WILEY | Hindawi

Review Article

Fairies in the Box: Children's Perception and Interaction towards Voice Assistants

Yu Tong [ID],[1,2] **Fuxing Wang** [ID],[1,2] **and Weijun Wang** [ID][1,2]

[1]*School of Psychology, Central China Normal University, Wuhan, Hubei 430079, China*
[2]*Key Laboratory of Adolescent Cyberpsychology and Behavior, Ministry of Education, Wuhan, Hubei 430079, China*

Correspondence should be addressed to Fuxing Wang; fxwang@ccnu.edu.cn and Weijun Wang; wangwj@ccnu.edu.cn

In recent years, the growing popularity of smart speakers (e.g., Google Home and Alexa) has facilitated young children's interaction with internet-based devices and provided them with more opportunities to obtain access to online information. This review summarizes the current state of the research by examining smart speakers' core characteristics, children's conceptualization and interaction with smart speakers, and the influences on children's learning and habits. Our review shows that (a) the natural language processing technology and central computing system (Internet) contribute to the uniqueness of smart speakers; (b) although children tend to attribute human characteristics (e.g., smart and friendly) to smart speakers, they might judge these voice assistant devices as neither explicitly living nor nonliving in ontological perception; (c) children's overattributing certain knowledge (e.g., questions about personal information) to smart speakers does not necessarily mean that this device is believed to be omniscient; and (d) in terms of promoting children's learning, smart speakers might not be more effective than a real human, and the interaction with smart speakers may not be conducive to children's maintenance of civilized social norms. Implications for children's conceptualization and interaction of smart speakers and the design of children-oriented smart agents are also discussed.

1. Introduction

Following touchscreen devices, smart speakers (e.g., Google Home from Google, Alexa from Amazon, and Tmall Genie from Alibaba) have become popular for children in the past five years. According to the latest survey released by Common Sense Media in 2020, 41% of young children aged 0-8 in the US have a smart speaker in their homes, while the proportion was only 9% in 2017 [1]. However, whenever new mass media or technological devices (e.g., wireless radio, television, smartphones, or tablets) are widely introduced into public life [2, 3], parents and educators always display strong concern: how would these emerging devices affect our children's development? In recent years, as children have increasingly gained access to smart speakers, this emerging internet-based device has not escaped the attention and doubt of researchers and educators. Though some researchers have focused on this field, existing reviews mainly concentrated on children's voice search [4] or outlined some potential directions about how smart speakers might influence children's development [5–8]. Given that comprehensive understanding of how young children understand and interact with smart speakers is of great importance for parents and educators to help their children interact with this emerging technology rationally, in this review, based on the latest empirical findings, we summarize the current state of research on children and smart speakers and hope to answer several questions of interest to researchers as well as parents: (1) the core features of smart speakers; (2) how children aged 3-12 years conceptualize smart speakers; (3) how children interact with smart speaker devices; and (4) how smart speaker devices impact children's learning and daily behavior.

To find studies that fit the scope of this review, the literature was identified by searching for "*child * AND (smart speaker OR voice assistant OR conversational agent OR voice search)*" from Web of Science, ACM, and Google Scholar. Studies must

address how children ages 3-12 understand and interact with smart speakers. Researches that focused on abnormal children (e.g., children with ASD) or described a new program design about smart speaker were excluded. Once relevant studies were identified, their reference sections were examined for other relevant studies. Overall, 40 research articles, conference presentations, or published proceedings were identified.

2. Core Features of Smart Speakers

According to "the basic model of human behavior with technologies" proposed by Yan (2020) [9], how users interact with technology determines their activities and effects (see Figure 1). Therefore, the human-computer interaction mode allowed by technologies should be the core feature of the device. In terms of interaction, smart speakers widely adopt voice-driven interfaces based on natural language processing technology [6]. Unlike earlier speech recognition systems, which plagued users with their poor flexibility, natural language processing no longer requires users to use specific words or follow certain patterns when giving commands or asking questions. Therefore, smart speakers could provide children who have limited literacy skills and immature fine motor skills unprecedented opportunities to interact with emerging internet-based devices [6, 10, 11].

Second, compared to other entities that also utilize voice interaction (e.g., social bots based on a specific corpus), smart speakers operate in a networked environment, suggesting that this kind of emerging device is equipped with a centralized central processing system (i.e., Internet) and thus makes it possible to process a much richer set of user commands. Furthermore, under the same wireless local area network, users can indirectly control other connected devices (such as smart TV, lighting, and air conditioning) through smart speakers, which, while expanding the potential capabilities of smart speakers [6], may prompt young children to perceive smart speakers differently from other devices [12].

