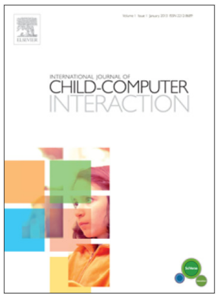

Sensing technologies and child-computer interaction: Opportunities, challenges and ethical considerations

Kshitij Sharma, Michail Giannakos

PII: S2212-8689(21)00052-0
DOI: https://doi.org/10.1016/j.ijcci.2021.100331
Reference: IJCCI 100331

To appear in: *International Journal of Child-Computer Interaction*

Received date : 14 November 2020
Revised date : 4 April 2021
Accepted date : 12 May 2021

Please cite this article as: K. Sharma and M. Giannakos, Sensing technologies and child-computer interaction: Opportunities, challenges and ethical considerations. *International Journal of Child-Computer Interaction* (2021), doi: https://doi.org/10.1016/j.ijcci.2021.100331.

eSensing Technologies and Child–Computer Interaction: Opportunities, Challenges and Ethical Considerations

Kshitij Sharma and Michail Giannakos
Norwegian University of Science and Technology, Trondheim, Norway

Corresponding author: Kshitij Sharma (kshitij.sharma@ntnu.no)

Abstract

This study presents the outcomes of a systematic literature review of empirical evidence on the capabilities of sensing technologies for child–computer interaction (CCI). This paper provides an overview of what and how sensing technologies have been used to explain, understand, and predict children's experiences with interactive devices and technologies and in what contexts. A search resulted in 44 papers that were included in the analysis. The results of the review depict the capabilities of sensing technologies for gauging children's performance, engagement, and experiences (while interacting with technology) and the ongoing advances and implications that emerge from the employment of sensors to capture and improve child behavior. In particular, we identified the four main objectives (i.e., engagement of children, recognition/prediction of special needs/behavior, explaining/understanding the behavior/attitude, and learning performance/experiences) that the CCI research has been focusing on with sensor data. We also summarize the implications derived from the reviewed articles and frame them within four thematic areas. Finally, this review stresses that future research should consider developing a framework that would enable sensor data capacities to be aligned with the ethical, social, and generalizability guidelines. These sensor capacities could also be utilized to advance theory and practice. Our findings set a baseline for supporting the adoption and democratization of sensor data within future interactive technology research and development for children.

Abbreviations:

ACM → Association for Computing Machinery; **ADHD** → Attention Deficit Hyperactivity Disorder
AHFE → Applied Human Factors and Ergonomics (conference); **AI** → Artificial Intelligence
ASD → Autism Spectrum Disorder; **BVP** → Blood Volume Pulse; **CCI** → Child–Computer Interaction; **CHI** → Computer–Human Interaction; **DIS** → Designing Interactive Systems (conference); **ECCE** → European Conference on Computing Education; **ECG** → Electrocardiogram; **EDA** → Electrodermal Activity; **EEG** → Electroencephalogram; **EMG** → Electromyogram; **FDA** → Food and Drug Administration; **FDG** → Foundations of Digital Games (conference); **GPS** → Global Positioning System; **GSR** → Galvanic Skin Response; **HCI** → Human–Computer Interaction; **HCII** → Human–Computer Interaction International (conference); **HR** → Heart Rate; **HRV** → Heart Rate Variation; **IBI** → Interbeat Interval; **IDC** → Interaction Design and Children (conference); **IEEE** → Institute of Electrical and Electronics Engineers; **IJCCI** → International Journal of Child–Computer Interaction; **IMWUT** → Interactive, Mobile, Wearable and Ubiquitous Technologies (journal); **LAK** → Learning Analytics and Knowledge (conference); **ML** → Machine Learning; **PPG** → Photoplethysmogram; **RSP** → Respiratory Patterns; **SEN** → Special Educational Needs; **SLR** → Systematic Literature Review; **ST** → Skin Temperature; **STEM** → Science Technology Engineering and Mathematics; **TACCESS** → Transactions on Accessible Computing

(journal); **TD** → Typically Developing; **TEI** → Tangible, Embedded and Embodied Interaction (conference); **THRI** → Transactions on Human–Computer Interaction (journal); **TiiS** → Transactions on Interactive Intelligent Systems (journal); **UAHCI** → Universal Access in Human–Computer Interaction (conference); **UMUAI** → User Modeling and User-Adapted Interaction (journal); **VAMR** → Virtual Augmented and Mixed Reality (conference); VRIC → Virtual Reality International Conference; **WSC** → World Simulation Consortium (conference)

1. Introduction

During the last decade, we have seen an enormous penetration of ubiquitous technologies, such as tablets, smartphones, and wearables, into children's lives (Kabali et al., 2015). Due to their ubiquitous nature, these technologies are accessible to young children, supporting their play, communication, learning, and other application areas. Technologies such as smart displays, motion-capture systems, and smart toys, enable children to perform more complex interactions, such as gestures, waving, handshaking, and other motion-based interaction. Many of these technologies, such as robots and other automated systems, are designed to support specific age groups and children's needs and abilities. Such technologies enable various data collections via their ability to sense, among other things, children's movements, gaze, and skin.

Sensing technologies' key affordances, such as their temporality or the ability to directly access new indicators of cognitive and affective processes (Cukurova et al., 2020), consist of relatively new and promising forms of information in CCI research. However, there is no consolidated overview of whether and how the technological developments of sensing are appropriate and beneficial for children's interactive experiences. In this literature review, we focus on the sensing capability of these technologies (i.e., monitoring children's skin conductance, heart rate, respiratory patterns, motion, gaze, blood volume pulse, facial features, postures, gestures, speech, and other special physiological data sources [e.g., neuroendocrine and other hormones]). These sensor-based data sources are often used in CCI research to inform researchers and practitioners about children's attentive, cognitive, and behavioral patterns. This study investigates the opportunities, challenges, and ethical considerations of the "sensing capabilities of technology"[1] for CCI research.

Technologies that enable sensing are now widely available and affordable. They provide the capacity to obtain and store data about an individual's experience with a technology. At the same time, these devices have made it possible to monitor more subtle phenomena, such as the quality of social interactions, students' mental health, and learning engagement (Wang et al., 2018). Worsley and Blikstein (2018) argued that sensing technologies could provide more meaningful insights into complex learning processes than are possible from traditional approaches. Similarly, Spikol et al. (2018) highlighted the importance and benefits of sensing to support open- ended tasks. Another example is quantified-self movement, which has shown its potential to utilize authentic and granular activity data in order to inform users about their lifestyle and fitness (Lee et al., 2016), with the aim of involving the user in self-monitoring and self-reflection processes to regulate different aspects of their life and behavior (Ruiz et al., 2016).

[1] Also referred to as a sensing technology throughout this paper.

These intriguing sociotechnical developments over the last decade, combined with the development of different technologies and sensors (e.g., wristbands, cameras), posit the sensing capability of technologies as an emerging paradigm that potentially provides opportunities and challenges for CCI research. For instance, sensing can lead to significant concerns among CCI researchers and stakeholders about the invasion of children's privacy; bias, fairness, accountability, and transparency of the sensor data; and the risk of enabling and nurturing at this very early age a "surveillance" culture through constant monitoring of children's behaviors.

In light of these potential benefits and challenges, this paper presents a systematic literature review (SLR), strictly following the guidelines of Kitchenham and Charters (2007). The aim of this SLR is to examine the empirical evidence on the status of sensing technologies in CCI research, as well as to identify the challenges, opportunities, and limitations of sensing technologies in CCI research. This SLR will allow us to provide information about the ways in which sensing has been utilized, and about its impact on CCI research. This paper presents an overview of what and how sensing technologies have been used and in what contexts. Although sensing technologies have not been introduced in the CCI field as much as in other HCI areas (e.g., immersive environments, mobile and ubiquitous computing), enough work has already been done to conduct a review. Specifically, the following research questions guide our work.
RQ1: What is the current status of sensing technologies for CCI research?

- **RQ1.1**: Which sensing technologies are used in CCI research?
- **RQ1.2**: What types of data do these sensing technologies capture?
- **RQ1.3**: What insights come from sensing technologies in CCI research?

RQ2: What are the challenges, opportunities, and limitations of sensing technologies in CCI research?

Leveraging sensing technologies to amplify children's interaction potential, or utilizing them as a mechanism to explore children's experiences while interacting with systems (Lee-Cultura et al., 2020) is a valuable capacity. The use of sensing technology has been employed in CCI research as well as in neighboring communities (e.g., the learning sciences) in different ways. For instance, the use of video recordings that were later coded by a human expert is an established approach in qualitative research, and the use of eye tracking or other sensor data is also not new. However, the use of sensing technology as an automated, unobtrusive, and continuous data collection approach that produces complementary insights (Thieme et al., 2020) for traditional HCI data collection might serve as a useful information resource for CCI research. At the same time, this use encompasses unprecedented concerns about the invasion of children's privacy, about bias, fairness, accountability, and transparency of the sensor data, and about the respective insights and extracted practices (Charisi et al., 2020; Cukurova et al., 2020; Frauenberger et al., 2019). All these concerns require in-depth investigation and debate in the CCI community.

2. Related Work

In recent years, the CCI community has engaged in discussions about the promised benefits and ethical issues of sensing and logging technologies in evaluating children's behavior (Hourcade et al., 2018). These technologies include tracking of movements and keystrokes, the capture of facial movements, eye tracking, and the use of biomarkers and quantified data, such as measurements of EDA, blood pressure, and heart rate. In some cases, other sophisticated sensing devices are employed. For example, Bian et al. (2013) used an impedance cardiogram to monitor the stress levels in a virtual instruction system. Similarly, Zhang et al. (2015) used EMG along with other sensors to quantify cognitive load in another instruction system.

However, using sensors in studies involving children entails several ethical and privacy-related implications that need to be considered (Dowthwaite et al., 2020; Kawas et al., 2020). Despite the challenges of sensing technologies as a method for evaluating children's behavior, previous works advocate the use of such technologies to capture complex interactions between children and the systems with which they engage in different contexts and for different purposes (e.g., Kourakli et al., 2017; Papavlasopoulou et al., 2018). In particular, Kourakli et al. (2017) used motion-based sensing to analyze how children with special educational needs engage with a motion-based touchless game. Papavlasopoulou et al. (2018) provided evidence based on eye tracking for the relation between joint attention (how much collaborating peers look at the same set of objects in the same time window) and children's performance in a collaborative coding task. In both these studies, there were special processes for preserving children's anonymity: Kourakli et al. (2017) used the session identifiers for individuals, and Papavlasopoulou et al. (2018) used only group-based identifiers. Therefore, sensing technologies were employed to serve specific and well-informed purposes that otherwise could not be explored using traditional means (e.g., identifying common gaze during programming or monitoring children's detailed movement patterns).

This line of research involves a rich set of multimodal data that can inform the researcher about temporal and sometimes unique qualities of children's experience and behavior (Crescenzi-Lanna, 2020). For instance, eye tracking enables children's attention to be observed, and it provides insights into their cognitive effort when engaging with a system (Papavlasopoulou et al., 2018). EDA and temperature can be used to infer engagement and stress (Di Lascio et al., 2018), while facial videos can, to a limited extent, reveal the emotions displayed by children throughout the different events of the interaction (Amos et al., 2016). Wristband sensors have been coupled with wearable applications to enable children and parents to track changes in their emotions (Betancourt et al., 2017). In a recent study, Lee-Cultura et al. (2020) utilized sensing technology in an informal game-based learning setting to showcase how multiple sensor data provide better opportunities to explain and predict children's behavior than the individual data sources.

There are several information-rich SLRs in the two fields of CCI and sensing technologies (individual and multimodal). However, no SLR has explicitly connected these two fields. For example, CCI-related conferences (e.g., IDC, CHI, LAK, TEI) have attracted a few literature reviews concerning the different aspects of CCI research. Although these works did not focus on the role of sensing technology in contemporary CCI research, certain connections and implications were extracted. For example, Van Mechelen et al. (2020), in a review of the ethical aspects of research involving children, highlighted aspects connected with the type of data collection, data protection, and data privacy, as well as with the child's and parent's awareness of the type of information collected by the researcher and why. In the same vein, Kawas et al. (2020), in a critical reflection on the values and ethics over the past decade, indicated the importance of upholding children's privacy and the role of sensitive data collection from children. Baykal et al. (2020) reviewed the literature on collaborative technologies, including those used with children with special needs, and addressed factors such as social interaction, engagement, and communication among the children and the supporting persons (e.g., parents, teachers). Another noteworthy contribution from a related field is de Barbaro's (2019) review, which showcased how sensing technology can track children's motion, obtain audio and video data, and utilize ML algorithms to extract meaningful markers such as children's stress, depression, and emotional adaptation, and parental stress and academic success.

