaining [23], questioning [9], assessment and feedback [15]—require *reflective knowledge building*, where students synthesize, structure and reflect [19]. Roscoe and Chi [19] proposed that teachable agents can serve as an infrastructure for testing different hypotheses about tutor behaviour. In computer-mediated learning applications, agents have mostly served as peers [20, 14] or tutors [12, 17], with only a handful of systems positioning the agent as a less knowledgeable peer that students teach [2, 4]. SimStudent [16] is a simulated learner used to study student-tutor learning in mathematics problem solving. In Betty's Brain [3, 2], students read articles, then teach and quiz a virtual agent (i.e., Betty) about causal relationships (e.g., burning fossil fuels increases CO_2) in science by manipulating concept maps.

Other teachable agent research involves physical robots. In Tanaka and Matsuzoe [22], young children (age 3-6) taught a humanoid robot English words, while simultaneously interacting with a human teacher. Later, they [21] also investigated how preschool children learn English by teaching Pepper, an adult-size humanoid robot, while receiving guidance from a human teacher demonstrating vocabulary-related gestures on a small screen. Yadollahi et al. [24] developed a collaborative story reading environment, where children (aged 6-7) can correct the robot's mistakes as it reads aloud. In other works [13, 6], children—working individually, in pairs and in groups—corrected a robot's handwriting. Finally, Chaffey et al. [5] had older students (mean age=20) teach a robot to solve math problems, and explored how dyadic stance affects student attitudes.

Curiosity Notebook

Our goal is to create a platform that facilitates student learning through teaching a conversational agent. Several design goals guided our development—To facilitate testing of hypotheses around learning by teaching, the platform should enable systematic modulation of the tutee characteristics hypothesized to be relevant to learning (e.g., types of question asked, accuracy), and provide students with choices of teaching tasks in order to allow for a quantitative characterization of their teaching strategies. To be feasible in real-world learning environments, the platform should support students teaching individually, in pairs or in larger groups, while providing equal access to teaching opportunities. Finally, to produce generalizable findings, the platform should support teaching conversations with agents with different embodiments (e.g., virtual, voice or physical agents) and learning tasks that can scale in complexity to different age groups (e.g., usable by elementary school *and* college students). These platform features, many of which are beyond what is provided by existing learning-by-teaching systems, enable researchers to ask a wide range of research questions about conversational agents within the context of diverse collaborative learning scenarios.

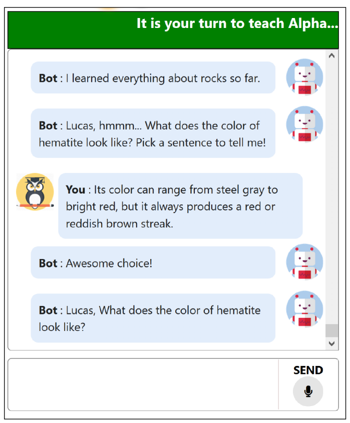

Figure 2: Example of Teaching Conversations about Rock Classification

Web Interface

The Curiosity Notebook provides a web interface that students use to read articles and teach a conversational agent how to classify objects, e.g., classifying paintings as impressionist, cubist, or realist art; animals as mammals, insects, and reptiles; or rocks as metamorphic, igneous or sedimentary. We chose classification tasks because they can be taught in a structured way—classifying objects involves identifying features and knowing how they map to each category. Classification tasks are also amenable to machine learning, allowing computational models of learning to be implemented in the agent [8].

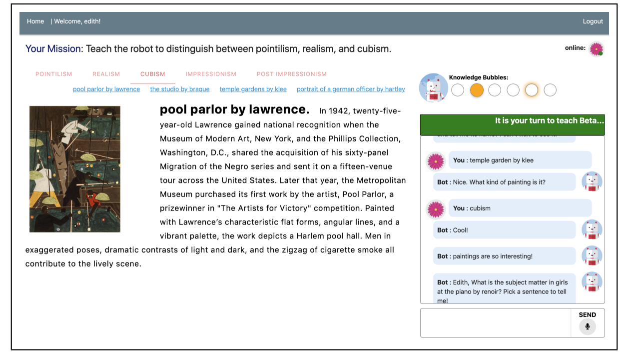

Figure 3: Teaching Interface

The main teaching interface consists of a reading panel (left), containing articles about objects (e.g., "The Gulf Stream" by Homer) belonging to different categories (e.g., realist paintings). Interactive functions allow students to highlight sentences to teach the agent. A chat window (bottom right), which can made visible or hidden for voice-only agents or physical robots, is used by the student teachers to converse with the agent. The state of the agent's learning is represented by *knowledge bubbles* (top right), each representing a feature that is relevant for the classification task at hand. For example, a feature relevant for distinguishing mammals and reptiles would be whether the animal lays eggs. A knowledge bubble gets filled when an agent has mastered/learned the associated feature, otherwise it remains empty.