3. How Children Conceptualize Smart Speakers

3.1. Is It Alive? Children's Perception of Smart Speaker's Ontological Categories. Progressing from touchscreen to natural speech processing utilized by smart speakers does not only imply a change in the mode of human-computer interaction. Given that speech is the most natural way for people to communicate with others and smart speakers have now employed this form, which was previously exclusive to human beings, children's understanding of whether smart speakers are alive may become blurred while they communicate with devices [5]. By observing how children aged 3-12 interacted with smart speakers, researchers found that the majority of children believed that smart speakers were smart, friendly, and trustworthy. While interacting, children often asked questions about their identity and personality and even joked with the device as if it was a real person indeed (e.g., "Google, can I eat you?"; [11, 13–16]). When asked to draw the smart speaker, elementary school children tended to display anthropomorphic features in their paint-

ings (e.g., eyes, limbs, and facial expressions; [17]). During communications, children showed nonverbal behaviors such as nodding, smiling, shrugging, and frowning, which the smart speaker actually could not actually recognize [18]. Hence, some researchers believed that children anthropomorphized smart speakers [19].

However, the view "young children anthropomorphize smart speakers" implies the default assumption that young children believe smart speakers are nonliving in ontological categories, which might not hold true for artificially intelligent entities. Although previous studies have found that children as young as age 3 could distinguish between living and nonliving kinds and that such discrimination is clear and stable [20], the emergence of robots, as well as other artificial intelligent entities in recent years, has increasingly challenged these existing ontological understandings. Although lacking anthropomorphic or animal-like appearance and unable to move autonomously, smart speakers have been equipped with the most prominent feature which was exclusive to human beings—speech. Kahn et al. [12] proposed that these artificially intelligent devices, such as smart speakers and other entities based on embodied social computational systems, might be perceived as a hybridization of living and nonliving, which do not belong to any existing ontological categories but form a completely new ontological category (NOC). After observing 3- to 6-year-old children's 40-minute interaction with a Google Home Mini and then examining their ontological reasoning, Xu and Warschauer [21] found that although the majority of children attributed human-like capacities to smart speakers (93% of children thought the Google Home Mini was smart, 86% thought it had memory, and 64% thought it had feelings), only 18% of children claimed it was indeed "human," while 25% believed "it was neither an artifact nor a living object." Likewise, Festerling and Siraj [14] suggested that children would interpret smart speakers as the superimposition of genuinely humanoid and nonhumanoid speakers. When smart speakers could not respond immediately or presented incorrect answers, children tended to believe that the device was controlled by a real human; however, when smart speakers displayed instant, standardized responses, and lack of common sense, children perceived the same smart speaker as a machine.

3.2. Omniscient God? How Children Perceive Smart Speaker's Scope of Knowledge. Connected to the Internet, smart speakers could provide young children with unprecedented opportunities to obtain access to online information. Rücker and Pinkwart [22] suggested that children were likely to view a single internet-based device as equivalent to the Internet and believed that one single device itself was a giant database into which they could save and extract all kinds of information but failed to realize the limitations of the scope of knowledge of the device as well as the Internet. Given that children sometimes asked smart speakers questions that were beyond the Internet knowledge domain (e.g., asking where their mom is[11, 23]), some researchers speculated that children would perceive smart speakers to be omniscient [11, 23].

However, considering the precise definition of "omniscient—knowing everything about everything," using the
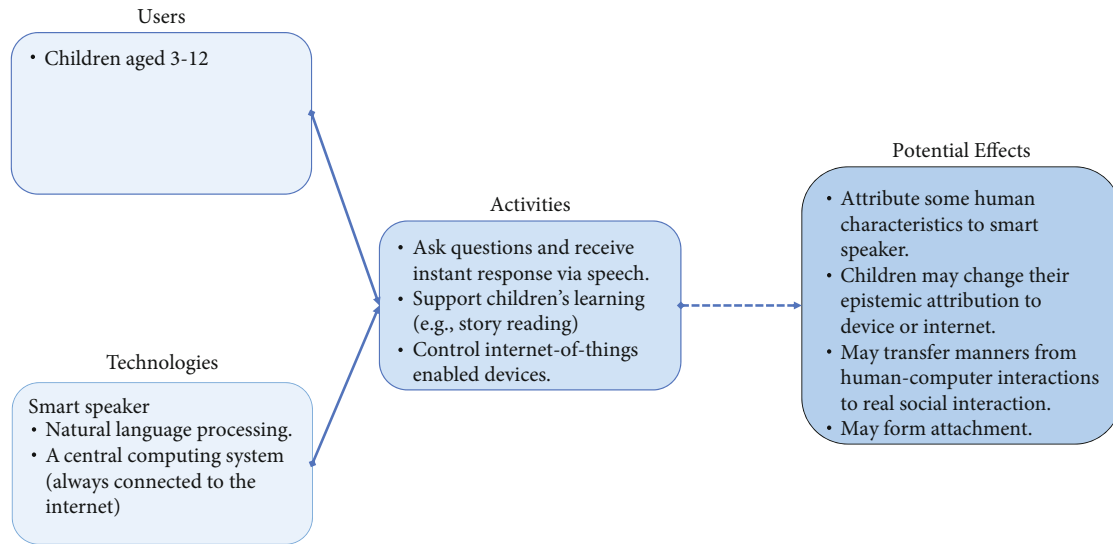
FIGURE 1: Applying the basic model to the smart speaker study.