There are also several SLRs on unimodal and multimodal sensor data and their specific usage. For example, Sharma and Giannakos (2020a) reviewed how multimodal sensor data have been used to understand and predict learners' behavior and/or performance, as well as the importance of these data for the future design of learning technologies. There have been several literature reviews on the use of eye-tracking research, most notably those by Ruhland et al. (2015) on the use of gaze in human–robot interaction and by Jarodzka et al. (2020) on the use of gaze in education. Similarly, there are contributions from the affective computing domain about using video and audio to detect expressions/emotions (Poria et al., 2017) and about the use of physiological, audio/video, and other data sources to detect emotions/expressions in educational settings (Yadegaridehkordi et al., 2019). Concerning EEG, there are several literature reviews based on the collection, processing, and usage of EEG data, among which the closest systematic reviews to CCI research are about the brain–computer interface for gaming (Vasiljevic and de Miranda, 2020) and use of EEG in neurohaptics (Alsuradi et al., 2020). Furthermore, Niknejad et al. (2020) provided a narrow systematic review on intelligent wearable devices, including smart watches, smart glasses, and head-mounted eye trackers.

The previous works and literature reviews in the field indicate the importance of research at the intersection of sensing technologies and CCI. This research domain inherently connects this technology with children's affordances (e.g., embodied interaction) and major debates (e.g., sensitive data). However, as noted above, there are no literature reviews on the confluence of these two major fields (i.e., CCI and sensing technologies). In the present work, we collect, filter, and analyze the research that addresses the intersection between children and technology that uses sensor data as its primary source of investigation. This study summarizes the insights and functions that sensing technology can offer in CCI, and it identifies the challenges and opportunities of sensing technology to further CCI research.

3. Methodology

In this SLR, we follow transparent and widely accepted procedures (especially in the area of computer science, and its sub-fields of HCI and educational technology) to minimize potential biases (researchers) and support the reproducibility of the results (Kitchenham and Charters, 2007).

3.1 Article Collection

Several procedures were followed to ensure a high-quality review of the literature on sensing technologies in CCI. The review includes "non-conventional" children's data (e.g., keystrokes, clickstreams, and other computer logs) coming from sensors, including gaze (general direction, pupilometric measurements, eye-tracking data), facial features (action units, landmarks, emotions, expressions), physiological data (e.g., EEG, EMG, BVP, HRV, hormonal readings, respiration-related measures), and motion-based data (posture, gesture). In order to have a more comprehensive review of the use of sensing technologies, we did not specify any behaviors or outcomes.

A comprehensive search of peer-reviewed articles was conducted on 30 May 2020 (posters, dissertations, editorials, and reports were excluded). The following keyphrase was used to search the databases: **“physiol* data” AND (“interaction design” OR “HCI” OR “human computer interaction”) AND (child* OR kid*)**. In addition, we manually checked the IDC proceedings and IJCCI papers that have been published up to June 2020. There was no time limit put on the year of publication, but the collected articles all come from the last decade. The following databases were searched: SpringerLink, Wiley, ACM Digital Library, IEEE Xplore, Science Direct, SAGE, and ERIC. The search process uncovered 696 peer-reviewed articles.

3.2 Inclusion and Exclusion Criteria

The selection phase determines the overall validity of the literature review, and thus it is important to define specific inclusion and exclusion criteria. We applied eight quality criteria informed by related works (e.g., Mangaroska and Giannakos, 2018)` known as critical appraisal skills program criteria” (for details of the checklist, see Dybå & Dingsøyr, 2008; Greenhalgh, 1997). As Dybå and Dingsøyr (2008) specified, the quality criteria should cover three main issues (rigor, credibility, and relevance) that need to be considered when evaluating the quality of the selected studies. We applied eight quality criteria informed by related works (e.g., Dybå & Dingsøyr, 2008):

1. Does the study clearly address the research problem?
2. Is there a clear statement of the aims of the research?
3. Is there an adequate description of the context in which the research was carried out?
4. Was the research design appropriate to addressing the aims of the research?
5. Does the study clearly determine the research methods (subjects, instruments, data collection, data analysis)?
6. Was the data analysis sufficiently rigorous?
7. Is there a clear statement of findings?
8. Is the study of value for research or practice?

Once the papers were selected based on the critical appraisal skills program criteria (Dybå and Dingsøyr, 2008; Greenhalgh, 1997), we applied the following additional criteria for further selection of papers:

1. CCI research covers end users, ranging from toddlers to adolescents (Giannakos et al., 2020), so the participants in the studies had to be 19 years old or younger (Kail, 2011).
2. The papers should present empirical studies with children as the end-user; studies that present design sessions where the children are co-designing sensing technologies were not included. The reason for excluding papers on design sessions with co-design activities is that such studies do not use sensor-based data (i.e., the objective of this review study). Further, review papers were also removed.
3. The analysis and outcomes of the papers should include the sensing data streams used and what measures were employed. This is because of the focus of this paper (e.g., see RQ1).

Finally, we selected 50 papers for further in-depth analysis. Figure 1 shows the breakdown of the selection process with the number of papers.

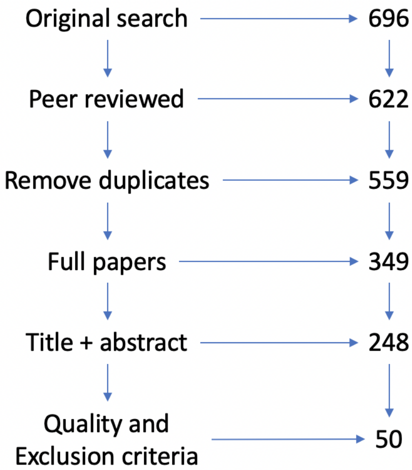

Figure 1. Number of papers at the different phases of the selection process

3.3 Data Analysis

To answer the two research questions through this SLR, we analyzed the selected papers using a coding scheme. This scheme allowed us to consolidate the essence and the main focus of the studies. We selected categories that represent the physiological data utilized as well as the objectives and content of the paper. This categorization enabled us to record all the details needed from the papers and to use them to address our research questions. In particular, each collected study was analyzed using the following elements:

1. **Experimental setting**: this is the space where the data was collected, such as a laboratory, hospital (dedicated special needs facilities), museum, school (as part of classroom activities), or outdoors (outside classroom).
2. **Number of children**: the number of participants in the empirical studies reported in the selected contributions (fewer than 10, 10–30, 30–50, 50–100, more than 100).
3. **Population**: the age range of the children and further categorization of the children's population, including infants (0–2 years), toddlers (2–4 years), preschool (4–6 years), school age (6–12 years), and teenagers (13–19 years).
4. **Type of population** (if SEN, then type of SEN): this coded whether and what type of special needs the participants had.
5. **Research design**: ethnography, evaluation, and quasi-experimental.
6. **Type of data collected**: this is the type of data collected, such as gaze, EDA, EEG, ECG, and PPG.
7. **Brands of the data collection technology**: this coded the brands (or the exact model names, if provided) of the technologies used to collect the sensor data, such as Biopac, Tobii, and Empatica. In case of the contributions using "in-lab" created technology, we coded it as "self-made tech".
8. **Main task of the study**: this is the core task of the contribution, such as basic communication, basic interaction, driving, game, learning task, memorization, none, participatory design, problem solving, robot interaction, role playing, social interaction, and video interaction.
9. **Performance assessment**: this refers to the methods used by the contributions, such as pre-post-questionnaire, pre-post-test, and task-based performance.

10. **Research methodology**: qualitative, quantitative predictive, quantitative inferential, and mixed methods.
11. **Research objective**: this refers to the main theme of the paper showing why the sensor data was used.
12. **Research outcome**: this refers to the main outcome of the contributions.

It is important to highlight that articles were coded based on reported information. Different authors reported information at different levels of granularity, while in some cases the information was missing from the paper. Overall, we endeavored to code the article as accurately and completely as possible. Details on the paper coding are shown in Appendix A1. While coding the initially selected 50 articles, we noticed that six of them did not provide most of the information to be coded across many of the above 12 points. Therefore, we decided to exclude these six papers from the SLR due to the limited information reported.

4. Research Findings

4.1 Population and Research Design

Concerning the sample size employed, a good amount (n=10) of the studies engaged 10 or fewer children, most (n=19) had 10–30 children, six studies had 30–50 children, and another six had 50–100 children. Three studies did not report the exact sample size and only two studies (Jiang et al., 2020; Xu et al., 2012) used a sample size of more than 100 participants. The median population size is 20–30.

Another interesting element of the methodology is the age of the children included in the selected studies. Most studies focused on teenagers, then school age, preschoolers, toddlers and infants, with the median age group being school age. Strangely, a good number of CCI papers did not specify the age of the children. Figure 2 provides the number of contributions for each of the age groups.

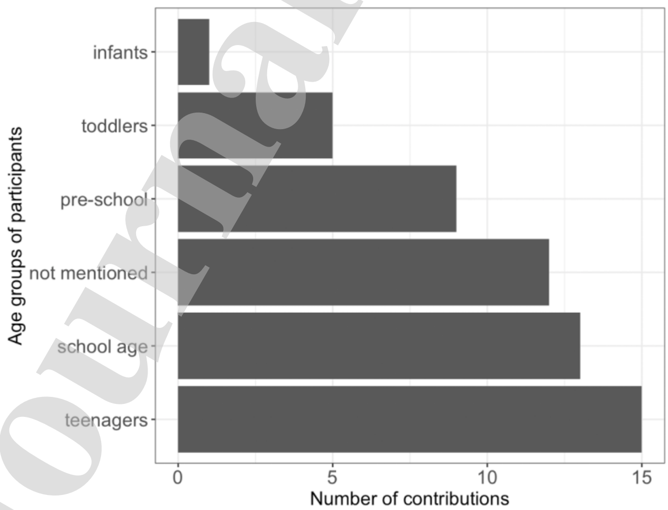

Figure 2. The number of contributions for each of the age groups

Most of the studies focused on typically developing (TD) children (n=22), while 12 contributions focused on SEN children, and seven studies recruited both TD and SEN children as participants. Two studies did not provide any information about the abilities of the participating children. Among the studies that included SEN children, most (n=16) had participants diagnosed with ASD, and two focused on children with ADHD. One study focused on children with Down's syndrome (Saadatzi et al., 2013), another on children with motor planning difficulty (Johnson and Picard, 2017), and another on children with neurodevelopmental communication impairments (Betancourt et al., 2017). Finally, one contribution did not specify the special needs of the children participating in the experiment.

In terms of research design, most studies (n=25) were quasi-experimental or an evaluation study (n=14). Two contributions presented case studies (Dinet and Kitajima, 2018; Johnson and Picard, 2017), one study presented a design experiment (Hauser et al., 2020), and one study presented a usability study (Bekele et al., 2013).

4.2 Interaction Environment and Setting

Most studies (n=25) used laboratories as their experimental setting. Given that there is usually a need for specialized equipment and settings, the frequent use of established user-experience labs is not surprising. However, we also identified a good number of "in-the-wild" studies that took place in schools/classrooms (n=8) and hospitals (n=6) in which the experiments were conducted in a dedicated facility for SEN children. This, too, is unsurprising, since sensing is a technology traditionally used for children with special abilities and often it is necessary to keep the context of the hospital. Four contributions conducted the experiments as out-of-school/classroom activities, and two contextualized their experiments in a museum setting. Figure 3 (left panel) shows the distribution of the papers.

We found a wide range of intended goals/tasks for which the sensing technology was used in the studies. The most frequent goal was connected with gameplay and driving tasks (used in 10 and nine contributions, respectively). These goals were followed by basic interactions with various technologies (n=7), problem solving (n=6), and interactions with robots (n=5). Other goals used in the remaining contributions were learning-based goals (Jeon et al., 2019; Woolf et al., 2009) and video interaction (Castellano et al., 2014; Sonne et al., 2017). Four goals were associated with one contribution each: storytelling (Redata et al., 2017), social interaction (Melo et al., 2019), role playing (Silva et al., 2013), and participatory design (Clegg et al., 2017). A detailed account of our analysis is presented in Figure 3 (right panel).