Design of Teaching Conversation

The agent begins each teaching conversation by asking the students to show it a physical artifact (e.g., "Can you pick an animal and tell me its name? I can't wait to see it!"). After selecting an object, the teaching conversation proceeds with the agent highlighting one of the knowledge bubbles and asking a series of 4 or 5 questions about the corresponding feature. The agent is designed to ask a mix of low and high level questions, as inspired by classification schemes for categorizing questions [18, 11, 10]. Low-level thinking questions include questions about features (e.g., "What does the skin of mammals look like?"), examples (e.g., "Can you give me an example of a cubism painting?") and facts (e.g., "Select a sentence to tell me about frogs and how they lay eggs.") High-level thinking questions include why questions (e.g., "Why is a snake a reptile?"), synthesis questions (e.g., "Do all reptiles look the same?"), and questions that prompt students to repeat/rephrase or explain the meaning of a word, e.g., "Can you help me understand what you just said better?"

The system automatically generates a variety of questions by filling in pre-defined sentence templates with names of objects, features and categories that the students are currently learning about. The questions are sequenced so that lower-level thinking questions always precede higher-level thinking questions. Some randomness was introduced for ordering questions to prevent the conversation from being too mechanical. The teachable agent occasionally seeks feedback from students about its learning, by asking questions about its general intelligence (e.g., "Am I smart?"), its learning progress (e.g., "Am I learning?", "Do you think I know more now than before?"), or how well it might perform if tested ("Will I do well in a test?").

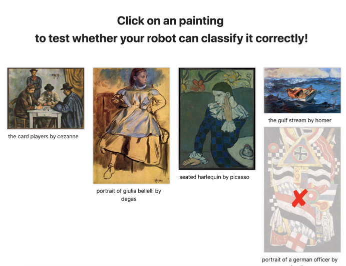

Figure 4: Testing Interface

After 4-5 rounds of questioning and answering, the knowledge bubble is filled and students are rewarded with confetti on the screen letting them know that the agent has "learned" that feature, and the next feature is randomly chosen for the student to teach. Between teaching conversations, students can also choose to test the agent's knowledge. The testing interface shows a set of images representing objects (e.g., images of paintings) to be classified. Students can click on an image, and the agent will attempt to classify it (e.g., saying "I think it is an impressionist painting" or "I don't know"). The system provides corrective feedback—an overlay over each image will show a green checkmark if the agent is correct, and a red "x" otherwise. Our teachable agent is currently *simulated* to learn. As the system currently lacks natural language processing (NLP), the agent does not understand students' responses to questions, and always pretends to learn what students have taught. Initially, the agent is unable to answer any test questions correctly; its ability to answer test questions increases with the number of features it has learned (i.e., number of completed teaching conversations).

Coordinated Group-Based Teaching

Students can teach the agent individually or in groups of arbitrary size, and their group placement can be configured by teachers or researchers through an administrative interface. If a student is placed in a group and their group members are present, their view of the system is synchronized—that is, if one student navigates to another interface (e.g., teaching vs testing), all students will be automatically brought to the same screen.

Agent Embodiment

Our platform can support virtual, voice-only, as well as physical robot agents. To connect to NAO, a separate program written in Python (a NAO-supported API) would ping the Curiosity Notebook's database and send messages to the robot to speak a sentence or deliver a gesture. This setup enables the platform to be connected to any robotic platform (e.g., NAO, Pepper) or speakers (e.g., Google Home, Alexa) via a script unique to that device. This clean separation between the conversational agent's logic and delivery devices also allows us to, in future work, easily test hypotheses about different embodiments of the conversational agent and their effects on learning by teaching.

Exploratory Study

We conducted a 4-week exploratory study with 12 fourth and fifth grade students at a local school. Students (7M/5F) participated in the study over 4 weeks. Six of the parents reported English as their only primary language at home. Two reported another language alongside with English, and four indicated the primary language at home is a language other than English. On a 5-point scale, participants reported moderate amount of experience with computers and smartphones (m=3.9, sd=1.2) and little experience with robots (m=1.9, sd=1.2). Enrollment was on a first come first serve basis. No monetary compensation was provided; instead, students were given a “Certified Robot Teacher” certificate as a token of appreciation.