evidence "children sometimes asked smart speaker questions beyond its capacity" to infer the conclusion "children would believe internet-based devices such as smart speaker are omniscient" may be overgeneralizing. Inspired by the constructive theories of cognitive development, children's understanding of novel entities and phenomena strongly relied on their existing concepts [24–26]. This suggests that when children try to understand the knowledge state of smart speakers or the Internet, it is possible for them to make inferences based on their existing schema of a real human [27–29]. Given that even young children have realized the limitation of people's knowledge state through daily interaction, it might be difficult for children to confirm that there did exist a truly omniscient agent or entity [30]. Indeed, some empirical studies found that children did not believe that the Internet and internet-based devices would know everything [31, 32]. When children explored the boundaries of smart speakers or the Internet knowledge state by raising questions, the frequency of seeking facts (e.g., technology, history, language, spelling, and translation) increased with age, while the proportion of asking private information and other requirements beyond the scope of online information decreased [11, 16]. Thus, it is reasonable to speculate that children asking smart speakers inappropriate questions might be children's exploration or use of trial-and-error. Children may conceptualize the knowledge state of smart speakers by firsthand observation and experimentation as well as secondhand testimony, similar to how they understand an individual's knowledge state [33].

4. How Children Interact with Smart Speakers

Unlike screen media, the natural language processing interface used in smart speakers allows young children to interact with the device in a way that most closely resembles real human interaction. However, ultimately, smart speakers are machines rather than real persons. They could not reply to children's commands in a personalized way that incorporates young children's background knowledge and emotional state, which might be attributed to a unique pattern of computer-human interaction.

How do children interact with a smart speaker? There are various sorting methods based on different perspectives and observations. Some researchers concentrated on the goal of children's use of smart speakers and classified children's activity into "Exploration," "Information Seeking," and "Functional Tasks and Commands" [23]. Other researchers have focused on smart speakers' functions and divided the interaction into "sending and reading text messages, making phone calls, and sending and receiving emails," "answering basic questions," "setting timers, alarms, reminders, schedules, etc.," "controlling other media and other connected devices," and "telling jokes or stories" [6]. However, the above classification only focused on the one-way commands of children to smart speakers while weakening the continuous and alternating interaction between the users and devices. To highlight the dynamic interaction process between children and smart speakers, we attempt to summarize the interaction between children and smart speakers from two perspectives: how children command smart speakers and how they react to smart speakers' responses.

Previous studies have suggested that when children interacted with smart speakers, requesting information was the most common activity [11, 16, 23, 34, 35]. Based on children's purpose, children's questions could be categorized as "exploring the identity of the device" and "using smart speaker to seek another information" [23], including asking about personal preferences and characteristics of voice assistants [13, 14], background information [14], location [23], exploring children's own personal information related to their family, and asking facts about science and technology, culture, language (e.g., word spelling), modus operandi (e.g., recipes and navigation), and other objective knowledge [11]. Apart from questioning, users can also operate the smart speaker for other extended functions and indirectly control other smart devices that are connected under the same local wireless network. For children, although a series

of observations have suggested that children could use smart speakers to send messages [23], make phone calls [23], listen to music, and command the device to tell jokes [11, 36], the frequency of children's use of these extended functions is much lower than that of seeking knowledge or information from the smart speaker device [11, 23, 35]. This may be because children primarily view smart speakers as a source of information rather than an entertainment tool [35]. Another possible explanation might be that most of the relevant studies are based on usage in real family settings, where children in full parental care are likely not to need to command smart speakers themselves.

In terms of children's responses to the smart speaker, when the smart speaker is able to accurately recognize and respond to children's speech and instructions, a series of observations found that children had a tendency to actively respond to the smart speaker's questions and commands, showing active participation similar to that of a real person, as well as displaying basic features of interpersonal dialog such as shared attention and coordinated reciprocity [10, 37]. In addition, when responding to questions from the smart speaker, children not only actively used verbal communication but also often displayed nonverbal expressions such as shaking heads and shrugs [18]. These results suggest that smart speakers can trigger children's natural interpersonal responses as a real human might do.