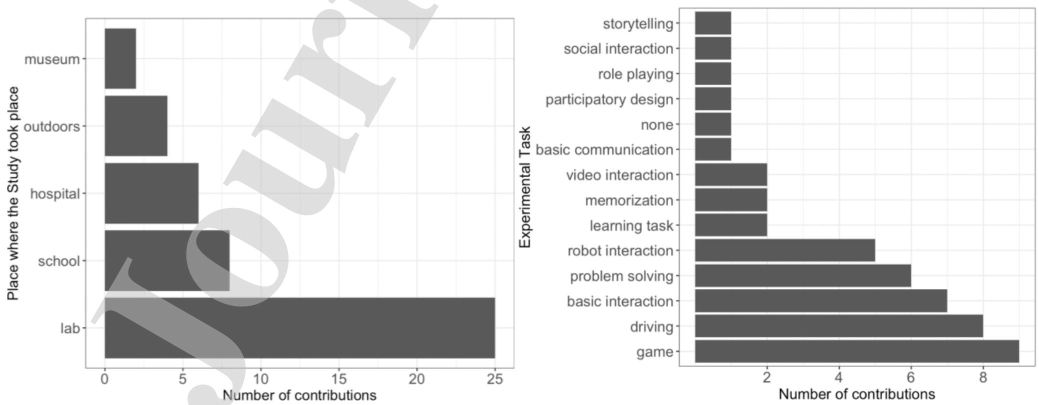

Figure 3. The number of papers based on the contextual setting (left) and the experimental task (right)

4.3 Data Collected and Performance Assessment

In terms of data collection, there were 25 different data modalities used (some being similar but measured with different sensors, such as ECG, HRV, and HR; PPG and BVP; and EDA and GSR). The distribution is shown in Figure 4 (left panel). Most studies used two or more data sources (27 out of 44), while 17 used only one modality. In terms of number of data modalities collected, nine contributions used two data streams, and eight contributions used three data streams. Furthermore, three studies utilized four data streams, four studies used seven data streams, and two data studies used eight data streams. For individual data streams, GSR (n=14) and respiratory patterns (n=13) were the most used data sources, followed by ECG and PPG (n=10). The next most popular data stream is children's gaze (n=8), followed by EEG, EMG, skin temperature (all with six contributions each), EDA, HR, and motion (all with five contributions each). Four studies recorded facial data, while BVP and physical activity were used by three studies each and HRV and speech data were recorded by two studies each. Finally, there were individual cases of using an accelerometer and IBI (Goodwin et al., 2016), GPS and sleeping patterns (Kuzminykh and Lank, 2019), phonocardiogram and impedance cardiogram (Bian et al., 2013), and posture (Woolf et al., 2009). In most of the studies, the data modality was associated with the intended goal of the study; for example, children's gaze was used to capture their attention, reading/writing behavior, and information processing, while motion data was used to capture children's workout activity and embodied interaction.

In terms of brands or models of the technological solutions, there were 11 contributions that did not provide this information, and five contributions used their lab-made technologies for the empirical purposes (Hauser et al., 2020; Johnson and Picard, 2017; Sonne et al., 2017; Woolf et al., 2009; Xu et al., 2012). Among the contributions that used off-the-shelf solutions, 13 used a Biopac product (BioNomadix, Biopac MP150, BioPLUX) and nine used a Tobii eye tracker (TX120, eye-tracking glasses, others that were not mentioned). Empatica and Emotive were used by five and four contributions, respectively. The rest of the off-the-shelf technologies were used in more sparse ways that the four providers just mentioned. These providers (models) included Zephyr BioHarness (two contributions), Affectiva, B-alert, Gazepoint, Neurosky, and Polar (all with one contribution each). To summarize, most of the studies employed highly reliable sensing technology (e.g., medical-quality FDA approved devices) and sometimes we also encountered lab-made technologies to satisfy the needs of the study. Figure 4 (right panel) presents in detail the data collected and the various technologies employed for these data collections.

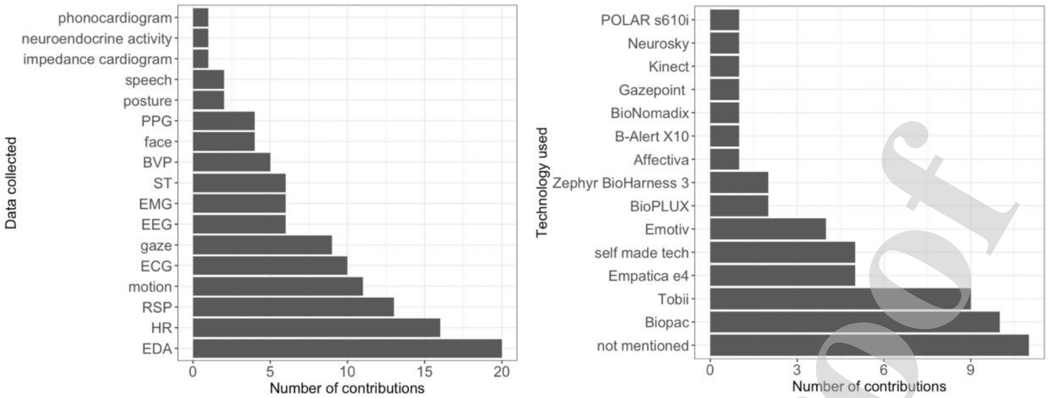

Figure 4. The number of papers based on the sensor data collected (left) and the technology used (right)

An important aspect of research that uses sensor data is connected with what is known in the ML and AI disciplines as the "ground truth". This is either a stable value that depicts either a reliably certain "truth" (e.g., correctness of the child when responding to a multiple-choice question) or a "truth" observed and annotated by humans (e.g., annotating when the child interacts with the system based on a video recording). This information is used together with the information provided by inference from the sensor data to train, test, and, ultimately, evaluate our models. When the method of inference is statistical and not based on ML, this ground truth is replaced by a dependent variable that is decided a priori (in the design phase).

The performance assessment (either ground truth or the dependent variable) in the selected contributions was either missing (n=21, with two others not mentioning the method used for performance assessment) or it used one of the following three techniques for assessing it: (1) a group of studies used task-based performance (e.g., game score, number of mistakes, matched figures) as the performance assessment technique (n=17); (2) three studies used pre-post-tests (Redata et al., 2017; Wade et al., 2015, 2016); and (3) two studies used pre-post-questionnaires (Bian et al., 2019; Sonne et al., 2017) to measure the experiment-based performance of the participating children. Although the absence of ground truth would have been unacceptable in other communities (e.g., AI, ML), in the context of a multidisciplinary research community such as CCI with a substantial body of research focusing on qualitative and exploratory studies, this outcome is not surprising. Several researchers used insights from sensing technology for exploratory purposes, or to complement their understanding and formulate hypotheses that were tested in another research study.

4.4 Data Analysis, Research Objectives, and Outcomes

Among the 44 articles, there is a fairly skewed distribution of the data analysis methodologies toward use of quantitative methods. Twelve studies used mixed methods, 10 used qualitative methods, and 22 used quantitative methods. Among the studies using quantitative methods to analyze the data, 11 used inferential statistics, 10 used predictive modeling, and one study used both the inferential statistics and predictive modeling (Yiannakakis et al., 2008). In other words, 34 out of 44 studies used at least one quantitative data analysis method (including mixed methods). This is expected because sensing technologies provide an ample amount of numerical data that can be analyzed quantitatively to accomplish the research objective.

In terms of the primary research objectives of the contributions, we observe four prominent themes emerging from the selected papers:

1. Engagement of children (n=9): In this category, the dependent variable of the contribution was the engagement of the children. It was either measured/predicted using the sensor data, or the sensor technology was used to intervene to improve the engagement levels of the children while interacting with the specific technology.
2. Recognition/prediction of special needs/behavior (n=11): In this category, the main focus was either on differentiating between TD and SEN children, or on recognizing behavior indicative of a special need.
3. Explain/understand the behavior/attitude (n=17): This category of papers focused mainly on explaining, understanding, or predicting different behaviors (affective, cognitive, immersive) or on explaining different attitudes (sympathy, deception, specific attitudes toward the interactive system).
4. Learning performance/experiences (n=7): These papers focused on explaining and/or predicting participants' learning task performance or experiences while interacting with the system.

We report the research outcomes and insights of these four themes in the following subsections.

4.4.1 Engagement of children

Among the studies that are concerned with children's engagement, different aspects and/or points-of-view of engagement are examined or explored. For example, Bian et al. (2016, 2019) and De Wet and Potgieter (2019) used the "flow theoretic" view of engagement (Csikszentmihalyi et al., 2018). In the studies by Sonne et al. (2017), Morrison et al. (2015), De Wet and Potgieter (2019), and Castellano et al. (2014), engagement was investigated in a context where a new technology was presented to the children (e.g., games, robots, and tactile interaction). Furthermore, Yannakakis et al. (2008) studied social engagement in a more naturalistic setting such as a playground, while Javed et al. (2019) studied social engagement in a robot-mediated interaction setting. Table 1 presents the contextual details of these papers (the data collected, the goal of the study, and the age group targeted).

Table 1. Details of the studies where children's engagement was the main theme

Reference	Data	Goal	Age group
Sonne et al. (2017)	Respiratory patterns	To keep children engaged with a game while the doctors are taking blood samples from them	7–12
Betancourt et al. (2017)	EDA	Engagement with the biosensors themselves, measured by the desensitization period	2–11
Bian et al. (2019)	PPG, GSR, and respiration	Engagement with a driving task	14–17
Castellano et al. (2014)	Facial expressions	Evaluation of a multilayered engagement detector	8–9
Morrison et al. (2015)	Heartbeat and breath rate	Describe participant experience and engagement (enjoyment and	12–19

		pleasure) with the new tactile interactive wall	
Javed et al. (2019)	Facial data	Identify key design features that can improve social engagement in children	4–12
Bian et al. (2016)	ECG, ICG, PPG, PCG, EDA, GSR, EMG	Engagement with a driving task	Not mentioned
Yannakakis et al. (2008)	ECG	Estimation of the degree to which games provided by the playground engage the players	8–10
De Wet and Potgieter (2019)	EEG	Excitement, engagement, and frustration while manipulating a robot with a brain–computer interface	13–19

The studies used physiological data both as a measurement and as an interaction modality. For example, in a biofeedback game, children easily got engaged with the game through controlled breathing even during the longer and more complicated procedures (Sonne et al., 2017). In an experiment with a virtual driving adaptive platform, Bian et al. (2019) showed that the engagement-sensitive group had a statistically significantly higher engagement than participants in the performance-sensitive group, and that the physiology-based data-driven adaptive mechanism may be more effective at keeping the user in the "flow state" when compared to the performance-based data-driven adaptive mechanism. Betancourt et al. (2017) reported that younger children (aged 2–4) required a period of engagement with using the biosensors before wearing them. They further reported that, despite the initial reluctance of many younger participants, most of the children started wearing the sensors consistently in subsequent sessions. In another study that examined engagement with tactile interaction, Morrison et al. (2015) reported that participants found the experience engaging and understood what was required of them.

To understand the key design features required to improve children's social engagement, Javed et al. (2019) found that touch-based interaction was more engaging than gaze-based interaction for children with ASD. Further, in an attempt to predict children's engagement with a robot using physiological data, Castellano et al. (2014) used features such as a smile detector, a smile feature, manually annotated gaze features (look at robot high/medium/not), game state, game evolution, and captured pieces. They further reported F1-scores of 92% and 83% for high/low and high/medium/low engagement levels. In a similar setting, Bian et al. (2016) reported an accuracy of 78% for predicting children's engagement levels. Moreover, heart-rate-based features (average, maximum, range, and approximate entropy) have been positively correlated with children's engagement in the playground (Yiannakakis et al., 2008).

To summarize, sensing technology has been used either for engaging children (e.g., using sensor data as an interaction modality that engages children) or for quantifying children's engagement (e.g., using sensing data to accurately estimate and predict children's engagement). Either of the two approaches seems to support CCI research and provides valid evidence that would otherwise be difficult to obtain with mainstream interaction modalities and measurements.

4.4.2 Recognition/prediction of special needs/behavior

Table 2 presents the contextual details of these papers (the data collected, the goal of the study, and the age group targeted).