Figure 5: Physical Artifacts for Classification

The study was conducted in an after-school club, which ran once a week for 1.5 hours each. Four NAO robots were used in each session; to personalize the experience, each robot has a name tag hung around their neck with a gender-neutral name (i.e., Alpha, Beta, Gamma and Delta). Students formed groups of 3, and taught the robot about a different topic (i.e., animals, rocks, paintings) each week, then all topics during the last week. Each student was given a chromebook, and sat together with their group members facing the robot, which was positioned in a sitting posture in front of the students on the table (as shown in Figure 1). Each group of students was joined by a student researcher, who observed the group and answered questions if issues arise with the platform. We provided physical artifacts (Figure 5) for each classification task, namely animal figurines, rocks and minerals, and postcards of different styles of paintings from NYC Metropolitan Museum. During the session, we piloted a variety of surveys, iteratively re-designed the platform, made detailed observations, and interviewed students about their learning by teaching experience. We report the lessons learned as follows.

Dimensions of Tutee Characteristics

During the post-study interview, we asked the students—“Is your robot a good student? Why or why not?”—a question that gave us a window into what children see as the main attributes of a good learner. Students mentioned that the agent is a good learner because of its **attentiveness** (e.g., “because it pays very close attention”, “because he/she ... sits in one spot and doesn’t get distracted”), **curiosity** (e.g., “because it’s curious”, “because Delta asks questions, just like a human student”), and its **ability and eagerness to learn** (e.g., “because he got everything right”, “because he’s always ready to learn”). The more negative rationale mentions that “talking too much” as the reason for the agent for not being a good student. These tutee characteristics can be easily parameterized; for example, attentiveness can be modulated through eye gaze, talkativeness can be modulated by the number of sentences the agent says and the length of the sentences, and curiosity can be expressed through frequent question asking.

The extent to which the agent is a “good” student and what the agent says can both affect students’ perception of their own competence as teachers. The majority of the students in our study saw themselves as good teachers. Students attributed their success at teaching to not only the learning progress of the robot (e.g., “because my robot has learned a lot”, “because we got all the bubbles for animals”), but also to the positive feedback they received from the robot (e.g., “because the robot told me so”, “because Delta always says good choice”). This suggests that the verbal repetoire of the teachable agent should include feedback to the students about their teaching.

Quantification of Teaching Strategies
One of our design goals is to capture more objectively the teaching strategies of the student teachers, e.g., quantitative data about the students' teaching vs testing schedule. We observed, for example, that some groups filled as many knowledge bubbles as possible before they tested the knowledge of their robot; whereas other groups tested the robot often, e.g., after each filled bubble. However, our current platform limits the tutors' choice in *how* they teach—the agent controlled the entire sequence of teaching interactions by posing questions for the tutors to answer. Quite a number of students expressed a desire to be more *proactive* in how they teach; specifically, they wanted to be able to pose their own questions to the agent. In future work, providing a way for students to choose the next teaching topic and ask questions that the agent can *appear to* intelligently answer can help us understand the students' *process* of teaching and how their process affects learning.

Complexity of Learning Task and Material
Over the course of the study, students had the opportunity to experience different topics. When asked which topic (animals vs. rocks vs. paintings) they liked the most, the responses can be clustered into two groups: (1) students who liked teaching a topic because it was easier, because they knew more about it, and because they perceived the robot to be learning more/better about that topic, and (2) students who liked teaching a topic because they knew less about it. This observation implies that personality trait (e.g., the desire for challenge, growth vs fixed mindset) can critically affect students' preferences of topics to teach and how much they enjoy the teaching experience.

Coordination of Group-Based Teaching
During the initial session, the Curiosity Notebook gave students complete freedom to choose *what* and *when* to teach the robot. This setup was too open ended—students had great difficulty narrowing down what content matters and dividing the teaching task. Subsequently, the agent was redesigned to control turn taking—it determines which group member is online and active, asks the student who has participated the least number of turns to teach next. When a student is stuck (e.g., picked a sentence unrelated to what the agent is asking about), the agent will also delegate the task to the next student, asking him/her to help.