However, due to children's (especially preschoolers') limited verbal expression skills, when they communicate with electronic devices based on voice interaction, such as smart speakers, approximately half of their commands cannot be accurately recognized by the device, resulting in a breakdown in human-machine communication [11]. At this point, children actively employ a range of conversation repair strategies, including changing the topic of the question and the type of question asked [35], repeating instructions, increasing the volume, highlighting pronunciation and adding pauses between words or syllables, adding background information, replacing words, and reorganizing word order [10, 13, 23, 38]. Moreover, compared to adult users, children showed less frustration when faced with human-machine communication failures, were less likely to attribute communication failures to the device, and showed more attempts to use various repair strategies [10]. In addition, different repair strategies employed by users may reflect their attributions to smart speaker failures. For example, the strategies "repeat command" and "raise volume" which are most commonly used by children may indicate that children tend to attribute the failure of the breakdown to the smart speaker's misunderstanding of characteristic words and the absence of attention in communication [10].

5. How Smart Speaker Affect Child Development

5.1. Is a Smart Speaker Helpful for Children's Early Learning?
By observing smart speakers telling stories to children, researchers found that children used both verbal and nonverbal expressions to actively respond to smart speakers' questions, showing interaction patterns similar to parent-child

communication [18]. Hence, some parents and educators expect smart speaker devices to boost children's learning and development as well as interactions with real people do.

When interacting with children, smart speakers can only use synthesized sounds to answer children's questions but lack all nonverbal communication cues (e.g., fixation, facial expressions, body posture, gestures, and emotional feedback). They have no physical body and cannot move, let alone respond to the user's actions. On the one hand, according to Social Presence Theory [39] and Social Agency Theory [40, 41], the social characteristics of media or agents should have an important impact on learners' communication and learning effects. When smart speakers are involved in children's learning process, they may not be regarded as humanoid social partners and cannot create a sense of social presence similar to the interaction with a real person. Thus, smart speakers should not facilitate children's learning as well as real people do. On the other hand, the social features that smart speakers lack but real people have are likely to increase learners' cognitive load, which occupies students' attention which is necessary for processing the learning material [42]. Therefore, compared with real people, it is also possible for smart speakers to promote children's learning by promoting cognitive resource investment in the content itself and the learning process. Xu and her colleagues [43] compared how a smart speaker and a real person impacted 3- to 6-year-olds' content comprehension and language expression in a storybook reading. And results indicated that a real person companion was more effective than a smart speaker in terms of helping children's story comprehension. This finding could provide support for Social Presence Theory and Social Agency Theory, which suggest that mechanized smart speakers could not be as effective as a real human in facilitating children's learning engagement. In addition, when communicating with children, parents and educators often use rhetorical questions and guided interactions, which can serve as a scaffold for children to explore and further process information on their own. However, while in the dialog with the smart speaker, when children ask questions, smart speakers always present answers directly. Although some researchers have attempted to set scaffolding questions in smart speakers to help children understand [37], the number and role of scaffolds are still limited [36]. These results again suggested that smart speakers might not facilitate children's learning and cognitive processing as well as real people do.

Based on these differences between smart speakers and real people, some researchers attempted to increase the social features of smart speakers to improve their facilitation of children's learning but failed to achieve the desired results. For example, Yuan et al. [38] used the Wizard of Oz (where a real person simulates the inner workings of a smart speaker thus excludes the effect of confounding misrecognition) to add "personalization (changing language details for different children such as calling a child's name)" and "anthropomorphization (giving the device a human identity)" to smart speakers and asked children aged 5-12 years to interact with a "question-only smart speaker," "naming personalization of the device," and "smart speaker with personalization and anthropomorphization." The results showed that although

children aged 5-12 years preferred smart speakers with personalization, the personalization and anthropomorphization of the speaker had no effect on promoting the effectiveness of children's questions. Other researchers have also added subjective words related to feelings and emotions to the statements provided by smart speakers; but compared to the devices that only provided objective facts, the effectiveness of communication between 5- to 6-year-olds and smart speakers was not significantly improved [44]. Some researchers suggest that "Social Agency Theory" requires a certain threshold: smart speakers are too different from a real person in physical appearance; thus, adding only verbal human-like features is not enough to lead children to perceive the smart speaker as a "real person" [44, 45].

Overall, although the idea of using devices such as smart speakers to assist young children's learning is very appealing, given the shortcomings that still exist in the design and flexibility, smart speakers cannot yet fully replace the role of parents and educators [46]. Although smart speakers will not hinder children's learning, they should not be viewed as a substitute for parental companionship and social connection; instead, parents should teach their children to realize that it is only a tool or a machine [8].

It is also worth noting that though smart speakers may not facilitate children's deep cognitive processing as well as real people. When researchers examined speech patterns used by young children, they found that children aged 3-6 years expressed clearer and more comprehensible language during conversations with smart speakers than with real people [43]. This result indicated that smart speakers might have potential advantages in developing young children's verbal communication skills, especially promoting clear articulation and improving speech intelligibility [5, 47]. However, this facilitation is probably a byproduct of the inadequate maturity of natural language processing technology with a low recognition rate of young children's speech and vocabulary.