Table 2. Details of the studies where recognizing/predicting special behavior of the children was the main theme

Reference	Data	Goal	Age group
Jiang et al. (2020)	Motion	Achieve a comprehensive coverage of all ADHD-related symptoms in DSM-5	7–13
Goodwin et al. (2016)	BVP, IBI, EDA, acceleration	Predicting potentially dangerous aggressive behavior toward others in children with ASD	6–17
Kuzminykh and Lank (2019)	GPS tracking, activity, and sleeping patterns	Understanding information needs of parents with children with ASD	Not mentioned
Melo et al. (2019)	Speech, ECG, movement	Understanding effect of robot-mediated therapeutic activities involving children with ASD	3–6
Lundberg et al. (1993)	Systolic and diastolic blood pressure, heart rate, and neuroendocrine activity	Comparing physiological data when the children are at daycare and while they are at home	3–6
Bekele et al. (2013)	Gaze, ECG, PPG, SKT, GSR, EMG, and respiration	Difference between ASD and TD children when they perform a facial expression detection task	13–17
Saadatzi et al. (2013)	ECG, PPG, SKT, GSR	Examination of affective and physiological variation among children with ASD, in response to manipulated social parameters	13–19
Bian et al. (2013)	ECG, ICG, PPG, PCG, EDA, GSR, EMG	Develop a driving simulator using VR technology capable of flexibly responding to subtle affective changes in teenagers with ASD	Not mentioned
Sonne and Grønbæk (2015)	Physical activity	Investigate the potential of using a wearable-sensor system to provide in situ assistance to children with ADHD in regaining attention in school contexts	Not mentioned
Mohammad and Nishida (2010)	GSR, respiration, and BVP	Distinguishing natural and unnatural partner behavior in a close encounter situation	Not mentioned
Wade et al. (2016)	EEG, physiological	Evaluate a gaze-contingent driving intervention system with drivers with ASD	13–18

These studies address the behavior and needs of children with special needs. These efforts can be further divided into three sub-categories: (1) finding behavioral patterns that are peculiar to a given special need; (2) finding the behavioral differences between TD and SEN children; and (3) detecting special needs for children.

Finding behavioral patterns that are peculiar to a given special need: Jiang et al. (2020) showed that there are significant differences in the motion data of children with ADHD and TD children. They also demonstrated that these differences appear even in tasks that do not require a lot of motion, such as reading, and the differences are more pronounced as the requirement of motion-based actions increase for the completion of a given task such as limb reaction. In another study on robot interaction, Bekele et al. (2013) revealed differences between the gaze patterns of ASD and TD children. ASD children focused on the forehead, whereas TD children focused on the eyes of the virtual character. Using physiological data, Bekele et al. (2013) further found a 94% accuracy on ASD versus TD classification. Saadatzi et al. (2013) showed that children with ASD had higher anxiety levels while reading social media texts than TD children.

Finding the behavioral differences between TD and SEN children: Goodwin et al. (2016) combined physical (movement) and physiological (HRV, BVP, EDA) features to predict the onset of aggressive behavior in children with ASD. Melo et al. (2019) showed that children with ASD have a higher heart rate during an obstacle completion task, especially when a child is struggling to complete a task (such as removing the obstacle). Moreover, the children with ASD also reported seeing the movement of the robot to be unnatural when it was controlled by a human (wizard) than when it was automatic (Melo et al., 2019). Bian et al. (2013) developed a driving simulator using VR technology to flexibly respond to subtle affective changes in teenagers with ASD. The system was highly accepted among the children with ASD (Bian et al., 2013).

Detecting special needs of children: Lundberg et al. (1993) compared physiological data for when children stayed at home with when they were in daycare. Compared to levels reported at home, daycare was associated with an increased heart rate, epinephrine, and norepinephrine excretion, and with decreased cortisol levels. The authors concluded that the daycare centers present more challenges to the children than activities at home, such as interaction with other children, adults, and participation in various activities (Lundberg et al., 1993). In a similar vein, Sonne and Grønbæk (2015) developed a system for using wearable-sensor data to lower the hyperactivity levels of children with ADHD so that they can re-engage with school activities.

In summary, all three categories used sensing data either to explore or to identify a particularity that the human eye would be unable to capture, detect as a pattern, or use comparatively. To detect a pattern and perform a comparison, the studies employed prediction and classification techniques that are now part of the common methodological toolkit of a CCI researcher. It is likely that in the coming years there will be an increase in the use of these techniques due to the rise of both sensing technologies and AI and ML.

4.4.3 Explain/understand the behavior/attitude

The contributions under the theme of explaining/understanding the behavior and/or attitudes of the children are primarily concerned with various aspects of children's cognitive, affective, and attention-related behavioral patterns. For example, cognitive load, affective states (frustration, boredom, stress, deception), emotional states (happy, sad, angry), attention while performing a specific task, and joint attention in cooperative/collaborative tasks. In most of these papers, the sensor data was used to infer/detect these patterns, and the papers under this theme did not necessarily have a "learning objective" (unlike those described in Section 4.4.4). Table 3 presents the contextual details of these papers (the data collected, the goal of the study, and the age group targeted).

Table 3. Details of the studies where explaining/understanding the behavior of typically developing children was the main theme

Reference	Data	Goal	Age group

Lopes et al. (2018)	Speech	Understanding symptoms of high cognitive load	Not mentioned
Dinet and Kitajima (2018)	Motion	Effect of immersive environment on (dangerous) behavior	Mean age 9.5
Xu et al. (2012)	Physical activity	Understand the attitude toward sustainability, adaptability, and sociability goals of the activity-based games	Not mentioned
Pike et al. (2016)	EEG	Attention while watching films	Not mentioned
Okita et al. (2011)	Facial data	Examine the different features of humanoid robots and the influence on children's affective behavior	4–10
Sridhar et al. (2018)	GSR, HRV	Understand cognitive-affective states while performing cognitive tasks	4–7
Christensen and Biskjaer (2018)	HR	Understand children's playing behavior	8–9
Jyoti and Lahiri (2020)	PPG, GSR, respiration	Understand the effect of joint attention cues	5–8
Sridhar et al. (2019)	GSR, HRV	Understand cognitive-affective states while performing cognitive tasks	4–7
Zhang et al. (2014)	PPG, GSR, gaze	Affective states	13–17
Jeon et al. (2019)	Gaze	Find a relationship between eye movement and degree of sympathetic behavior	Not mentioned
Bekele et al. (2014)	Gaze, ECG, PPG, SKT, GSR, EMG, and respiration	Model the context-relevant psychological state	Not mentioned
Silva et al. (2013)	BVP, EDA	Understanding deceptive behavior during multimedia interaction	Not mentioned
Zhang et al. (2015)	EEG, ECG	Measure cognitive load	13–17
Woolf et al. (2009)	Posture, movement, grip, and emotions	Emotional states	Not mentioned

This group of contributions used physiological data streams to explain the behavior during and/or attitudes toward the interaction with the technology. One prominent facet of this category of papers is explaining/predicting cognitive load using physiological data. For example, Zhang et al. (2015) reported an accuracy of 81.75% for cognitive load (expert label) prediction using a fusion of physiological (individual stream accuracy: 79.29%), EEG (individual stream accuracy: 80.58%), and gaze (individual stream accuracy: 71.92%) features. Similarly, Jeon et al. (2019) predicted children's cognitive load while reading by using gaze features, such as frequency of repeated visits to specific important words and phrases. Lopes et al. (2018) used physiological data to detect cognitive load and employed a dialogue-based agent to mitigate the adverse effects of high cognitive load. The results show that the entropy of system prompts in high-load condition was significantly higher, and that there was also significantly higher skewness for intensity in the low cognitive load.

Another key aspect of this category of research outcomes is prediction/analysis of affective state. For example, Zhang et al. (2014) used pupil diameter, blink rate, and physiological data to predict four affective states (engagement, enjoyment, anxiety, and boredom) using a decision tree. The authors reported a very high accuracy of 83.09% (binary classification). Similarly, Bekele et al. (2014) used gaze, ECG, PPG, ST, GSR, EMG, and RSP to predict children's affective states. The F1-score for predicting physiological affective states (expert labeling) with four classes was reported to be as high as 0.96.

This group of papers shifts away from the traditional multidisciplinary nature of CCI and adopts an ML stance, with the main goal of predicting various affective states of the children. This stream is closely associated with the affective computing community, and it seems able to provide various insights in traditional CCI research.

4.4.4 Learning performance/experiences

This theme encompasses the contributions that have "explicitly" collected children's responses about their performance either through a pretest and a post-test, or because the experimental task had an explicit performance measurement (e.g., shape placement performance: Johnson and Picard, 2017; driving task: Wade et al., 2014, 2016). This theme also includes those contributions that predict/analyze the game scores in a game-based learning setting (An et al., 2018; Stone et al., 2014). The final set of contributions explicitly ask children to report their experiences, such as excitement (Clegg et al., 2017) and anxiety (Wade et al., 2014), within a specific learning setting (e.g., simulations, inquiry-based learning). All the contributions in this theme had a specific "learning objective", as opposed to those discussed in Section 4.4.3. Table 4 presents the contextual details of these papers (the data collected, the goal of the study, and the age group targeted).

Table 4. Details of the studies where explaining/understanding the learning performance and/or learning experiences of children was the main theme

Reference	Data	Goal	Age group
Johnson and Picard (2017)	EDA	Shape placement performance	2–5
Wade et al. (2016)	EEG, eye tracking, EMG, ECG, GSR, PPG, ST, and respiration	Performance in driving tests	13–18
Clegg et al. (2017)	Heart rate, breathing rate, and movement	STEM learning experiences for children in inquiry learning scenario	6–11
An et al. (2018)	Gaze	Game score	18–19
Stone et al. (2014)	ECG, EEG	Game score	Not mentioned
Wade et al. (2014)	ECG, EMG, respiratory patterns, SKT, PPG, GSR, eye tracking	Learning experience (anxiety) in driving simulator	13–17
Radeta et al. (2017)	Skin conductance	Learning gains in a game-based system	9–10

These contributions utilized the physiological data not only to differentiate between good and poor performers but also to provide feedback. For example, Johnson and Picard (2017) used a feedback system to improve shape recognition, with their data showing a clear improvement in shape placement over time but only with the motivating sensory feedback. This indicates that feedback is a necessary feature. In a live physiological sensing and visualization (LPSV) system, Clegg et al. (2017) showed that allowing for incremental integration of new variables and life-relevant components in inquiry experiences for young learners helped them focus on aspects relevant to the inquiry. The authors also showed that the data from LPSV can help educators mitigate the range of sensitive discussions, physical activity, and noise levels.

Using gaze to differentiate between good and poor performers in a game-based learning setting, An et al. (2018) concluded that participants generally spent more time fixating on selected options than on any other option, that the participants spent more time fixating on the correct options, and that participants who exhibited higher fixation duration percentages on their selected answers scored higher. In another game-based learning environment, Stone et al. (2014) showed that tutees had higher HR and higher engagement than tutors. However, tutor–tutee synchrony was consistently responsible for most of the variance explained in the game score.

Physiological data has also been used to explain performance in driving simulators (Wade et al., 2015, 2016). In an experiment in which the participants were rated not only on how they performed on the driving tasks but also on whether they looked at the salient features of the environment, the participants showed improvement in success in both scanning the driving environment and in performing the task (Wade et al., 2015). This was not the case in another condition where the participants were rated only on their driving performance (Wade et al., 2016).

This last group of papers focuses on the intended goal (performance). The papers used sensing technology to explain performance and differentiate groups with different needs (e.g., low vs. high performers), and some went one step further to utilize the information collected via sensing to provide feedback. Therefore, this category of papers utilized sensing technology to create a complete cycle of interaction with and support for children.

5. Discussion

After selecting and coding the 44 contributions in this SLR, we observed certain trends that are clear and common among them, based on which technologies were used (**Research Question 1.1**), what data was captured (**Research Question 1.2**), and what insights emerged (**Research Question 1.3**). For example, most of the papers had a small sample size (fewer than 30 children, irrespective of the number of conditions); most studies focused on teenagers and schoolchildren; and among the papers that included populations with special needs, most of the children had ASD. Furthermore, most of these studies were carried out in laboratories or schools (33 out of 44), and most were based on games, problem solving, or virtual driving simulators (24 out of 44). The most common method to measure performance was task-based performance (17 out of 22 that measured performance). Finally, the four major themes for the research objectives were: (1) engagement of children; (2) recognition/prediction of special needs/behavior; (3) explaining/understanding the behavior/attitude; and (4) learning performance/experience. Several implications emerged from this literature review, which we discuss next.

5.1 Implications for Theory and Practice: Challenges and Opportunities

In this section, we discuss the challenges and opportunities for research that uses physiological sensing in the context of CCI (**Research Question 2**). We divide the challenges and opportunities into four groups: (1) populations and data quality; (2) usability and effectiveness; (3) new sensors and techniques; and (4) ethical and social concerns.

5.1.1 Populations and data quality

In Section 4.1, we showed that most of the studies had a low to moderate effect size, with 28 out of 44 contributions having fewer than 30 children participating in the study. A majority of studies involving children with special needs (e.g., ASD, ADHD) reported that these special needs covered a diverse set of requirements for the individual children. Participating children lived with a wide range of conditions; therefore, such studies provide insights from a diverse set of children (Bian et al., 2019; Sonne and Grønbæk, 2015; Sonne et al., 2017). However, the generalizability and the internal validity of these studies are always at risk. Although this is not necessarily the goal for many of these papers (e.g., qualitative, exploratory works), it is advisable that future studies consider the confounds in their population when analyzing sensor data from a low sample size with diverse special needs. This will allow us to achieve some level of generalizability and to extract models that can be used in different populations and create products that can support end users outside the experimental settings.