Overall, the turn taking mechanism enabled multiple student groups of different sizes to simultaneously teach different robots in the same classroom. Some students took the initiative to offer help to their teammates when it was not their turn to teach, while others were impatient at having to wait. Interestingly, the amount of attention that the robot gives to each student teacher seems to also affect students' perception of their own teaching ability; one student said "Student *X* teaches way better because the robot chooses *X* more." Together, these observations suggest a more personalized approach to managing group-based teaching that takes into account each student's ability to work in team and his/her unique need for attention from the agent.

Conclusion
In this work, we introduce a platform called the Curiosity Notebook that allows students to learn by teaching a conversational agent. Findings from our exploratory study provide insights into factors that matter in the design of conversational, group-based learning by teaching scenarios. Already, the platform has been deployed to study question generation [1] and learning by teaching in crowdsourcing contexts [8]. Future work involves re-designing the platform to facilitate proactive teaching, personalization based on individual and group characteristics, and the use of the agent's internal states to scaffold learning and teaching.

REFERENCES

[1] Mehdi Alaimi, Edith Law, Kevin D. Pantasdo, Pierre-Yves Oudeyer, and Hélène Sauzeon. 2020. Pedagogical Agents for Fostering Question-Asking Skills in Children. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '20)*. 1–10. DOI: http://dx.doi.org/10.1145/3313831.3376776

[2] Gautam Biswas, Krittaya Leelawong, Daniel Schwartz, Nancy Vye, and The Teachable Agents Group at Vanderbilt. 2005. Learning by Teaching: A New Agent Paradigm for Educational Software. *Applied Artificial Intelligence* 19, 3-4 (2005), 363–392. DOI: http://dx.doi.org/10.1080/08839510590910200

[3] Gautam Biswas, James R. Segedy, and Kritya Bunchongchit. 2016. From Design to Implementation to Practice a Learning by Teaching System: Betty's Brain. *International Journal of Artificial Intelligence in Education* 26, 1 (2016), 350–364. DOI: http://dx.doi.org/10.1007/s4059

[4] Sean Brophy, Gautam Biswas, Thomas Katzlberger, John Bransford, and Daniel Schwartz. 1999. Teachable Agents: Combining Insights from Learning Theory and Computer Science. 50 (1999), 21–28.

[5] Tricia Chaffey, Hyeji Kim, Emilia Nobrega, Nichola Lubold, and Heather Pon-Barry. 2018. Dyadic Stance in Natural Language Communication with a Teachable Robot. In *Proceedings of the Companion of the ACM/IEEE International Conference on Human-Robot Interaction (HRI '18)*. ACM, New York, NY, USA, 85–86. DOI: http://dx.doi.org/10.1145/3173386.3176979

[6] Shruti Chandra, Raul Paradeda, Hang Yin, Pierre Dillenbourg, Rui Prada, and Ana Paiva. 2017. Affect of Robot's Competencies on Children's Perception. In *Proceedings of the Conference on Autonomous Agents and Multi-Agent Systems (AAMAS '17)*. 1490–1492.

[7] Catherine C. Chase, Doris B. Chin, Marily A. Oppezzo, and Daniel L. Schwartz. 2009. Teachable Agents and the Protégé Effect: Increasing the Effort Towards Learning. *Journal of Science Education and Technology* 18, 4 (01 Aug 2009), 334–352. DOI: http://dx.doi.org/10.1007/s10956-009-9180-4

[8] Nalin Chhibber. 2019. *Towards the Learning, Perception and Effectiveness of Teachable Conversational Agents*. Master's thesis. University of Waterloo, Waterloo, Ontario, Canada.

[9] Meredith D. Gall. 1970. The Use of Questions in Teaching. *Review of Educational Research* 40, 5 (1970), 707–721. DOI: http://dx.doi.org/10.3102/00346543040005707

[10] James J. Gallagher and Mary Jane Aschner. 1963. A Preliminary Report on Analyses of Classroom Interaction. *Merrill-Palmer Quarterly of Behavior and Development* 9, 3 (1963), 183–194.