5.2. Is a Smart Speaker Harmful for Children's Social Interaction? Regardless of how close the speech used by smart speakers is to real human communication, it is still ultimately a mechanical device; thus, the behavioral patterns developed by young children's interactions with smart speakers may not be applied to real social interaction (e.g., politeness, which is of great value in real social human interaction and is meaningless in communication with smart speakers). Although it has been found that new manners learned by children from one specific smart speaker were easily transferred to another smart speaker, these behavior patterns were not easily transferred to real settings, especially when children interacted with strangers [48]. However, empirical research is still lacking on the more realistic question of whether behavior patterns that are suppressed in real situations but could play a role in interactions with smart speakers will be reinforced in real social situations. For example, studies have found that children often used loud voice and even intimidation tactics when smart speakers failed to recognize their language [23, 49]. Some parents are concerned about whether children would transfer these aggressive verbal expressions into conversations with real people, thus impeding the acquisition and maintenance of good language habits and norms of daily behavior [15, 36]. Secondly, children's interactions with smart speakers are immediate (children give commands and smart speakers respond immediately), which is contrary to "delay gratification" [5]. Thus, too much interaction with and commands to smart speakers may be detrimental to children's development of self-control skills. Third, because smart speakers are equipped with many characteristics of a good partner (e.g., patient listening, no ridicule, and keeping secrets), children may perceive the device as a trustworthy partner with whom they can build emotional and attachment relationships [5, 15, 23, 36, 50], which may reduce children's social need in the real world and thus be detrimental to their social development.

However, as voice-based internet-based devices for public use in the home, smart speakers are more likely to be supervised by parents than other screen-based media (e.g., smartphones and tablets) when used by children [51]. Several studies even suggested that parents made full use of smart speakers to actively cultivate children's daily habits [47]. Therefore, how smart speakers affect children's behavioral habits in real contexts and whether children would transfer behavioral patterns from their interactions with smart speakers to real contexts still require direct evidence from empirical studies, especially tracking studies based on real use situations.

6. Implications

The rapid popularity of smart speakers among children has led businesses to think about how to design products to make them more popular [52]. Considered that most current smart speakers are designed as square or cylindrical, and some surveys suggest that the anthropomorphic shape of smart speakers can appeal to children [53] and meet the expectations of some children [15], some smart speakers were designed to anthropomorphic shapes. However, given that an entity's appearance plays an important role in children's ontological classification [20], it is possible that smart speakers with anthropomorphic shapes might blur children's ontological categorization and be more conducive for children to transfer poor communication habits formed in human-computer interaction to real interpersonal interactions. Moreover, smart speakers with anthropomorphic shapes might make it easier for children to form emotional attachments and establish social connections with mechanized devices, thereby hindering children's interactions in real social contexts. Thus, whether smart speakers with anthropomorphic shapes should be adopted widely remained to be seen. Secondly, studies have found that unlike smartphones or tablets, young children were more likely to view smart speakers as an information source rather than an entertainment tool [35], suggesting that children have a tendency to learn facts from these devices. Thus, it might be important to strengthen the quality control of information conveyed by smart speakers or embed filtering programs to help children shield inappropriate online information.

As for parents, despite that smart speakers allow young children to interact with the device independently through natural speech, considered that young children lack a sophisticated understanding of smart speakers and that social interaction could help children improve their understanding of emerging

technologies [54], parents should increase their joint media engagement with their children. In addition, considered that in the process of interacting with smart speakers, young children may have a tendency to form attachments or master-servant relationships with devices [12] or display several behavioral patterns that are not encouraged in the real society (e.g., raise their volume to repair breakdown in human-machine communication), parents should pay attention to children's interaction with the device and guide them properly. Perhaps in the future, smart speakers may identify children's voiceprint, record it automatically, and then permit parents to review their children's interaction with the device, which provides convenience for parents to monitor their children continuously.

7. Future Directions

While parents and educators have begun to pay attention to how children interact with smart speakers, there are many questions that need to be further discussed.

First, most children in existing studies either had no experience with smart speaker use prior to participating in the experiment (e.g., [11, 37]), or this vital context was not clearly reported in the paper (e.g., [10, 44]). This may lead to the possibility that behaviors exhibited by children in interacting with smart speakers are more likely an exploration of smart speakers. That is, the behavior patterns might not be relatively stable patterns based on their current level of cognitive development and experience but rather an exploration process of children facing new devices. Since behavior is a reflection of the current cognitive state, children could update their cognitive schema continually and then further adjust their behavior patterns [35]. Therefore, based on the observations of how children who are unfamiliar with devices interacted with smart speakers, children's understanding of smart speaker devices may be underestimated.

Second, the gradual entry of smart speakers into the children's lives may have a lasting impact on their learning and development in a subtle way. The existing findings are mainly based on observations of less than one month, and their conclusions are still limited to the revelation of short-term effects. Future studies can examine the long-term effects of smart speakers on children's development through delayed tests.