Another risk of having a wide population is that their ages, genders, and special needs conditions (among other factors) impact the data collection (see Section 4.1). In particular, this is the case if the data are collected using physiological sensors such as, EEG, EDA, heart rate, pupil dilation, and temperature. Consequently, studies offer only superficial assessments of measurements emerging from these data sources and/or they had to discard a high percentage of the data. Betancourt et al. (2017) provided a helpful set of guidelines that could improve both the support for the claims and the data-processing steps. These guidelines include various methods for normalizing, de-noising, and synchronizing the different sensor data.

Another challenge with respect to the participating population emerges when the studies include both typically developing children and children with special needs. In our SLR, we have found seven such studies (see Section 4.1). Most of these contributions had the research objective to recognize the special needs and/or classify the children with special needs (Bekele et al., 2013; Goodwin et al., 2016; Jiang et al., 2020). In such cases, the challenge is to keep the typically developing children engaged at their own levels and not to have the mirroring effect from the children with special needs. For example, if the typically developing children stop interacting with the technology, this should not hinder the interaction of the typically developing children. For this challenge to be overcome, there is a need for further testing of the extracted models and classifiers with increased populations and different settings. The sample sizes and data quality needed for this endeavor cannot easily be obtained in typical CCI studies, so large research consortia and data-sharing techniques might be needed to overcome this challenge.

5.1.2 Usability and effectiveness

Involving children as participants (this SLR includes only studies where children are participants, and not research partners or in other roles) presents certain technical and practical challenges while conducting the studies. A few studies reported certain unexpected behavior from children (Javed et al., 2019; Sonne et al., 2017), which could have added a new confounding variable to the experiment. Moreover, there can be certain delays and unforeseen factors when including children in the studies (Morrison et al., 2010). Such cases can have a high impact on the data quality and reproducibility of the results, and they highlight the importance and necessity of qualitative studies and the multidisciplinary research community that accounts for such high ecology as well as occasionally low data-quality studies. CCI's nature (Giannakos et al., 2020) allows us to bring together various disciplines (e.g., design, computer science, learning sciences) and methodological traditions (e.g., from the social sciences and engineering) that help develop a clear epistemological position and further our knowledge landscape and horizons by embracing the particularities of our discipline.

A common factor in the design of these studies is that most of them had one or two sessions per child. This could mean that the experiment remained within the novelty zone of interaction and that the engagement was short lived. Although these studies highlight the salient and enjoyable aspects of the technology and interaction, it would be more informative for longitudinal skills development to conduct studies over a longer period of time. However, conducting longitudinal studies that utilize sensing technology is not always feasible. Another way to increase the engagement with the short-term studies is to allow participants to control their own timing of interaction with the system. In addition, the initial time spent with the technology can be increased to reduce the novelty effect. In this way, children can build proprietorship from their speculative interaction and experimentation (Morrison et al., 2010). This challenge can also be addressed both by engaging with research traditions and methods that allow the researcher to employ longitudinal studies and by utilizing sensing technology via mobile and ubiquitous devices for in situ self-reports and systematic sensing data collection (e.g., the mobile-assisted experience sampling method: Van Berkel et al., 2017).

Another facet of this challenge concerns wearable devices. Most of the off-the-shelf sensing equipment are in the form of wearables (EEG caps, eye-tracking glasses, and wristbands to capture EDA, HR, BVP, and temperature), as shown in Section 4.3. This adds an extra level of complexity, especially when children are involved. The effectiveness of the wearable devices is limited by how well the device is attached to the user. Moreover, the convenience of wearing the device has a significant impact on the practical usage of the device (Bian et al., 2019; Jiang et al., 2020). This convenience needs to be realized by optimizing not only the hardware devices but also the number and wearing position of sensing devices, which would achieve non-intrusive perception (see Section 4.4.2). In addition, the use of wearables might hinder children's participation, since the child or parent might not want to wear something that is not socially acceptable.

Furthermore, the choice of the wearable and the data collected should depend on the age group of the children included in the experiment (Hedman, 2014). This is further illustrated in Figure 6, which shows the relation between the age groups and the data collected in the studies included in this SLR. For example, when electrodes are placed on fingers, and when an individual moves their knuckles, it can produce artifacts. This can also be extended to reflect the dependency of the sensing technology used and the other factors defining the context of a study, such as the task type and the location of the experiment (e.g., school, museum, or laboratory).

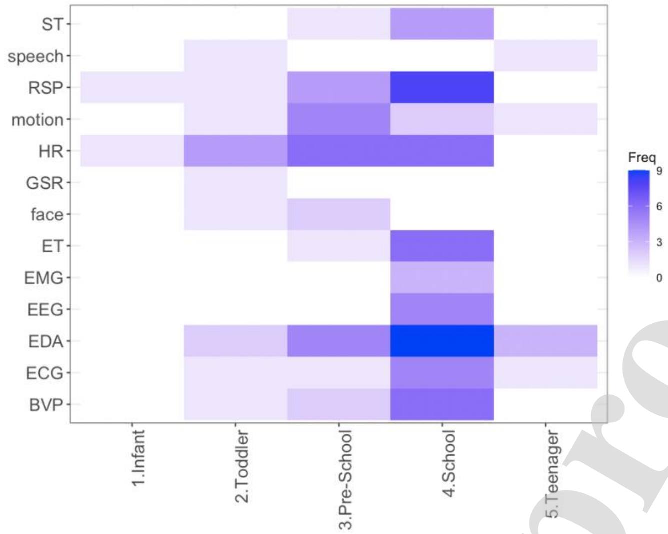

Figure 6. Mapping between the age groups and the data collected in the contributions

Finally, one of the most important factors in the practical design of the studies is the ecology of the study and the potential inherent biases of wearables (e.g., children's cognitive load might increase when wearing equipment) (Almjally et al., 2020; Christensen and Biskjaer, 2018; Papavlasopoulou et al., 2019). Experimental activities for children can induce biases such as increasing the cognitive load, which might affect the children's performance and general experience with the tasks (see Section 4.4.3). Proper organization and integration of the tasks, activities, and experimental materials, with a coherent representation of the related technology and instruction in how to use it, are suggested in order to prevent any unnecessary streams of information and cognitive overload.

5.1.3 New sensors and techniques

The most prominent opportunity in the field of sensing technologies and CCI is the set of new sensors and data-processing/analytical techniques. These sensors and technologies are not novel per se, but they are seldom used in the context of CCI. Here, we discuss a few recommendations emerging from this SLR.

An under-utilized category of sensors in the contributions included in this SLR is that of proximity-related sensors. As Section 4.3 makes clear, none of the studies used such sensors. Using a close-range (up to a meter) proximity detector can open new avenues of information, especially in longitudinal in situ studies that take place in museums, schools, or other open environments, rather than in a laboratory. For example, it is possible to record when children approach a certain technology or exhibit in a museum and to combine that information with the velocity with which they engage. In such cases, if children approach but do not engage, it may indicate that the task is too challenging or the feedback is not enticing enough. Such insights might help researchers design better methods of communicating and providing feedback to the children. In addition, such sensors can allow us to investigate group dynamics and children's behavior in collaborative tasks. Such data are also ecological (children do not wear anything) and do not breach children's privacy (no personal data of the child is captured).

Another important sensor and technology combination is speech recognition. However, despite its advantages (Yuan et al., 2019), it is seldom used in the context of CCI studies (only two studies used speech, as mentioned in Section 4.3). This technology might be most beneficial for helping teenagers engage with the technology. Another advantage of using speech recognition is that it can help anticipate the children's responses or changes in engagement levels, enabling the system to be triggered to produce more engaging and enjoyable interactions. It might be crucial for the technology to have early indications of whether the user is confused, engaged, afraid, or bored. With speech recognition, the system can then decide to dynamically alter/modify the communication and information presentation style, while simultaneously attempting to fulfill the task goals. However, for this technology to be used, several challenges need to be addressed. Proper speech development age (Markopoulos et al., 2021), effective discourse formulation (Yuan et al., 2019), and the use of different languages are some of the most common obstacles. However, it is important to take care of the testing granularity of such systems so that these speech-based technologies are as close to human parity as possible (Schuller and Schuller, 2020). When using advanced techniques, researchers should also take care that these techniques do not become difficult to explain from a human-centric point of view (Schuller and Schuller, 2020).

Some recommendations can also be made for the sensor data-processing/analyzing techniques that are making their way into related fields, such as HCI, affective computing, learning analytics, and educational data mining. These techniques are not often used in CCI when sensing technologies are involved. The first such technique involves extracting the fine-grained features that can better reflect the interactive process. Most of the papers in this SLR used aggregated measurements of the sensor data to analyze the interaction between children and the task/technology (see Section 4.4). Aggregated data provide the researcher with a certain decision-making capability (based on predictions or inference), but such data do not completely represent the whole interaction. The temporality of sensor data, such as EDA, heart rate, and EEG, provides a larger amount of information about engagement, stress, attention, affect, and cognitive processes (Giannakos et al., 2020a, 2020b; Sharma et al., 2019, 2020a) than is provided by aggregate measurements. Therefore, it is advisable to use the temporal information from the collected sensor data in order to leverage the inherent benefits of sensing technology (e.g., temporality, direct access to insights into a child's cognitive and affective processes).

Furthermore, physiological sensor data provides a unique opportunity for CCI researchers to apply state-of-the-art ML techniques, since the data have different granularities and are sufficient for properly training the algorithms. This SLR contains some examples of ML being used to predict the performance/engagement of children while interacting with the technology. However, most of the contributions used basic ML algorithms. For example, Goodwin et al. (2016) used a logistic regression classifier to predict the aggressive nature of the children (see Section 4.4.2). Wade et al. (2014) used support vector machines to classify children's performance and affective states (see Sections 4.4.3 and 4.4.4). Saadatzi et al. (2013) also used support vector machines to classify children's anxiety levels (see Section 4.4.2). It can be noted that the methods used in these papers are classical and provide good prediction quality, but they do not exploit the full potential of the data gathered from the physiological sensors.

Moreover, most of the methods used in the studies either used the aggregated version of the physiological data across the dependent variable without using time as a factor, or they employed discrete classes/clusters of behavior when temporal analysis was conducted. Both these methods have their limitations. For example, clearly aggregating the data at the dependent variable level does not tell us anything about children's behavior trajectories. On the other hand, using discrete classes/clusters of behavior does not produce a holistic portrayal. In related fields such as learning analytics and educational data mining, there are examples of using deeplearning methods with temporal data to provide better predictability than is possible with clustering or aggregation (Prieto et al., 2018; Olsen et al., 2020; Sharma et al., 2020b). These methods could be further exploited to further our understanding of children's interaction with the technologies.

Finally, the last set of recommended techniques involve the adaptation and personalization of the interactive experience for the children. By incorporating state-of-the-art adaptation and personalization techniques (Korhonen et al., 2020; Oh and Kang, 2020; Sharma et al., 2020c), CCI researchers can make the interaction more engaging and effective for the children. Only a few papers in this SLR mentioned the need for personalization and adaptation (Dinet and Kitajima, 2018; Hauser et al., 2002; Jiang et al., 2020). Higher levels of personalization can be achieved through effective feature selection and robust models for recognizing human cognitive states (see Section 4.4.3). Currently in CCI, each participant needs to complete the predefined set of tasks to make accurate predictions. However, there might be redundancy among these experimental tasks. Therefore, the systems should have a dynamic component to adjust the task number, task order, and difficulty based on the real-time responses and sensor data of the subjects. For example, a heuristic task selection method can be used. This heuristic could use the results of each subject under the current task and the current cognitive-affective state, in order to recommend the next task until certain diagnosis results are obtained.

Concerning adaptation in CCI, there is an increasing consensus that implementing the tasks in an adaptive manner (also known as skill training) might significantly improve long-term support for children with special needs (Bradshaw et al., 2019; Pathak et al., 2019). Due to the alarmingly increasing prevalence of disorders and the lack of trained therapists, technology-based assistive ASD intervention has gained momentum in recent years (for a review, see Jaliaawala and Khan, 2020). We recommend such protocols be followed in the task selection and adaptation for studies both with children with special needs and with typically developing children (adaptive systems with typically developing children were also found to be more engaging than the non-adaptive systems; Almasri et al., 2019).

5.1.4 Ethical and social considerations

Ethical constraints are even more important when children are involved than for any other participant population. CCI researchers have always been cautious about privacy and ethical concerns (Dowthwaite et al., 2020; Kawas et al., 2020). Van Mechelen et al. (2020) systematically presented the ethical constraints in CCI research. In this section, we focus on recommendations relating to the sensing technologies and the support sphere of the children (parents, caregivers, teachers).