[11] Arthur C. Graesser and Natalie K. Person. 1994. Question Asking during Tutoring. *American Educational Research Journal* 31, 1 (1994), 104–137. DOI: http://dx.doi.org/10.3102/00028312031001104

[12] Arthur C. Graesser, Katja Wiemer-Hastings, Peter Wiemer-Hastings, Roger Kreuz, and Tutoring Research Group. 1999. AutoTutor: A Simulation of a Human Tutor. *Cognitive Systems Research* 1, 1 (1999), 35–51. DOI: http://dx.doi.org/10.1016/S1389-0417(99)00005-4

[13] Deanna Hood, Séverin Lemaignan, and Pierre Dillenbourg. 2015. When Children Teach a Robot to Write: An Autonomous Teachable Humanoid Which Uses Simulated Handwriting. In *Proceedings of ACM/IEEE International Conference on Human-Robot Interaction (HRI '15)*. 83–90. DOI: http://dx.doi.org/10.1145/2696454.2696479

[14] Takayuki Kanda, Takayuki Hirano, Daniel Eaton, and Hiroshi Ishiguro. 2004. Interactive Robots as Social Partners and Peer Tutors for Children: A Field Trial. *Human–Computer Interaction* 19, 1-2 (June 2004), 61–84. DOI: http://dx.doi.org/10.1207/s15327051hci1901&2_4

[15] Chinmay Kulkarni, Koh Pang Wei, Huy Le, Daniel Chia, Kathryn Papadopoulos, Justin Cheng, Daphne Koller, and Scott R. Klemmer. 2013. Peer and Self Assessment in Massive Online Classes. *ACM Transaction on Computer-Human Interaction* 20, 6, Article 33 (Dec. 2013), 31 pages. DOI: http://dx.doi.org/10.1145/2505057

[16] Noboru Matsuda, Evelyn Yarzebinski, Victoria Keiser, Rohan Raizada, William W. Cohen, Gabriel J. Stylianides, and Kenneth R. Koedinger. 2013. Cognitive Anatomy of Tutor Learning: Lessons Learned with SimStudent. *Journal of Educational Psychology* 105, 4 (2013), 1152–1163. DOI: http://dx.doi.org/10.1037/a0031955

[17] Roxana Moreno, Richard E. Mayer, Hiller A. Spires, and James C. Lester. 2001. The Case for Social Agency in Computer-based Teaching: Do Students Learn More Deeply hen They Interact with Animated Pedagogical Agents? *Cognition and Instruction* 19, 2 (2001), 177–213. DOI: http://dx.doi.org/10.1207/S1532690XCI1902_02

[18] Taffy E. Raphael and P. David Pearson. 1985. Increasing Students' Awareness of Sources of Information for Answering Questions. *American Educational Research Journal* 22, 2 (1985), 217–235. DOI: http://dx.doi.org/10.3102/00028312022002217

[19] Rod D. Roscoe and Michelene T.H. Chi. 2007. Understanding Tutor Learning: Knowledge-building and Knowledge-telling in Peer Tutors' Explanations and Questions. *Review of Educational Research* 77, 4 (2007), 534–574. DOI: http://dx.doi.org/10.3102/0034654307309920

[20] Jong-Eun Roselyn Lee, Clifford Nass, Scott Brenner Brave, Yasunori Morishima, Hiroshi Nakajima, and Ryota Yamada. 2006. The Case for Caring Colearners: The Effects of a Computer-mediated Colearner Agent on Trust and Learning. *Journal of Communication* 57, 2 (2006), 183–204. DOI:http://dx.doi.org/10.1111/j.1460-2466.2007.00339.x

[21] Fumihide Tanaka, Kyosuke Isshiki, Fumiki Takahashi, Manabu Uekusa, Rumiko Sei, and Kaname Hayashi. 2015. Pepper Learns Together with Children: Development of an Educational Application. In *Proceedings of the IEEE-RAS International Conference on Humanoid Robots (Humanoids '15)*. IEEE, 270–275. DOI:http://dx.doi.org/10.1109/HUMANOIDS.2015.7363546

[22] Fumihide Tanaka and Shizuko Matsuzoe. 2012. Children Teach a Care-Receiving Robot to Promote Their Learning: Field Experiments in a Classroom for Vocabulary Learning. *Journal of Human-Robot Interaction* 1, 1 (2012), 78–95. DOI: http://dx.doi.org/10.5898/JHRI.1.1.Tanaka

[23] Jörg Wittwer and Alexander Renkl. 2008. Why Instructional Explanations Often Do Not Work: A Framework for Understanding the Effectiveness of Instructional Explanations. *Educational Psychologist* 43, 1 (2008), 49–64. DOI: http://dx.doi.org/10.1080/00461520701756420

[24] Elmira Yadollahi, Wafa Johal, Ana Paiva, and Pierre Dillenbourg. 2018. When Deictic Gestures in a Robot Can Harm Child-robot Collaboration. In *Proceedings of the ACM Conference on Interaction Design and Children (IDC '18)*. ACM, New York, NY, USA, 195–206. DOI: http://dx.doi.org/10.1145/3202185.3202743