Third, when children judge real people's knowledge state and decide whether to ask them questions, their choices are influenced by the child's own level of cognitive development (e.g., theory of mind; [55–57]) and informants' past performance [58]. Future studies may examine whether boundary conditions that influence children's judgments of the knowledge states of real people could also be applied to children's understanding of the knowledge and capability states of smart speakers [7].

8. Conclusion

Based on the core features of smart speakers, this review systematically summarizes how children aged 3-12 understand and interact with smart speakers and the impact of smart speaker devices on children's learning and daily behavior. We conclude that natural language processing and network information processing, on which smart speakers are based, play a crucial role in children's smart speaker use and understanding. Children tend to assign human characteristics to smart speakers but do not simply classify them as fully alive or inanimate. Although children ask questions beyond the capabilities of the smart speaker, whether they consider it "omniscient" remains to be explored. In the process of guiding children's reading and speech development, smart speakers are not yet sufficient to facilitate development as well as a real human might. Moreover, children's interactions with smart speakers may negatively affect their cognitive development and the formation or maintenance of good social norms.

Data Availability

Data sharing is not applicable to this article as no datasets were generated or analyzed during the current study.

Ethical Approval

Ethics approval statement is not applicable to this article as no human participants were involved in the current study.

Conflicts of Interest

The researchers declare that there is no conflict of interest involved in this work.

Acknowledgments

References

[1] V. Rideout and M. B. Robb, *The Common Sense Census: media use by kids age zero to eight*, Common Sense Media, San Francisco, CA, 2020, http://www.commonsensemedia.org/sites/default/files/uploads/research/2020_zero_to_eight_census_final_web.pdf.

[2] S. L. Calvert and B. J. Wilson, *The Handbook of Children, Media and Development*, Blackwell Publishing, 2008.

[3] C. Critcher, "Making waves: historical aspects of public debates about children and mass media," in *The International Handbook of Children, Media and Culture*, K. Drotner and S. Livingstone, Eds., pp. 91–104, SAGE Publications Ltd., United States, 2008.

[4] S. B. Lovato and A. M. Piper, "Young children and voice search: what we know from human-computer interaction research," *Frontiers in Psychology*, vol. 10, 2019.

[5] C. Biele, A. Jaskulska, W. Kopec, J. Kowalski, K. Skorupska, and A. Zdrodowska, "How might voice assistants raise our children?," in *Intelligent Human Systems Integration 2019. IHSI 2019. Advances in Intelligent Systems and Computing*, W. Karwowski and T. Ahram, Eds., vol. 903, pp. 162–167, Springer, Cham, 2019.

[6] M. B. Hoy, "Alexa, Siri, Cortana, and more: an introduction to voice assistants," *Medical Reference Services Quarterly*, vol. 37, no. 1, pp. 81–88, 2018.

[7] G. Murray, "Who is more trustworthy, Alexa or mom? Children's selective trust in a digital age," *Technology, Mind & Behavior*, vol. 2, no. 3, 2021.

[8] B. K. Wiederhold, ""Alexa, are you my mom?" the role of artificial intelligence in child development," *Cyberpsychology, Behavior and Social Networking*, vol. 21, no. 8, pp. 471-472, 2018.

[9] Z. Yan, "A basic model of human behavior with technologies," *Human Behavior and Emerging Technologies*, vol. 2, pp. 410–415, 2020.

[10] Y. Cheng, K. Yen, Y. Chen, S. Chen, and A. Hiniker, "Why doesn't it work?: voice-driven interfaces and young children's communication repair strategies," in *Proceedings of the 17th ACM Conference on Interaction Design and Children*, pp. 337–348, New York, 2018.

[11] S. B. Lovato, A. M. Piper, and E. A. Wartella, ""Hey Google, do unicorns exist?": conversational agents as a path to answers to children's questions," in *Paper presented at the Proceedings of the 18th ACM International Conference on Interaction Design and Children*, pp. 301–313, Boise, ID, USA, 2019.

[12] P. H. Kahn, H. E. Gary, and S. Shen, "Children's social relationships with current and near-future robots," *Child Development Perspectives*, vol. 7, no. 1, pp. 32–37, 2013.

[13] S. Druga, R. Williams, C. Breazeal, and M. Resnick, ""Hey Google is it OK if I eat you?": Initial explorations in child-agent interaction," in *Paper presented at the Proceedings of the 2017 Conference on Interaction Design and Children*, pp. 595–600, Stanford, California, USA, 2017.

[14] J. Festerling and I. Siraj, "Alexa, what are you? Exploring primary school children's ontological perceptions of digital voice assistants in open interactions," *Human Development*, vol. 64, no. 1, pp. 26–43, 2020.