With regard to privacy, data such as camera images and recognized facial expressions/emotions, as well as their usage within interactive applications, play an essential role. The contextual use of these technologies can provoke disengagement in the children. These affective technologies (Bekele et al., 2013; Javed et al., 2019; Okita et al., 2011) utilize facial expression and emotions as a key factor in their protocols (see Section 4.4.3). Such technology can have different effects on the children's perception of their privacy if the interaction is taking place in different settings, such as in school, outdoors, or at home. We recommend that the children are informed about the usage of their data so that such contextual biases can be mitigated. In terms of using other sensitive sensor data (e.g., heart rates, GPS locations, blood pressure), we propose two different use-cases to handle the data appropriately: a short-term use, where the sensitive data is used to drive the momentary adaptation as a part of interaction (e.g., task completion, enjoyment); and a long-term use with proper anonymization processes to create more long-term effectiveness routines (e.g., skill improvement, longitudinal engagement). These long-term routines can be properly vetted by the parents and teachers, depending on the context of the study.

Concerning the social aspects of the CCI studies using the sensing technologies, there are certain roles that the teachers, parents, or caregivers (in the case of children with special needs) can play (see Section 4.4.1). These roles will not only allow the studies to be conducted in a smooth fashion but might also increase children's acceptance of and engagement with the sensing technologies. For example, for collaborative or cooperative technologies, the feedback can only be triggered by those children who are using the sensors, since it is not necessary for all the children in a group to wear a wristband or eye-tracking glasses. In such cases, only the player being tracked can trigger the feedback, which might require further cooperation and/or social exchanges between children. In a similar manner, a teacher or caregiver could handle the prompting (for more cooperation/social exchange), further extending the social dynamic.

Furthermore, for activities where the children are creating artifacts, adults have even broader roles than in the previous example. For such scenarios, Papavlasopoulou et al. (2019) proposed that teachers and assistants be more involved than just providing instructions and giving occasional help. The authors called for more "honest teaching relationships" with the children. This might impact the acceptance of the wearable sensors by the children, as they will be more engaged in the task (and forget about the sensors) when there is more involvement from the "adults in the room". The teachers and caregivers can also help the children not only to understand the requirements of the task but also to make them comfortable with the wearables and other sensors present in the experimental setting (e.g., cameras and motion sensors).

Finally, with regard to children with special needs such as ASD and ADHD, an overwhelming proportion of such children often fail to achieve conventional independence as adults in terms of behavioral markers (Shattuck et al., 2012). Additionally, traditional intervention approaches might not be sufficient for creating opportunities for addressing these skills and deficits within and across naturalistic settings in appropriately intensive sessions (Goodwin, 2008). Adding sensing technology might elevate these impairments (e.g., by participating in new environments and processing information about the new wearables, while at the same time learning skills related to functional independence). In such cases, the support sphere (parents, caregivers, and teachers) can provide much-needed help in conducting the studies in a seamless manner. Kuzminykh and Lank (2019) argued that people in the support sphere can act as secondary users of the sensor data and help the children understand the novelty in their environment in a simplified manner.

It should also be acknowledged that utilizing sensing technology in CCI research and practice can lead to significant concerns among stakeholders about the invasion of privacy through various modalities of data collected during the interaction between children and computational and communication technologies. There might also be concerns about the potential bias built into the computational modeling approaches used in the analysis and interpretation of sensor data. Another common concern is associated with the potential "surveillance" culture encouraged by the use of sensing technologies in CCI. Such concerns are also associated with different forms of mainstream data collected in HCI (e.g., clickstreams, keystrokes), and they are valid and of particular importance for the vulnerable end-user group of children. As CCI researchers, we can neither control nor predict future technological advances, but we nevertheless have a moral responsibility to be aware of and reflect on possible implications of the research we conduct, the technologies we envision, and the impacts we may facilitate in children's lives (Antle et al., 2021). This is why future research needs to consider the ethical and practical issues relating to the use of sensing technology in CCI research and practice (Kawas et al., 2020). When making a real-world impact on children's lives, it is essential to respect values and ethics.

5.2 Challenges and Opportunities of Sensing Technologies in CCI Research and Practice

5.2.1 Risks and challenges of using sensing technologies with children

Apart from the aforementioned ethical and social considerations, there are risks and challenges of using sensing data with children. First, as Section 4.4 reveals, most studies are quantitative in nature, and there was little attempt to include qualitative data from the studies to improve the insights by triangulating methods. When qualitative data are lacking or has less impact on the final outcomes, certain types of insights (e.g., those that are either non-quantifiable or require different data collection technique) might be missed. Therefore, it is important for sensing technologies to be employed to serve well-informed purposes within well-justified lines of research (e.g., eye tracking to investigate children's reading behavior), and, if possible, for sensor data to be triangulated with additional data collections in order to strengthen its richness and interpretation (e.g., think-aloud data or interviews to identify the reasons behind certain behaviors).

The second challenge of such methods of investigation is rooted in an over-reliance on the "ground truth" labels that act as the training for the predictive/explanatory models. Like the first risk, reliance on the ground truth risks a shallow understanding of the theoretical concepts. For example, cognitive load has been meticulously defined as having intrinsic, extraneous, and germane components (Sweller, 2011), but when it comes to the studies measuring cognitive load (e.g., Lopes et al., 2018) the sensor data was used to measure a superficial proxy of cognitive load without much information about which actual component of cognitive load was the focus of the contribution. Therefore, researchers need to consider the over-simplification of complex constructs, imperfect quantitative models, and, occasionally, potential prioritization of dimensions that can be quantified using sensor data at the expense of equally or more important dimensions (which cannot yet be quantified).

Another major challenge is the lack of contextual information input by the data-processing and analysis algorithms. The studies using sensor data inherently employed context-independent measurements to understand a construct relating to children's affective, cognitive, and behavioral processes. It is difficult to gain a deeper understanding of children's interaction with technology without some information in the predictive/explanatory models. For example, in a context of robot-mediated communications (e.g., Javed et al., 2019; Melo et al., 2019; Okita et al., 2011), the constructs, such as engagement with the robot and joint attention with a peer (in the case of a collaborative setting), were studied using sensor data without much attention paid to the contextual information. However, engagement and joint attention are constructs that can vary a great deal depending on the (rich) contextual information from the setting under investigation, and this information sometimes seems to be missing from the models used in the selected contributions.

With real-time low-level data accessibility, there is a high probability of an increase in “micro-management” in CCI research. For example, EDA and HRV-related measurements of engagement used (e.g., Bian et al., 2016; Sonne et al., 2017) in a game/simulation can lead to short-term adaptation in the interaction without considering the long-term impact of those changes. Such efforts could also result in over-prioritization of dimensions that can be quantified using sensor data over other dimensions that cannot be easily quantified. For example, learning performance (see Section 4.4.4) has been typically measured with a low-level post-test, covering mostly procedural and declarative knowledge, whereas higher-level learning constructs, such as analysis and synthesis, are generally untouched when sensing technologies are used in CCI research.

Based on all the aforementioned potential challenges and risks of using sensor technologies, future research in CCI that uses sensing technologies should take a cautious approach to the intersecting domain of investigation. On the one hand, sensor data provides researchers with an opportunity of real-time, automatic proxies of the affective, cognitive, and behavioral dimensions of CCI; however, without both proper theoretical grounding and contextual information, it is possible to arrive at faulty conclusions. CCI researchers need to carefully consider the strengths and weaknesses of sensing technology and purposefully apply it, whether for investigating a construct that would otherwise be non-observable, for supporting children with special abilities, or for another well-motivated purpose. In addition, given the paramount importance of the contextual information in CCI research, we should leverage inherently contextual techniques, such as human-sensors (e.g., via observations, interviews) to complement and interpret insights coming from sensor data.

5.2.2 Opportunities of using sensing technologies with children

Sensing technology enables CCI researchers to employ passive and continuous captured data about children’s behaviors. Such sensor data can be used to enrich CCI research measurements (e.g., attention, cognitive load), to support technology’s functionalities (e.g., affective systems), and to propel an interaction modality between the system and the child (e.g., gesture, motion). Contemporary HCI research (Thieme et al., 2020) recognizes the value of sensing data coupled with ML approaches, and it emphasizes the importance of seeing the complementarity of these methods, instead of merely seeing them as competitors. Sensor-based insights generated during a child’s interaction with the technology have the potential to help us to identify children’s real needs (e.g., Castellano et al., 2014) and sometimes even support these needs (e.g., Kourakli et al., 2017).

Traditionally, CCI research has employed either qualitative methods or has used mainstream data collection (e.g., clickstreams, keystrokes, self-reports) (Markopoulos et al., 2021). Mainstream data collection focuses on children’s actions (e.g., pressing a button) and consciousness (e.g., their opinion about something) and ignores important processes such as information processing while reading a book or arousal levels during gameplay. Sensing technology offers the possibility of giving us access to these processes, allowing us to complement insights extracted from these otherwise “unseen” processes. Enriching our understanding about children’s experience during their interaction with computational and communication technologies allows CCI research to consider children’s otherwise unseen needs (e.g., moments of disengagement or negative affectivity). Complementing our research with sensor data can provide us with a more holistic picture that allows us to support childcenteredness (i.e., viewing the child as a protagonist; Iversen et al., 2017), and to design and develop technologies that are aware of and account for children's experience.

To be able to utilize sensing technology in CCI research, and to seize the potential opportunities, future work needs to tackle several challenges. The sensor devices, such as wearables (e.g., wristbands and glasses), need to embrace children as end users. For example, children's constant movements during their play need to be considered when designing these devices, which are mainly made for adults (e.g., in terms of their size, weight, tolerance). Moreover, the algorithms need to be considered and sometimes changed to address CCI research objectives. Nevertheless, the reviewed studies managed to use various contemporary sensing devices (e.g., wristbands and eye trackers) in CCI research. Therefore, despite the additional challenges mentioned in Section 5.2.1, future research and practice needs to consider the development of sensing technology that accounts for children's needs (e.g., in relation to data privacy, ethics, and size).

5.3 Limitations and Future Work

Along with its contributions, this work has some limitations. First, we had to make some methodological decisions (e.g., selection of databases, the search query) that might lead to certain biases in the results. However, we endeavored to avoid such biases by considering all the major databases and following Kitchenham and Charters's (2007) guiding steps.

Second, the selection of empirical studies and coding of the papers might create another possible bias. The focus was clearly on the empirical evidence, and the coding was performed by two independent researchers. However, we acknowledge that the authors of the contributions could have used completely different terminology in their published works (e.g., the name of the individual sensing device, such as eye tracking, EEG) and avoided using the term "physiological data". However, most of the studies used both the name of the particular sensing device (e.g., EEG) and the term "physiological data", so those studies would have been captured with our search strategy. In addition, the most relevant venues (e.g., IDC, IJCCI) were searched thoroughly.

Third, some elements of the papers were not described accurately, leading to some missing information in the coding of the papers. However, the amount of missing information was small and could not affect the results significantly.

Sensing technology offers promising functionalities to CCI research, both as an interaction modality and as a data collection mechanism. As our literature review has shown, sensing technology is becoming increasingly prevalent in CCI research due to its inherent benefits (e.g., it is automatic, pervasive, and yields temporal insights), as well as to its ability to be employed with and complement more traditional research methods. Despite the growing use of sensing technology in CCI and its potential, the CCI literature recognizes that thorough consideration of the ethical underpinnings is necessary. From a practical standpoint, future work should consider the preparation required for employing sensing technology in CCI studies. In addition, future work should consider ethical, social, and privacy concerns, identify how to communicate the information provided by this data collection to the children and parents, and allow them to provide an informed assent/consent. This is particularly important since most children and parents are new to some of the sensing technologies. Therefore, future research needs to provide frameworks for CCI researchers that allows them to plan and utilize sensing devices in their research, since describing the details in the consent form or running sensing-based studies in the same way that we implement traditional measurements is probably not sufficient. Rather, it is extremely important to engage in discussion with the children and parents and to explain the rationale and added value of such data collection.

6. Conclusions

We have presented an SLR of 44 contributions in the field of CCI and sensing technologies from recent years. We analyzed the papers from the perspective of the study design (learning context, environment, population, and so on) and the insights the research provides about the children's task-based performance/outcome, engagement, special needs, and behavior. We categorized the main findings of the selected papers in four thematic areas and discussed the challenges and opportunities emerging from the current review in terms of both the sensing technologies used and the impact they could have on our understanding of children's outcomes and behavior. Finally, based on the current state of the field, we have proposed four different strands of further possible advances.