[15] R. Garg and S. Sengupta, "Conversational technologies for in-home learning: using co-design to understand children's and parents' perspectives," in *Proceedings of the 2020 CHI conference on human factors in computing systems*, pp. 1–13, New York, NY, USA, 2020a.

[16] S. B. Lovato, *"Hey Google, is God real?": conversational agents answering children's questions*, Paper presented at the SRCD Biennial Meeting, Baltimore, Maryland USA, 2019.

[17] M. Landoni, E. Murgia, T. Huibers, and M. S. Pera, "You've got a friend in me: children and search agents," in *Adjunct Publication of the 28th ACM Conference on User Modeling, Adaptation and Personalization*, pp. 89–94, New York, NY, USA, 2020.

[18] Y. Xu and M. Warschauer, "Exploring young children's engagement in joint reading with a conversational agent," in *Paper presented at the Proceedings of the 20th ACM International Conference on Interaction Design and Children Conference*, pp. 216–228, London, United Kingdom, 2020a.

[19] J. H. Danovitch, "Growing up with Google: how children's understanding and use of internet-based devices relates to cognitive development," *Human Behavior and Emerging Technologies*, vol. 1, no. 2, pp. 81–90, 2019.

[20] J. L. Jipson and S. A. Gelman, "Robots and rodents: Children's inferences about living and nonliving kinds," *Child Development*, vol. 78, no. 6, pp. 1675–1688, 2007.

[21] Y. Xu and M. Warschauer, "What are you talking to?: understanding children's perceptions of conversational agents," in *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, New York, NY, USA, 2020b.

[22] M. T. Rücker and N. Pinkwart, "Review and discussion of children's conceptions of computers," *Journal of Science Education and Technology*, vol. 25, no. 2, pp. 274–283, 2016.

[23] S. B. Lovato and A. M. Piper, ""Siri, is this you?": understanding young children's interactions with voice input systems," in *Proceedings of the 14th international conference on interaction design and children*, Medford, MA, USA, 2015.

[24] S. A. Gelman, "Learning from others: children's construction of concepts," *Annual Review of Psychology*, vol. 60, pp. 115–140, 2009.

[25] A. Gopnik and H. M. Wellman, "Reconstructing constructivism: causal models, Bayesian learning mechanisms, and the theory theory," *Psychological Bulletin*, vol. 138, no. 6, pp. 1085–1108, 2012.

[26] J. Piaget, *Science of Education and the Psychology of the Child. (D. Coltman, Trans.)*, Oxford, England, Orion, 1970.

[27] N. Epley, A. Waytz, and J. T. Cacioppo, "On seeing human: a three-factor theory of anthropomorphism," *Psychological Review*, vol. 114, no. 4, pp. 864–886, 2007.

[28] J. D. Lane, "Constructing ideas of the supernatural," *Journal of Cognition and Development*, vol. 22, no. 3, pp. 343–355, 2021.

[29] J. D. Lane, H. M. Wellman, and E. M. Evans, "Children's understanding of ordinary and extraordinary minds," *Child Development*, vol. 81, no. 5, pp. 1475–1489, 2010.

[30] J. D. Lane, H. M. Wellman, and E. M. Evans, "Approaching an understanding of omniscience from the preschool years to early adulthood," *Developmental Psychology*, vol. 50, no. 10, pp. 2380–2392, 2014.

[31] S. Eisen and A. S. Lillard, "Young children's thinking about touchscreens versus other media in the US," *Journal of Children and Media*, vol. 11, no. 2, pp. 167–179, 2017.

[32] F. Wang, Y. Tong, and J. H. Danovitch, "Who do I believe? Children's epistemic trust in internet, teacher, and peer informants," *Cognitive Development*, vol. 50, pp. 248–260, 2019.

[33] P. L. Harris, M. A. Koenig, K. H. Corriveau, and V. K. Jaswal, "Cognitive foundations of learning from testimony," *Annual Review of Psychology*, vol. 69, no. 1, pp. 251–273, 2018.

[34] E. Beneteau, Y. Guan, O. K. Richards et al., "Assumptions checked: how families learn about and use the echo dot," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 4, no. 1, pp. 1–23, 2020.

[35] C. Oranç and A. Ruggeri, ""Alexa, let me ask you something different" Children's adaptive information search with voice assistants," *Human Behavior and Emerging Technologies*, vol. 3, no. 4, pp. 595–605, 2021.

[36] R. Garg and S. Sengupta, "He is just like me: a study of the long-term use of smart speakers by parents and children," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 4, no. 1, pp. 1–24, 2020.

[37] Y. Xu and M. Warschauer, "Young children's reading and learning with conversational agents," in *Extended abstracts of the 2019 CHI conference on human factors in computing systems*, Scotland, UK, 2019.

[38] Y. Yuan, S. Thompson, K. Watson et al., "Speech interface reformulations and voice assistant personification preferences of children and parents," *International Journal of Child-Computer Interaction*, vol. 21, pp. 77–88, 2019.

[39] J. Short, E. Williams, and B. Christie, *The social psychology of telecommunication*, John Wiley, London, 1976.

[40] R. E. Mayer, "Principles based on social cues in multimedia learning: personalization, voice, image, and embodiment," in

The Cambridge Handbook of Multimedia Learning, R. E. Mayer, Ed., pp. 345–370, Cambridge University Press, 2014.

[41] F. Wang, W. Li, and T. Zhao, "Multimedia learning with animated pedagogical agents," in *The Cambridge Handbook of Multimedia Learning*, R. E. Mayer and L. Fiorella, Eds., pp. 452–462, Cambridge University Press, 2022.

[42] F. Wang, W. Li, R. E. Mayer, and H. Liu, "Animated pedagogical agents as aids in multimedia learning: effects on eye-fixations during learning and learning outcomes," *Journal of Educational Psychology*, vol. 110, no. 2, pp. 250–268, 2018.

[43] Y. Xu, D. Wang, P. Collins, H. Lee, and M. Warschauer, "Same benefits, different communication patterns: comparing children's reading with a conversational agent vs. a human partner," *Computers & Education*, vol. 161, article 104059, 2021.

[44] S. Aeschlimann, M. Bleiker, M. Wechner, and A. Gampe, "Communicative and social consequences of interactions with voice assistants," *Computers in Human Behavior*, vol. 112, article 106466, 2020.

[45] S. Schneider, M. Beege, S. Nebel, L. Schnaubert, and G. D. Rey, "The cognitive-affective-social theory of learning in digital environments (CASTLE)," *Educational Psychology Review*, pp. 1–38, 2021.

[46] Y. L. R. Chassiakos, J. Radesky, D. Christakis, M. A. Moreno, and C. Cross, "Children and adolescents and digital media," *Pediatrics*, vol. 138, no. 5, article e20162593, 2016.

[47] E. Beneteau, A. Boone, Y. Wu, J. A. Kientz, J. Yip, and A. Hiniker, "Parenting with Alexa: exploring the introduction of smart speakers on family dynamics," in *Proceedings of the 2020 CHI conference on human factors in computing systems*, New York, NY, USA, 2020a.

[48] A. Hiniker, A. Wang, J. Tran et al., "Can conversational agents change the way children talk to people?," in *Interaction Design and Children*, pp. 338–349, New York, NY, USA, 2021.

[49] E. Beneteau, O. K. Richards, M. Zhang, J. A. Kientz, J. Yip, and A. Hiniker, "Communication Breakdowns between Families and Alexa," in *Proceedings of the 2019 CHI conference on human factors in computing systems*, New York, NY, USA, 2019.

[50] A. Hoffman, D. Owen, and S. L. Calvert, "Parent reports of children's parasocial relationships with conversational agents: trusted voices in children's lives," *Human Behavior and Emerging Technologies*, vol. 3, no. 4, pp. 606–617, 2021.

[51] A. Sciuto, A. Saini, J. Forlizzi, and J. I. Hong, ""Hey Alexa, what's up?": studies of in-home conversational agent usage," in *Proceedings of the 2018 Designing Interactive Systems Conference Pages*, pp. 857–868, Hong Kong, China, 2018.

[52] S. Bay, *Children's smart speaker full inventory, segmentation of the market melee war fierce fight*, 2018, https://zhuanlan.zhihu.com/p/37492803.

[53] X. Li and Y. Sung, "Anthropomorphism brings us closer: the mediating role of psychological distance in user–AI assistant interactions," *Computers in Human Behavior*, vol. 118, article 106680, 2021.

[54] J. H. Danovitch and R. L. Severson, "Children's understanding of emerging technologies: introduction to the special issue," *Human Behavior and Emerging Technologies*, vol. 3, no. 4, pp. 464–467, 2021.

[55] A. J. Lucas, C. Lewis, F. C. Pala, K. Wong, and D. Berridge, "Social-cognitive processes in preschoolers' selective trust: three cultures compared," *Developmental Psychology*, vol. 49, no. 3, pp. 579–590, 2013.

[56] M. A. Sabbagh and D. A. Baldwin, "Learning words from knowledgeable versus ignorant speakers: links between preschoolers' theory of mind and semantic development," *Child Development*, vol. 72, no. 4, pp. 1054–1070, 2001.

[57] H. M. Wellman and D. Liu, "Scaling of theory of mind tasks," *Child Development*, vol. 75, no. 2, pp. 523–541, 2004.

[58] Y. Tong, F. Wang, and J. H. Danovitch, "The role of epistemic and social characteristics in children's selective trust: three meta-analyses," *Developmental Science*, vol. 23, no. 2, article e12895, 2020.