In essence, sensing technology approaches can provide explicit and comprehensible ways to advance CCI research. Presenting sensor-based information to children, researchers, parents, and other stakeholders to make them more informed decision-makers offers great social benefits. This can contribute to the vision of Mike Eisenberg by offering a new form of "transhumanist technologies" (Eisenberg, 2017) that enable children, teachers, parents, and CCI researchers to perceive important insights that augment their capacities. Therefore, sensing technologies should be tightly coupled with researchers, respect children's values, and ultimately enhance children's various capabilities.

References

Almasri, A., Ahmed, A., Almasri, N., Yousef S. , Mahmoud, A. Y., Zaqout, I.S., Akkila, A. N., & Abu-Naser, S. S. (2019). Intelligent tutoring systems survey for the period 2000–2018.

Almjally, A., Howland, K., & Good, J. (2020, June). Investigating children's spontaneous gestures when programming using TUIs and GUIs. In Proceedings of the Interaction Design and Children Conference (pp. 36-48).

Alsuradi, H., Park, W., & Eid, M. (2020). EEG-based neurohaptics research: A literature review. *IEEE Access*, *8*, 49313–49328.

An, B., Kim, I., Pakdamanian, E., & Brown, D. E. (2018, December). Exploring gaze behavior to assess performance in digital game-based learning systems. In *2018 Winter Simulation Conference (WSC)* (pp. 2447–2458). IEEE.

Antle, N. A., Frauenberger, F., Landoni, M., & Fails M. A. (2021). Ethics in CCI [Special issue], Int. Journal of Child–Computer Interaction.

Baykal, G. E., Van Mechelen, M., & Eriksson, E. (2020, April). Collaborative technologies for children with special needs: A systematic literature review. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (pp. 1–13).

Bekele, E., Wade, J. W., Bian, D., Zhang, L., Zheng, Z., Swanson, A., ... & Sarkar, N. (2014, June). Multimodal interfaces and sensory fusion in VR for social interactions. In *International Conference on Virtual, Augmented and Mixed Reality* (pp. 14-24). Springer, Cham.

Bekele, E., Zheng, Z., Swanson, A., Davidson, J., Warren, Z., & Sarkar, N. (2013, July). Virtual reality-based facial expressions understanding for teenagers with autism. In *International Conference on Universal Access in Human-Computer Interaction* (pp. 454–463). Springer, Berlin, Heidelberg.

Betancourt, M. A., Dethorne, L. S., Karahalios, K., & Kim, J. G. (2017). Skin conductance as an in situ marker for emotional arousal in children with neurodevelopmental communication impairments: Methodological considerations and clinical implications. *ACM Transactions on Accessible Computing (TACCESS)*, *9*(3), 1–29.

Bian, D., Wade, J., Swanson, A., Weitlauf, A., Warren, Z., & Sarkar, N. (2019). Design of a physiology-based adaptive virtual reality driving platform for individuals with ASD. *ACM Transactions on Accessible Computing (TACCESS)*, *12*(1), 1–24.

Bian, D., Wade, J., Warren, Z., & Sarkar, N. (2016, July). Online engagement detection and task adaptation in a virtual reality based driving simulator for autism intervention. In *International Conference on Universal Access in Human–Computer Interaction* (pp. 538–547). Springer, Cham.

Bian, D., Wade, J. W., Zhang, L., Bekele, E., Swanson, A., Crittendon, J. A., ... & Sarkar, N. (2013, July). A novel virtual reality driving environment for autism intervention. In *International Conference on Universal Access in Human-Computer Interaction* (pp. 474–483). Springer, Berlin, Heidelberg.

Bradshaw, J., Shic, F., Holden, A. N., Horowitz, E. J., Barrett, A. C., German, T. C., & Vernon, T. W. (2019). The use of eye tracking as a biomarker of treatment outcome in a pilot randomized clinical trial for young children with autism. *Autism Research*, *12*(5), 779–793.

Castellano, G., Leite, I., Pereira, A., Martinho, C., Paiva, A., & Mcowan, P. W. (2014). Context-sensitive affect recognition for a robotic game companion. *ACM Transactions on Interactive Intelligent Systems (TiiS)*, *4*(2), 1–25.

Charisi, V., Malinverni, L., Schaper, M. M., & Rubegni, E. (2020). Creating opportunities for children's critical reflections on AI, robotics and other intelligent technologies. In *Proceedings of the 2020 ACM Interaction Design and Children Conference: Extended Abstracts* (pp. 89–95).

Christensen, A. L., & Biskjaer, M. M. (2018, September). Designing digital interactive instructions for children's construction play. In *Proceedings of the 36th European Conference on Cognitive Ergonomics* (pp. 1–8).

Clegg, T., Norooz, L., Kang, S., Byrne, V., Katzen, M., Velez, R., ... & Bonsignore, E. (2017, May). Live physiological sensing and visualization ecosystems: an activity theory analysis. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (pp. 2029–2041).

Crescenzi-Lanna, L. (2020). Multimodal learning analytics research with young children: A systematic review. *British Journal of Educational Technology*, *51*(5), 1485–1504.

Csikszentmihalyi, M., Montijo, M. N., & Mouton, A. R. (2018). Flow theory: Optimizing elite performance in the creative realm.

Cukurova, M., Giannakos, M., & Martinez-Maldonado, R. (2020). The promise and challenges of multimodal learning analytics.

de Barbaro, K. (2019). Automated sensing of daily activity: A new lens into development. *Developmental Psychobiology*, *61*(3), 444–464.

de Wet, L., & Potgieter, L. (2019, July). The impact of binaural beats on user performance and emotions using a BCI for robot control. In *International Conference on Applied Human Factors and Ergonomics* (pp. 93–105). Springer, Cham.

Di Lascio, E., Gashi, S., & Santini, S. (2018). Unobtrusive assessment of students' emotional engagement during lectures using electrodermal activity sensors. Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, 2(3), 1-21.

Dinet, J., & Kitajima, M. (2018, April). Immersive interfaces for engagement and learning: Cognitive implications. In *Proceedings of the Virtual Reality International Conference-Laval Virtual* (pp. 1–8).

Dowthwaite, L., Creswick, H., Portillo, V., Zhao, J., Patel, M., Vallejos, E. P., ... & Jirotka, M. (2020, June). "It's your private information. It's your life." Young people's views of personal data use by online technologies. In *Proceedings of the Interaction Design and Children Conference* (pp. 121–134).

Dybå, T., & Dingsøyr, T. (2008, October). Strength of evidence in systematic reviews in software engineering. In Proceedings of the Second ACM-IEEE international symposium on Empirical software engineering and measurement (pp. 178-187).

Eisenberg, M. (2017). The binding of Fenrir: Children in an emerging age of transhumanist technology. In *Proceedings of the Sixteenth International Conference of Interaction Design and Children, IDC '17*, (pp. 329–333). ACM, New York, NY.

Frauenberger, C., Landoni, M., Fails, J. A., Read, J. C., Antle, A. N., & Gourlet, P. (2019). Panel: Broadening the Discussion of Ethics in the Interaction Design and Children Community. In *IDC'19: Proceedings of the 18th ACM International Conference on Interaction Design and Children*.

Giannakos, M., Papamitsiou, Z., Markopoulos, P., Read, J., & Hourcade, J. P. (2020a). Mapping child–computer interaction research through co-word analysis. *International Journal of Child-Computer Interaction*.

Giannakos, M. N., Sharma, K., Papavlasopoulou, S., Pappas, I. O., & Kostakos, V. (2020b). Fitbit for learning: Towards capturing the learning experience using wearable sensing. *International Journal of Human–Computer Studies*, *136*, 102384.

Giannakos, M. N., Sharma, K., Pappas, I. O., Kostakos, V., & Velloso, E. (2019). Multimodal data as a means to understand the learning experience. *International Journal of Information Management*, *48*, 108–119.

Goodwin, M. S. (2008). Enhancing and accelerating the pace of autism research and treatment: The promise of developing innovative technology. *Focus on Autism and Other Developmental Disabilities*, *23*(2), 125–128.

Goodwin, M. S., Özdenizci, O., Cumpanasoiu, C., Tian, P., Guo, Y., Stedman, A., ... & Ioannidis, S. (2018, May). Predicting imminent aggression onset in minimally-verbal youth with autism spectrum disorder using preceding physiological signals. In *Proceedings of the 12th EAI International Conference on Pervasive Computing Technologies for Healthcare* (pp. 201–207).

Greenhalgh, T. (1997). How to read a paper: Assessing the methodological quality of published papers. *BMJ*, *315*(7103), 305–308.

Hauser, S., Suto, M. J., Holsti, L., Ranger, M., & MacLean, K. E. (2020, April). Designing and evaluating Calmer, a device for simulating maternal skin-to-skin holding for premature infants. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (pp. 1–15).

Hedman, E. B. (2014). *Thick psychophysiology for empathic design* (Doctoral dissertation, Massachusetts Institute of Technology).

Hourcade, JP, Zeising, A., Iversen, OS, Pares, N., Eisenberg, M., Quintana, C., & Skov, MB (2017). Child–computer interaction sig: Ethics and values. In *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems* (pp. 1334–1337).

Iversen, O. S., Smith, R. C., & Dindler, C. (2017, June). Child as protagonist: Expanding the role of children in participatory design. In *Proceedings of the 2017 Conference on Interaction Design and Children* (pp. 27–37).

Jaliaawala, M. S., & Khan, R. A. (2020). Can autism be catered with artificial intelligence-assisted intervention technology? A comprehensive survey. *Artificial Intelligence Review*, *53*(2), 1039–1069.

Jarodzka, H., Skuballa, I., & Gruber, H. (2020). Eye-tracking in educational practice: Investigating visual perception underlying teaching and learning in the classroom. *Educational Psychology Review*, 1–10.

Javed, H., Burns, R., Jeon, M., Howard, A. M., & Park, C. H. (2019). A robotic framework to facilitate sensory experiences for children with autism spectrum disorder: A preliminary study. *ACM Transactions on Human–Robot Interaction (THRI)*, *9*(1), 1–26.

Jeon, J. H., Kim, G., & Kim, J. A. (2019, July). A study on the development of the psychological assessment a using eye-tracking: Focused on eye gaze processing of literacy text. In *International Conference on Human-Computer Interaction* (pp. 74–83). Springer, Cham.

Jiang, X., Chen, Y., Huang, W., Zhang, T., Gao, C., Xing, Y., & Zheng, Y. (2020, April). WeDA: Designing and evaluating a scale-driven wearable diagnostic assessment system for children with ADHD. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (pp. 1–12).

Johnson, K. T., & Picard, R. W. (2017, June). SPRING: Customizable, motivation-driven technology for children with autism or neurodevelopmental differences. In *IDC* (pp. 149–158).

Jyoti, V., & Lahiri, U. (2020). Human–computer interaction based joint attention cues: Implications on functional and physiological measures for children with autism spectrum disorder. *Computers in Human Behavior*, *104*, 106163.

Kabali, H. K., Irigoyen, M. M., Nunez-Davis, R., Budacki, J. G., Mohanty, S. H., Leister, K. P., & Bonner, R. L. (2015). Exposure and use of mobile media devices by young children. *Pediatrics*, *136*(6), 1044–1050.

Kail, R. V. (2011). *Children and their development (6th Edition) (Mydevelopmentlab Series)*. Prentice Hall, Englewood Cliffs, NJ:.

Kawas, S., Yuan, Y., DeWitt, A., Jin, Q., Kirchner, S., Bilger, A., ... & Yarosh, S. (2020, June). Another decade of IDC research: Examining and reflecting on values and ethics. In *Proceedings of the Interaction Design and Children Conference* (pp. 205–215).

Kitchenham, B., & Charters, S. (2007). Guidelines for performing systematic literature reviews in software engineering.

Korhonen, O., Väyrynen, K., Krautwald, T., Bilby, G., Broers, H. A. T., Giunti, G., & Isomursu, M. (2020). Data-Driven personalization of a physiotherapy care pathway: Case study of posture scanning. *JMIR Rehabilitation and Assistive Technologies*, *7*(2), e18508.

Kourakli, M., Altanis, I., Retalis, S., Boloudakis, M., Zbainos, D., & Antonopoulou, K. (2017). Towards the improvement of the cognitive, motoric and academic skills of students with special educational needs using Kinect learning games. *International Journal of Child–Computer Interaction*, *11*, 28–39.

Kuzminykh, A., & Lank, E. (2019). How much is too much? Understanding the information needs of parents of young children. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, *3*(2), 1–21.

Lee, V. R., Drake, J. R., & Thayne, J. L. (2016). Appropriating quantified self technologies to support elementary statistical teaching and learning. *IEEE Transactions on Learning Technologies*, *9*(4), 354–365.

Lee-Cultura, S., Sharma, K., Papavlasopoulou, S., & Giannakos, M. (2020, August). Motion-based educational games: Using multi-modal data to predict player's performance. In *2020 IEEE Conference on Games (CoG)* (pp. 17–24). IEEE.

Lopes, J., Lohan, K., & Hastie, H. (2018, October). Symptoms of cognitive load in interactions with a dialogue system. In *Proceedings of the Workshop on Modeling Cognitive Processes from Multimodal Data* (pp. 1–5).

Lundberg, U., Westermark, O., & Rasch, B. (1993). Cardiovascular and neuroendocrine activity in preschool children: Comparison between day-care and home levels. *Scandinavian Journal of Psychology*, *34*(4), 371–378.

Mangaroska, K., & Giannakos, M. (2018). Learning analytics for learning design: A systematic literature review of analytics-driven design to enhance learning. *IEEE Transactions on Learning Technologies*, *12*(4), 516–534.

Markopoulos, P., Read, J., Giannakos, M. (2021, in press). *Design of Digital Technologies for Children, Handbook of Human Factors and Ergonomics*, 5th edition.

Melo, F. S., Sardinha, A., Belo, D., Couto, M., Faria, M., Farias, A., ... & Luz, L. (2019). Project INSIDE: towards autonomous semi-unstructured human–robot social interaction in autism therapy. *Artificial Intelligence in Medicine*, *96*, 198–216.

Mohammad, Y., & Nishida, T. (2010). Using physiological signals to detect natural interactive behavior. *Applied Intelligence*, *33*(1), 79–92.

Morrison, A., Manresa-Yee, C., & Knoche, H. (2015, September). Vibrotactile Vest and The Humming Wall: "I like the hand down my spine". In *Proceedings of the XVI International Conference on Human Computer Interaction* (pp. 1–8).

Niknejad, N., Ismail, W. B., Mardani, A., Liao, H., & Ghani, I. (2020). A comprehensive overview of smart wearables: The state of the art literature, recent advances, and future challenges. *Engineering Applications of Artificial Intelligence*, *90*, 103529.

Oh, J., & Kang, H. (2020). User engagement with smart wearables: Four defining factors and a process model. *Mobile Media & Communication*.

Okita, S. Y., Ng-Thow-Hing, V., & Sarvadevabhatla, R. K. (2011). Multimodal approach to affective human-robot interaction design with children. *ACM Transactions on Interactive Intelligent Systems (TiiS)*, *1*(1), 1–29.

Olsen, J. K., Sharma, K., Rummel, N., & Aleven, V. (2020). Temporal analysis of multimodal data to predict collaborative learning outcomes. *British Journal of Educational Technology*, *51*(5), 1527–1547.

Papavlasopoulou, S., Giannakos, M. N., & Jaccheri, L. (2019). Exploring children's learning experience in constructionism-based coding activities through design-based research. *Computers in Human Behavior*, *99*, 415–427.

Papavlasopoulou, S., Sharma, K., & Giannakos, M. N. (2018). How do you feel about learning to code? Investigating the effect of children's attitudes towards coding using eye-tracking. *International Journal of Child-Computer Interaction*, *17*, 50–60.

Pathak, M., Bennett, A., & Shui, A. M. (2019). Correlates of adaptive behavior profiles in a large cohort of children with autism: The autism speaks Autism Treatment Network registry data. *Autism*, *23*(1), 87–99.

Pike, M., Ramchurn, R., Benford, S., & Wilson, M. L. (2016, May). # scanners: Exploring the control of adaptive films using brain-computer interaction. In Proceedings of the 2016 chi conference on human factors in computing systems (pp. 5385-5396).

Poria, S., Cambria, E., Bajpai, R., & Hussain, A. (2017). A review of affective computing: From unimodal analysis to multimodal fusion. *Information Fusion*, *37*, 98–125.

Prieto, L. P., Sharma, K., Kidzinski, Ł., Rodríguez-Triana, M. J., & Dillenbourg, P. (2018). Multimodal teaching analytics: Automated extraction of orchestration graphs from wearable sensor data. *Journal of computer assisted learning*, *34*(2), 193–203.

Radeta, M., Cesario, V., Matos, S., & Nisi, V. (2017, November). Gaming versus storytelling: understanding children's interactive experiences in a museum setting. In International Conference on Interactive Digital Storytelling (pp. 163-178). Springer, Cham.

Ruhland, K., Peters, C. E., Andrist, S., Badler, J. B., Badler, N. I., Gleicher, M., ... & McDonnell, R. (2015, September). A review of eye gaze in virtual agents, social robotics and HCI: Behaviour generation, user interaction and perception. *Computer Graphics Forum 34*(6), 299–326.

Ruiz, S., Charleer, S., Urretavizcaya, M., Klerkx, J., Fernández-Castro, I., & Duval, E. (2016, April). Supporting learning by considering emotions: tracking and visualization a case study. In *Proceedings of the Sixth International Conference on Learning Analytics & Knowledge* (pp. 254–263).

Saadatzi, M. N., Welch, K. C., Pennington, R., & Graham, J. (2013, July). Towards an affective computing feedback system to benefit underserved individuals: an example teaching social media skills. In *International Conference on Universal Access in Human-Computer Interaction* (pp. 504–513). Springer, Berlin, Heidelberg.

Schuller, D. M., & Schuller, B. W. (2020). A review on five recent and near-future developments in computational processing of emotion in the human voice. *Emotion Review*.

Sharma, K., & Giannakos, M. (2020a). Multimodal data capabilities for learning: What can multimodal data tell us about learning?. *British Journal of Educational Technology*, *51*(5), 1450–1484.

Sharma, K., Giannakos, M., & Dillenbourg, P. (2020b). Eye-tracking and artificial intelligence to enhance motivation and learning. *Smart Learning Environments, 7*, 1–19.

Sharma, K., Papamitsiou, Z., & Giannakos, M. (2019). Building pipelines for educational data using AI and multimodal analytics: A “grey-box” approach. *British Journal of Educational Technology, 50*(6), 3004–3031.

Sharma, K., Papamitsiou, Z., Olsen, J. K., & Giannakos, M. (2020c). Predicting learners' effortful behaviour in adaptive assessment using multimodal data. In *Proceedings of the Tenth International Conference on Learning Analytics & Knowledge* (pp. 480–489).

Shattuck, P. T., Narendorf, S. C., Cooper, B., Sterzing, P. R., Wagner, M., & Taylor, J. L. (2012). Postsecondary education and employment among youth with an autism spectrum disorder. *Pediatrics*, *129*(6), 1042–1049.

Silva, H. P. D., Alves, A. P., Lourenço, A. R., Fred, A., Montalvão, I., & Alegre, L. (2013). Towards the detection of deception in interactive multimedia environments. In *Human–Computer Interaction and Knowledge Discovery in Complex, Unstructured, Big Data-Proceedings of Third International Workshop, HCI-KDD 2013* (Vol. 7947, pp. 65–76). Springer, Berlin Heidelberg.

Sonne, T., & Grønbæk, K. (2015, September). Designing assistive technologies for the ADHD domain. In *International Symposium on Pervasive Computing Paradigms for Mental Health* (pp. 259–268). Springer, Cham.

Sonne, T., Merritt, T., Marshall, P., Lomholt, J. J., Müller, J., & Grønbæk, K. (2017, June). Calming children when drawing blood using breath-based biofeedback. In *Proceedings of the 2017 Conference on Designing Interactive Systems* (pp. 725–737).

Spikol, D., Ruffaldi, E., Dabisias, G., & Cukurova, M. (2018). Supervised machine learning in multimodal learning analytics for estimating success in project- based learning. Journal of Computer Assisted Learning, 34(4), 366-377.

Sridhar, P. K., Chan, S. W., Chua, Y., Quin, Y. W., & Nanayakkara, S. (2019). Going beyond performance scores: Understanding cognitive–affective states in Kindergarteners and application of framework in classrooms. *International Journal of Child–Computer Interaction*, *21*, 37–53.

Sridhar, P. K., Chan, S. W., & Nanayakkara, S. (2018, June). Going beyond performance scores: Understanding cognitive-affective states in kindergarteners. In *Proceedings of the 17th ACM Conference on Interaction Design and Children* (pp. 253–265).

Stone, B., Skinner, A., Stikic, M., & Johnson, R. (2014, June). Assessing neural synchrony in tutoring dyads. In *International Conference on Augmented Cognition* (pp. 167–178). Springer, Cham.

Sweller, J. (2011). Cognitive load theory. In *Psychology of Learning and Motivation* (Vol. 55, pp. 37–76). Academic Press.

Thieme, A., Belgrave, D., Sano, A., & Doherty, G. (2020). Machine learning applications: Reflections on mental health assessment and ethics. *Interactions*, 27(2), 6–7.

Van Berkel, N., Ferreira, D., & Kostakos, V. (2017). The experience sampling method on mobile devices. *ACM Computing Surveys (CSUR), 50*(6), 1–40.

Van Mechelen, M., Baykal, G. E., Dindler, C., Eriksson, E., & Iversen, O. S. (2020, June). 18 Years of ethics in child-computer interaction research: a systematic literature review. In *Proceedings of the Interaction Design and Children Conference* (pp. 161–183).

Vasiljevic, G. A. M., & de Miranda, L. C. (2020). Brain–computer interface games based on consumer-grade EEG Devices: A systematic literature review. *International Journal of Human–Computer Interaction*, *36*(2), 105–142.

Wade, J., Bian, D., Fan, J., Zhang, L., Swanson, A., Sarkar, M., ... & Sarkar, N. (2015, August). A virtual reality driving environment for training safe gaze patterns: application in individuals with ASD. In *International Conference on Universal Access in Human-Computer Interaction* (pp. 689–697). Springer, Cham.

Wade, J., Bian, D., Zhang, L., Swanson, A., Sarkar, M., Warren, Z., & Sarkar, N. (2014, June). Design of a virtual reality driving environment to assess performance of teenagers with ASD. In *International Conference on Universal Access in Human-Computer Interaction* (pp. 466–474). Springer, Cham.

Wang, R., Wang, W., DaSilva, A., Huckins, J. F., Kelley, W. M., Heatherton, T. F., & Campbell, A. T. (2018). Tracking depression dynamics in college students using mobile phone and wearable sensing. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, *2*(1), 1–26.

Wade, J., Zhang, L., Bian, D., Fan, J., Swanson, A., Weitlauf, A., ... & Sarkar, N. (2016). A gaze-contingent adaptive virtual reality driving environment for intervention in individuals with autism spectrum disorders. *ACM Transactions on Interactive Intelligent Systems (TiiS)*, *6*(1), 1–23.

Woolf, B., Dragon, T., Arroyo, I., Cooper, D., Burleson, W., & Muldner, K. (2009, July). Recognizing and responding to student affect. In *International Conference on Human-Computer Interaction* (pp. 713–722). Springer, Berlin, Heidelberg.

Worsley, M., & Blikstein, P. (2018). A multimodal analysis of making. International Journal of Artificial Intelligence in Education, 28(3), 385-419.

Xu, Y., Poole, E. S., Miller, A. D., Eiriksdottir, E., Catrambone, R., & Mynatt, E. D. (2012, May). Designing pervasive health games for sustainability, adaptability and sociability. In *Proceedings of the International Conference on the Foundations of Digital Games* (pp. 49–56).

Yadegaridehkordi, E., Noor, N. F. B. M., Ayub, M. N. B., Affal, H. B., & Hussin, N. B. (2019). Affective computing in education: A systematic review and future research. *Computers & Education*, *142*, 103649.

Yannakakis, G. N., Hallam, J., & Lund, H. H. (2008). Entertainment capture through heart rate activity in physical interactive playgrounds. *User Modeling and User-Adapted Interaction*, *18*(1–2), 207–243.

Yuan, Y., Thompson, S., Watson, K., Chase, A., Senthilkumar, A., Brush, A. B., & Yarosh, S. (2019). Speech interface reformulations and voice assistant personification preferences of children and parents. *International Journal of Child-Computer Interaction*, *21*, 77–88.

Zhang, L., Wade, J. W., Bian, D., Swanson, A., Warren, Z., & Sarkar, N. (2014, June). Data fusion for difficulty adjustment in an adaptive virtual reality game system for autism intervention. In *International Conference on Human-Computer Interaction* (pp. 648–652). Springer, Cham.

Zhang, L., Wade, J., Bian, D., Fan, J., Swanson, A., Weitlauf, A., ... & Sarkar, N. (2015, August). Multimodal fusion for cognitive load measurement in an adaptive virtual reality driving task for autism intervention. In *International Conference on Universal Access in Human–Computer Interaction* (pp. 709–720). Springer, Cham.

Conflict of Interest

Declaration of interests

☒ The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

☐The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: