Improving Preschoolers' Theory of Mind Skills With Mobile Games

Mariya Nikolayev¹*, Anya S. Evmenova¹, Stephanie M. Reich², Kevin A. Clark¹ and M. Susan Burns¹

¹ College of Education and Human Development (CEHD), George Mason University, Fairfax, VA, United States, ² School of Education, University of California, Irvine, Irvine, CA, United States

This single-case research study examined whether interactive touch screen apps enriched with Theory of Mind (ToM)—enhancing language would promote ToM skills in preschoolers. Six typically developing girls between the ages of 46- and 52-months participated in multiple sessions across the three phases of the study: In baseline, participants played games without voice-overs; in the original treatment phase, participants played games with embedded voice-overs; finally, in the modified treatment phase, participants first played games with embedded voice-overs, then engaged in the researcher-led conversation. All sessions across the three phases concluded with ToM assessments: two measures based on a continuous scale. The first measure included three tasks targeting earlier-developing ToM skills (diverse desires, diverse beliefs, and knowledge access), and the other measure had two tasks that assessed a later-developing ToM competency, false belief understanding. Results showed that apps with ToM-embedded language improved children's earlier-developing ToM skills (i.e., understanding that people can have different desires, beliefs, and knowledge access) in the phase where an adult-led conversation also followed voice-over-enriched app play. Apps with ToM-embedded language without a follow-up discussion were only marginally effective in promoting the earlier-emerging ToM skills. Across the conditions, apps were not effective in promoting children's later-developing ToM skills— false belief understanding. Our findings indicate that incorporating ToM conducive language in mobile apps can promote ToM development in preschoolers, especially when supplemented by an adult-led conversation.

Keywords: theory of mind, preschoolers, educational mobile apps, intervention, single-case design, joint media engagement, digital adult, language

INTRODUCTION

Theory of Mind (ToM) is an essential area of social-emotional development that enables children to recognize, understand, and predict feelings, intentions, beliefs, and desires of the self and others (Astington, 2003; Keenan, 2003). The level of ToM development across different ages has been found to predict children's positive and negative behaviors. Advanced mastery of ToM skills helps to make and keep friends (e.g., Slaughter et al., 2015), promotes persuasion and leadership skills (e.g., Peterson et al., 2018), contributes to the development of prosocial behaviors (e.g., Imuta et al., 2016), correlates with being liked by teachers (e.g., Slaughter et al., 2002) and being popular among peers (e.g., Fink et al., 2014; Slaughter et al., 2015), and promotes academic achievement (e.g., Lecce et al., 2014; Dore et al., 2018; Florit et al., 2020).

Conversely, delays and poor ToM development are associated with difficulties across developmental stages, e.g., problems in social-emotional functioning in preschool (e.g., Vissers and Koolen, 2016), aggressive behaviors in kindergarten (e.g., Renouf et al., 2010), bullying during middle childhood and teenage years (e.g., Shakoor et al., 2012), and feelings of loneliness in adolescence (e.g., Bosacki et al., 2020).

During ToM development, when learning about mental states of others, most Typically Developing (TD) children across the world follow a standard sequence of skills acquisition, with early abilities serving as precursors for later skills (Wellman and Liu, 2004; Shahaeian et al., 2011), including an understanding of false beliefs (i.e., being able to infer the incorrect belief of another person). Although most TD children demonstrate roughly the same level of false belief understanding by the end of preschool (Wellman and Peterson, 2013), the timing and rate of development, conceptual elaboration, and the degree to which children apply the skills in social situations vary among individual children (Cutting and Dunn, 1999; Wellman et al., 2001; Keenan, 2003; Charness et al., 2019). The variability in ToM skills continues to exist into adolescence (e.g., Caputi et al., 2012; Fink et al., 2015; Hughes and Devine, 2015), differentially contributing to the success of children's social and academic experiences across the life span (e.g., Devine et al., 2016; Dore et al., 2018; Peterson et al., 2018).

Given that this variation in ToM skills acquisition has an important impact on well-being (Hughes and Devine, 2015; Weimer et al., 2021) and the overall effectiveness of intervention programs (Hofmann et al., 2016; Roheger et al., 2022), some researchers suggest that ToM-enhancing programs should be provided not only to remedy ToM delays but also to prevent them (e.g., Hofmann et al., 2016). Supplemental early ToM learning may be especially significant for populations of children who are known to lag behind their peers in ToM development (Holmes et al., 1996; Tarullo et al., 2007; Dessen and de Hollanda Souza, 2014; Devine and Hughes, 2018; Charness et al., 2019).

Currently, most of the existing ToM interventions are provided through face-to-face interactions, with none, to our knowledge, being implemented beyond research, clinical, or school-based settings (see Mori and Cigala, 2015; Hofmann et al., 2016 for a review of the training programs). Furthermore, children who demonstrate deficits in ToM development, such as those from disadvantaged backgrounds (Charness et al., 2019; Ebert et al., 2020), are those who could potentially benefit from intervention programs, but also lack access to these interventions and educational opportunities (Hodgkinson et al., 2017; Griffith et al., 2019).

In 2011 the Global Child Development Steering Group named educational media a promising method for promoting early child development and addressing at-risk child populations (Engle et al., 2011). School districts across the country have invested millions of dollars into new technologies for education (Blackwell et al., 2013). A decade later, the presence of mobile devices became almost universal in the homes of children of all socioeconomic backgrounds as the popularity of educational gaming also increased (Griffith and Arnold, 2019; Rideout and Robb, 2020). This resulted in the production of hundreds of thousands of apps claiming educational benefits (Apple, 2021). The educational apps' prevalence and accessibility makes them a promising scalable method for delivering ToM educational content to various populations of children in need of ToM development support.

In this paper, we sought to examine whether the ToM-stimulating language in educational games for preschoolers can improve children's ToM performance with the support of the game's interactivity and parental engagement. There is evidence that interactivity and parental engagement are beneficial for children's learning of various skills, including coding, math, literacy, and some social-emotional, from digital games, yet whether they impact the ToM skills is unclear due to the lack of research on the topic. The preschool age of the participants was selected because ToM gains are most rapid between 3 and 5 years (e.g., Tompkins et al., 2019), and by four years of age, most children are capable of playing digital games independently but still benefit from adult support (Pempek and Lauricella, 2017 as cited in Bindman et al., 2021). Furthermore, given the potential of digital games to serve as a scalable platform for delivering ToM interventions, we proposed ToM-focused design suggestions to be used by edutainment game developers.

Background and Related Work

The review goes over the existing face-to-face language-based interventions designed to improve preschoolers' ToM skills and discusses scientific literature on children's learning from digital games. Intending to create effective ToM-educational content for mobile games, we focus on the language elements in face-to-face interventions that influence ToM development and various design and contextual factors of digital games that promote children's learning.

Language-Based ToM Interventions

A large body of literature highlights the critical role of language in developing ToM (e.g., see Tompkins et al., 2019 for a review), specifically the socio-linguistic environment and a child's abilities (De Villiers and De Villiers, 2000; Astington and Baird, 2005; Taumoepeau and Reese, 2013; Lillard and Kavanaugh, 2014). The concept of a socio-linguistic environment is based on the notion that language needed for ToM is facilitated by a child's social experiences, such as the content of conversations between parents and children, home literacy environment, and the frequency and content of conversations overheard by children (Astington and Baird, 2005; Slaughter et al., 2007; Ruffman, 2014; Tompkins et al., 2018; Lecce et al., 2021). The socio-linguistic environment and children's language are closely related and contribute both jointly and independently to ToM development (Astington and Baird, 2005). For example, studies have shown that the ToM of children from disadvantaged backgrounds, such as low SES or institutionalized settings, develops slower than children from more advantaged ones (e.g., Tarullo et al., 2007; Charness et al., 2019). Further, Ebert et al. (2020) demonstrated how various SES-related aspects of the home language and literacy environment contribute to children's ToM and language development.

The natural pace of ToM development is slow; children's false belief performance without intervention typically improves only slightly between the ages of 3.5 and 4 (Amsterlaw and Wellman, 2006). However, experimental intervention studies have accelerated children's ToM acquisition in as short a time frame as two weeks (e.g., Slaughter and Gopnik, 1996; Hale and Tager-Flusberg, 2003; Wellman, 2018), demonstrating that it is possible for the interventions to help children who are behind on ToM development to catch up with their more advanced peers.

Various methods have been used to deliver training content in face-to-face language-based ToM interventions (see Mori and Cigala, 2015; Hofmann et al., 2016 for review): some teach caregivers to reminisce about past events, pose questions in a specific way, and incorporate words for feelings, desires, and beliefs to promote children's expressive vocabulary (e.g., Taumoepeau and Reese, 2013; Spruijt et al., 2020). Others have teachers and researchers read storybooks enriched with mental-state language, show videos, or demonstrate puppet shows with the mental-state-verb-laden script (e.g., Esteban et al., 2010; Gola, 2012; Tompkins, 2015; San Juan and Astington, 2017). Finally, some have children complete false-belief tasks and then follow the task with evidence-based or corrective feedback and explanations (e.g., Slaughter and Gopnik, 1996; Clements et al., 2000).

Many studies have found that using ToM-promoting language alone is not enough to accelerate ToM (e.g., Peskin and Astington, 2004; Amsterlaw and Wellman, 2006; Ornaghi et al., 2011). The interactivity seems to be crucial for ToM mastery; in conversations, children observe how their own and other people's perspectives become clear as well as inconsistencies between their own and others' mental states and realities (De Villiers and De Villiers, 2014). Several studies have tried to make the children active participants rather than passive observers by engaging them in language-based activities and discussions containing mentalistic language (Lohmann and Tomasello, 2003; Ornaghi et al., 2011). For example, language that encompasses references to emotional states (e.g., *happy, sad, excited*), mental processes (e.g., *know, think, remember, understand, feel*), desires (e.g., *want, wish, hope*), and modulations of assertion (*e.g., guess, maybe, perhaps*) (Ruffman et al., 2002). Others use storybook readings followed by adult-led discussions and reflections about the mental states and behaviors of characters in the stories (e.g., Guajardo and Watson, 2002; Tompkins, 2015). Despite the common presence of mobile devices in the house of preschoolers and their potential in delivering educational content (e.g., Griffith et al., 2019), no studies, to our knowledge, have leveraged the interactive affordances of digital games to grant the users agency in decision-making and provide them with contingent feedback to teach ToM skills.

Educational Games and Learning

Despite the recognition by the educational and scientific communities of ToM as an essential set of social-emotional skills, it is largely overlooked by designers of educational games for young children (Nikolayev et al., 2015) and, along with other social skills, relatively understudied by digital media researchers (Flynn et al., 2019). The few existing studies of preschoolers and ToM focus on relations between ToM abilities and video

content (Reiß et al., 2019; Cingel et al., 2020), but not the interactive platforms. Only one recent study, that we are aware of, has examined ToM in relation to interactive gameplay in preschoolers, not as a dependent variable, but rather as a moderator between gameplay and prosocial behaviors (Shoshani et al., 2022). Although many educational apps marketed for preschoolers do not use optimal pedagogical approaches and are not rooted in developmental science (Callaghan and Reich, 2018; Meyer et al., 2021; Nikolayev et al., 2021), a growing body of literature demonstrates that digital apps that employ developmentally appropriate content and design elements have the potential to teach preschool children (ages 3–5) a wide variety of skills (Hirsh-Pasek et al., 2015; Herodotou, 2018; Flynn et al., 2019; Griffith et al., 2020; Kim et al., 2021; Papadakis, 2021b; Callaghan and Reich, 2022) including language (Teepe et al., 2017; Neumann, 2018; Dore et al., 2019; Kirsch, 2021), computational thinking (Papadakis, 2022), and executive function skills (Huber et al., 2018).

Beyond the content and design elements, children's learning from digital games is mediated by contextual factors of play (Guernsey, 2007; Takeuchi and Levine, 2014), such as joint media engagement (Takeuchi and Stevens, 2011). Meaningful adult-child co-play supports and enhances young children's learning from educational apps (Neumann and Neumann, 2014; Radesky et al., 2015; Sweeney, 2017; Neumann, 2018; Rasmussen et al., 2019; Toh and Lim, 2021). In the process of digital co-play, adults scaffold children's learning by engaging them in dialogs and explaining complex concepts (Yelland and Masters, 2007; Bindman et al., 2021), directing children's attention to the specific content and highlighting important information (Sobel et al., 2019), providing affection and encouragement (Yelland and Masters, 2007; Wood et al., 2016), and helping with technical and physical tasks, such as logging in, typing, and touching the screen (Reich et al., 2012; Wood et al., 2016). However, studies have found that joint media engagement does not require adults to be co-playing the game for it to be beneficial for learning (Eisen and Lillard, 2020; Musick et al., 2021). For example, Reich et al. (2012) observed children narrating digital gameplay and explaining their choices to playmates who were not actively gaming, and Bers (2020) as cited in Papadakis (2021a) demonstrates how collaboration when learning coding promotes not only computational thinking, but language and social-emotional skills.

Digital Games as Social Partners

Although digital co-play is believed to be instrumental for preschoolers and recommended by experts like the American Academy of Pediatrics (AAP Council on Communications and Media, 2016), children often engage in touchscreen media independently (see Ewin et al., 2021 for review). Promisingly, researchers are increasingly finding that digital games can function as social partners or "more knowledgeable others" to young players (e.g., Richert et al., 2011; Xu et al., 2021). Similarly, Xu et al. (2021) and Russo-Johnson et al. (2017) have demonstrated the ways in which interactivity *via* an app or conversational agent can support vocabulary learning and story comprehension, suggesting that interactive technology could

also facilitate ToM abilities. Recent work finds that artificial-intelligent (AI) agents (e.g., Alexa) supports preschoolers' deeper thinking and understanding (Xu et al., 2021). Xu et al. (2021) applied a sociocultural approach to create AI-mediated experiences to support children's language development through dialogic reading. They found that AI-powered conversational agents can indeed function as "more knowledgeable others" and provide the same benefits as dialogic reading with adult human partners.

Flynn et al. (2019), applied play framework by Zosh et al. (2018) to digital spaces and theorized that when design affords a specific type of interactivity, a game can assume the role of a "digital adult" and, in turn, may provide the benefits of adult-child co-engagement. One such type is contingent interactivity, also known as full interactivity in some studies (Peebles et al., 2018), which involves meaningful reciprocal exchanges between the player and the system and includes turn-taking, responsive contingent feedback, and device control. In contingent interactivity, the game assumes the role of a "digital adult" and initiates some activities within the game or directs the play.

A digital home literacy environment, i.e., shared and independent literacy activities using a digital device is a large part of children's everyday life (Segers and Kleemans, 2020). Preschoolers across different social-economic backgrounds benefit from educational app use (Arnold et al., 2021; Rowe et al., 2021) and spend, on average, at least 40 min daily on mobile devices (Rideout and Robb, 2020). From content access to co-engaging in digital use (see Papadakis et al., 2021 for a review), caregivers shape children's interactions with technology and differentially influence learning from educational media. In recent years, studies have examined scaffolding of children's digital learning by adults and the possibility of scaffolding by "digital adults" in the form of contingent interactivity. In this paper, we built on this premise. We designed language for apps to create a socio-linguistic environment to help promote children's ToM development with the support of games' contingent interactivity and real-life adult conversation.

PRESENT STUDY

Drawing on the promise that interactive technology can serve as a social partner, the present study examines whether an interactive, touch screen app that utilizes language known to promote children's ToM skills through face-to-face interventions and real-life interactions could be effective in boosting preschoolers' ToM skills.

Specific research questions were the following:

RQ1: Is there a functional relation between the use of digital apps with ToM -promoting language and children's understanding that people have different desires, beliefs, and knowledge sources?

RQ1a: Does children's understanding that people have different desires, beliefs, and knowledge sources increase when the use of games enriched with ToM-promoting

language is followed by an adult-led discussion about the games?

RQ2: Is there a functional relation between the use of digital apps with ToM-promoting language and children's understanding of false belief and knowledge sources?

RQ2a: Does children's false belief understanding increase when the use of games enriched with ToM-promoting language is followed by an adult-led discussion about the games?

MATERIALS AND METHODS

Single-case methods have been commonly employed in special education settings for many years and have been recognized as especially appropriate and valuable for identifying evidence-based practices in education research (Odom and Strain, 2002; Horner et al., 2005; Kratochwill et al., 2021). One of the benefits of a single-case research method is that it allows researchers to respond to individual differences and implement intervention modifications if needed (Ledford and Gast, 2018), making it especially valuable in researching technology-delivered personalized education. A multiple-baseline single-case design was used in this study for several reasons:

1. ToM skills are irreversible thus, withdrawal is not possible.
2. The design makes it possible to measure target responses to multiple assessments.
3. The design allows for control due to developmental maturation, which is important given that children's ToM skills improve with age.

There were three phases (conditions) in this study in the sequence of A—B—B+C, where A was the Baseline phase, B was the Voice-overs treatment phase, and B+C was Voice-overs combined with Discussion (VAD) modified treatment phase. Following the multiple-baseline across participants research logic, children were introduced to treatment in a staggered fashion to ensure that changes in the data patterns were due to the introduction of the treatment and did not have alternative explanations such as the multiple exposures to the assessment procedures or maturation (Ledford and Gast, 2018). In a multiple-baseline study, the experimental control and the functional relation between dependent and independent variables are established when participants' performance (e.g., level, trend, variability of data) changes only after they are introduced to treatment, while the performance of participants in baseline remains unchanged. The present study met the single-case design standards outlined by Kratochwill et al. (2021).

Participants

The participants for this study were selected from a preschool that serves low-income families in a Mid-Atlantic metropolitan area in the United States. Each participant was assigned a unique pseudonym, and all student identifying information was removed to maintain the confidentiality of the participants. Potentially identifying information about the preschool in which

the study was conducted was purposefully eliminated from the description. This research study was reviewed and approved by the Institutional Review Board (IRB) at George Mason University (GMU) to ensure the rights and welfare of the study participants. Parental consent/student assent was obtained prior to the beginning of the study. Permission was also obtained from the principal of the preschool.

For the preliminary selection, preschool director and classroom teachers were asked to identify typically developing children who were between 3 and 5 years of age and fluent in English from the pool of the children whose parents had consented to their child's participation in the research study. Fourteen children (eight girls, 6 boys) were recommended by the teachers for the stud. The first screening involved playing one of the five games without voice-overs and undergoing the Theory of Mind (ToM) assessment. Nine children who failed three or more of the five tasks in the ToM assessment were selected to participate in the intervention study. The data collection happened over the summer, and three children dropped out during the study for attendance reasons: two went on vacation during baseline and one of the phases of the testing, and one was sick for an extended period of time. The final sample included six girls between 42 and 54 months of age (see **Table 1**).

Mia was an only child. Social workers and teachers described Mia as social and talkative, able to explain her feelings, although with some challenges in interpersonal communications. Specifically, Mia seemed to be led by a friend who would not let her play with other children. Isabella came from a large Spanish-speaking, multigenerational family where she was "the baby" of the family. Teachers reported that Isabella appeared to be skilled in having relationships with adults but had trouble making and sustaining friendships with peers. Sienna was a very energetic, happy, and assertive girl. She was an only child of a single mother. Sienna and her mother spoke Spanish exclusively when together. Paula had an older sister, and her family spoke English and Spanish. Teachers described Paula as a well-liked girl who was very calm and patient but was constantly tired. Camilla came from a Spanish-speaking family and had an older brother. She was a friendly child who played well with others and made friends quickly. She could often be observed engaged in group play. Teachers described

TABLE 1 | Description of participants.

Name	Gender	Ethnicity	Age (in months)	Second language
Mia	Female	African American	49.5	Shona
Isabella	Female	Caucasian	51.5	Spanish
Sienna	Female	Hispanic	46	Spanish
Paula	Female	Hispanic	48.5	Spanish
Camilla	Female	Hispanic	46	Spanish
Emily	Female	Asian	48	Korean

In this study, a second language refers to any language other than English that children are exposed to on an everyday basis. Some participants exclusively spoke another language at home, whereas others only occasionally heard parents speaking another language.

Emily as quiet and reserved but friendly. Emily usually played with one friend, a quiet girl, who was beginning to learn English. Emily was the youngest of five children and the only girl in her family.

Data Collection Procedures

Children were visited three to five times a week for five weeks, for a total of 19–20 sessions. All children were trained and tested individually in a private room at the preschool. Only one researcher (the first author) collected the data. The researcher brought each child to the testing room (preschool library) and briefly went over the procedures and solicited verbal assent. All sessions were video recorded. Once the child was done with the session, the researcher brought them back to their class. Due to the single-case method, each child went through the three phases of testing in the same order (Baseline, Voice-over, VAD), but started treatment phases in a staggering fashion, e.g., participants 1 and 2 started treatment on the seventh visit, whereas participants 3 and 4 started on the eighth. According to the single case design standards, each phase for each participant should have at least five data points; and it's appropriate to change to the next phase of the intervention when data show stable patterns (Kratochwill et al., 2021). Each session, including Baseline condition and both types of treatment conditions (original and modified VAD), started with the child playing a game on an iPad and concluded with face-to-face assessment procedures. The sessions were numbered, and a specific game was assigned to each session.

Baseline Procedures (Phase I)
Baseline procedures began with participants playing one of the five LEGO® DUPLO® game apps on the iPad while wearing headphones connected to the researcher's computer. The computer volume was muted, and no sounds other than those from the iPad were played to the participants. The researcher sat across from the child at an angle that allowed her to see the child's screen and be available to help with the procedures and technical aspects of the play, but she did not initiate or support discussions about the game plot.

Voice-Over Training Procedures (Phase II)
During the voice-over training procedures, the participant and researcher sat at an angle to each other, each with their own device. The researcher could see the participant's iPad's screen, but the children could not see the researcher's laptop screen. Both the researcher and the participant wore headphones connected to the researcher's computer to hear the verbal component of the game. At the beginning of the first session in the original treatment phase, the researcher explained to the participants that she had forgotten to turn on the "sound" in the games before and that from now on, children would play with the turned-on sound. None of the participants expressed any concern or suspicion that the sound came from the researcher's computer and not the games. The sound in the games did not seem to change children's enjoyment of the game. The child played the game, and the researcher followed the gameplay and started the voice-overs from her laptop at

specific times. The game's original music and sounds were not muted on the iPad so that the gameplay would feel more natural. If a child missed a step, the researcher skipped the accompanying voice-over and introduced the next one at the appropriate time.

Voice-Overs and Discussion (VAD) Training Procedures (Phase III)

The VAD training procedures were identical to the voice-over treatment procedures, with one exception; right after the gameplay and before the assessment activities, the researcher engaged the child in a semi-structured conversation about the game. Depending on the child, each discussion lasted between 3 and 7 min. Children enjoyed the conversations and were eager to participate. After the discussion was over, the researcher started the ToM assessment procedures.

Independent Variable and Materials
Games

Five LEGO® DUPLO® game apps were used: LEGO® DUPLO® Circus, LEGO® DUPLO® Ice Cream, LEGO® DUPLO® ZOO, LEGO® DUPLO® FOREST, LEGO® DUPLO® FOOD®. LEGO® DUPLO® apps are distributed internationally, and as such, use sound effects and background music, rather than words, to be accessible to children who speak different languages. The games have a storyline and a goal for completion (e.g., help deliver a package), and require children to complete several mini games focusing on prosocial behaviors, social interactions, and decision-making.

Independent Variable: Voice-Overs

The voice-overs were designed based on research from the extant literature on verbal interactions that promote ToM. They included: (a) explanatory, causal, and contrastive talk about mental states (e.g., *Everyone in the audience thinks the acrobat may fall from the swing, but she knows she won't.*); (b) an abundance of mental verbs, specifically verbs referring to mental processes (e.g., *think, know, and remember*) that scaffold preschoolers' transition to belief-based thinking, and verbs of desire (e.g., *like, want*) to accommodate younger children who are still transitioning from desire-based to belief-based explanations of behaviors; (c) mental state verbs along with embedded sentential complements structures (e.g., *Bunny and Teddy think there is a green rock*, but click on it [wait for the child to click]—*it is really a turtle!*); (d) explanations of mental states underlying characters' behaviors (e.g., *The driver did not stop to help because he thinks you can put the food away all by yourself*); (e) references to events that occurred earlier in the game (e.g., *Guess what, Giraffe. Remember the Lion didn't see you getting the package? It means he does not know that you have it*); and (f) mental-state verbs directed at players were incorporated into statements, whereas utterances directed at other characters in the video were incorporated into questions (e.g., *Remember how the squirrel thought the box is full of candy and nuts? It turns out there was a DRUM inside*. vs *Giraffe, do you know what everyone likes?*). The voice-overs were embedded in narration: the narrator made

explicit positive assumptions about children's thought process regarding false belief situations presented in the games and commented on children's and characters' performance (e.g., "*You thought these were regular stars, but they are actually musical stars*"). Additionally, voice-overs were included in contingent feedback (e.g., *We think you are such a great builder; you made an awesome forest door!*) and the dialog between the characters (e.g.,—Giraffe*, do you think everyone saw the horse jumping through the fire?—Of course, Bunny, everyone thinks the trick was awesome*).

Voice-overs were recorded using the audio editor Audacity (Audacity Team, https://www.audacityteam.org) and embedded into a presentation slideshow as audio files with captions. Each slide corresponded to a different screen in the game and contained voice-overs for all the possible game scenarios so that the researcher could observe children playing and provide contingent feedback.

Modified Independent Variable: Voice-Overs With Discussion (VAD)

Following the first treatment phase using voice-overs, an additional treatment of voice-overs with follow-up discussions was introduced. Although conversations were structured around the game narrative, the interaction between participant and researcher (the first author) resembled a naturally occurring conversation and were different for each child. The researcher asked questions to help the child reconstruct narrative plots from the games, highlighted and repeated episodes that contained mental-state references and exchanges between characters, and engaged the child in discussing instances of deception and false beliefs. For example, in one session after the child played the LEGO DUPLO Circus game, the researcher began with the initial question, asking the child to recall that the circus came to town in the game, then asked about the characters' expectations about the show and performers and whether these expectations were met. Next, the researcher asked the child about the mental states underlying the characters' behaviors (e.g., *Why did the clown run away? What did he think about the tiger?*), and finally prompted the child to describe the circus audience's thoughts. When children injected their experiences into the conversation, the researcher supported them and then returned to discussing the events in the game. The researcher did not correct the participants if they made mistakes in attributing the false belief or misremembering details. Instead, the researcher prompted children to talk more about the scenario to help them think through the conflicting perspectives between the characters' expectations and reality.

ToM Dependent Variables

ToM was assessed using variations of tasks from the five-item developmental scale created by Wellman and Liu (2004) and the Location Change task developed by Wimmer and Perner (1983). Given that children's performance on False Belief tasks remains consistent across different task presentation formats and different types of tasks (e.g., Wellman et al., 2001; Hasni et al., 2017), to keep participants engaged, half of the tasks were presented

in a digital storybook format (created with iPad drawing apps), and the other half acted out with props. Presentation order was counterbalanced across sessions for task type and format.

Following the Gola's (2012) study, tasks were grouped into two categories. In the first category, three tasks assessed children's understanding that people can have diverse desires, beliefs, and knowledge about the same thing; and in the second category, two tasks assessed False Belief understanding. All tasks corresponded to a progression of milestones in children's development of ToM (the two False Belief Tasks are of similar difficulty), thus, the first category contained conceptually easier tasks than the second one. In the study these categories were used as two separate measures.

Measure 1: Desires, Beliefs, Knowledge. Three tasks assessed children's understanding that people can have different desires, beliefs, and knowledge access. These three skills were judged as either correct or not (1 = correct, 0 = incorrect). The scores were then added together for the Desires, Beliefs, Knowledge score, with the total score ranging from 0 to 3. The Desires, Beliefs, and Knowledge scenarios were presented randomly to prevent children from expecting the same order of items across the sessions.

- In the Diverse Desires task, a child must demonstrate an understanding that someone might have a different desire about the same object. The child is presented with a doll and pictures of two different snacks. The researcher asks for the child's preference of snack and subsequently states that the doll wants a different snack than the one selected by the child. The child is then asked which snack the doll would choose; the child must provide an answer to the target question that is different than what they desire.
- In the Diverse Beliefs tasks, a child must demonstrate an understanding that someone might hold a different belief about the same thing. The child is shown a doll and pictures of a garage and bushes. The child is told that Linda is looking for her cat; the researcher then asks the child where they think the cat is, in the bushes or in the garage. The researcher then says that Linda believes her cat is in a different location than indicated by the child and asks the child where Linda would look for her cat. The child must say the opposite of their belief.
- In the Knowledge Access (Seeing-Knowing) task, a child must correctly judge the knowledge of another person who does not have access to the information available to the child. The child is presented with a small box and asked what they think is in the box. After the child guesses or says that they do not know, the researcher lets the child open the box to see the contents (a Lego piece). The researcher then introduces a doll and says a doll has never seen inside the box and asks whether the doll knows what's inside the box. The child must say "No" to be correct.

Measure 2: False Belief Three tasks, Unexpected Contents False Belief and Explicit False Belief (Wellman and Liu, 2004), and Location Change task (Wimmer and Perner, 1983) were used to measure children's false belief understanding, with two different tasks used per session.

- In the Contents False Belief task, a child must reason how another person might misjudge the contents of a container. The child is provided with a familiar, easily identifiable container (e.g., a box of crayons) and is asked to guess what the contents are. After the child answers, "crayons," the box is opened, and a small wooden hippopotamus is revealed. The researcher then puts the toy back into the box and closes the lid. A doll appears, and the researcher states that the doll has not seen inside the box and asks the child what a doll thinks is in the box. The correct response to the question is "crayons."
- In the Explicit False Belief task, the child must decide where one would look for an object given the one's incorrect belief. The child is presented with a doll and two pictures, one of a backpack and another of a closet. The researcher then explains that a doll is looking for his mittens that are really in his backpack, but he thinks they are in his closet. The researcher then asks where a doll is going to look for his mittens. The correct response is in the "closet."
- In the Location Change Task children must decide where someone will be looking for an object given agent's information about the location. The task is based on a story of character A, who places an object (e.g., a book) in a specific location (e.g., a cabinet) and then leaves. Meanwhile, unbeknownst to character A, character B moves the object to a different location (e.g., a bookshelf), and character A then reappears. The child's task is to identify where character A will look for the object first. To be correct, the child must answer that character A will look in the original location (before the move).

To prevent children from getting used to solving the same type of scenario, each session contained either an Explicit False Belief or the Location Change Task. The tasks were randomly assigned to each session. All together there were two False Belief tasks per session and children received 1 point for each correct answer and could receive 0–2 points overall.

Reliability and Scoring
Procedural Reliability
An independent observer, trained in the procedures before the data collection, monitored session activities and compared them against a preplanned checklist of expected activities. Procedural reliability data were collected for 30% of the data for all participants across all three conditions (Baseline, Voice-overs, VAD). The number of correct actions was then divided by the number of planned actions and multiplied by 100%, yielding procedural reliability of 100%.

Interrater Agreement on ToM Outcomes
To ensure scoring reliability, an independent observer scored 30% of the assessment sessions. The independent observer was a child development professional with an extensive experience in experimental research. Inter-observer agreement was assessed for 33–35% of the observations of Desires, Beliefs, Knowledge (desires, beliefs, knowledge) scores and False Belief (False Belief) scores in the Baseline, Voice-overs, and VAD phases. The Total

Agreement formula was used to calculate interrater agreement; a smaller total of correct answers recorded by each observer was divided by the larger total and multiplied by 100% (Kennedy, 2005). The mean interobserver coefficient of agreement for Desires, Beliefs, Knowledge (desires, beliefs, knowledge) was 92% (range: 87–100%) for all participants. The average agreement for False Belief was 96% (range: 75–100%) for all participants. Thus, in most individual instances as well as in the group averages, the design standard for inter-rater agreement was met (Kratochwill et al., 2021).

Analytic Plan

Visual analysis of graphed data was used to examine the functional relation between voice-overs in the games, combination of VAD, and changes in participants' performance on Desires, Beliefs, Knowledge and False Belief (the latter two values are the scores per session). Specifically, we used the following procedure outlined by Kratochwill et al. (2013) to visually examine within and across phases changes in a) level (mean of all data points within the phase), b) trend (direction of the data slope), c) data variability (instability of data), d) immediacy of effect (degree of change from the last data point in one phase and the first data point in the next phase), and e) an index of data overlap between phases – Non-overlap of All Pairs (NAP) (Parker and Vannest, 2009; Manolov et al., 2016)—for each participant's Desires, Beliefs, Knowledge and False Belief data. The visual analysis allows for determination of (a) evidence of a functional relation between dependent and independent variables; as well as (b) the magnitude of that relation (Kratochwill et al., 2021). The decision is based on the changes within- and between-phases on six components of the visual analysis. NAP was also used to calculate the percentage of data that improved across participants for each measure. According to Kratochwill et al. (2021), there is strong evidence of a functional relation if at least three demonstrations of an effect are present at different time points; moderate evidence of a functional relation if at least three demonstrations of an effect are present with at least one demonstration of a non-effect; and no evidence of a functional relation if there are not at least three demonstrations of an effect.

Non-overlap of All Pairs

Though different indices of overlap exist, several studies (Manolov et al., 2011; Parker et al., 2011) demonstrate an advantage of a NAP index. It is derived from a non-parametric assessment procedure that involves individual comparison of all A to B data points and provides a percentage of all non-overlapping data points. NAP is appropriate for many different data types and distributions and is less susceptible to outliers than some other indices of data overlap (Parker and Vannest, 2009). The NAP is equivalent to the Mann–Whitney U statistic and ranges from 0 to 1, with 0.50 indicating a null effect of the treatment or a complete overlap between the baseline and intervention phases (Mann and Whitney, 1947; Parker and Vannest, 2009; Michiels et al., 2018). Values above or below 0.50 indicate improvement or regress in performance in the treatment phase in comparison to the baseline, with increasing degrees of

non-overlap (Parker et al., 2011; Berrett and Carter, 2018). We calculated nap with an online calculator available at http://www.singlecaseresearch.org/calculators/nap (Vannest et al., 2011).

RESULTS

Diverse Desires, Diverse Beliefs, Knowledge Access Skills

Based on the changes in such components of the visual analysis (level, trend, variability, overlap, and consistency), two of the six participants increased their Desires, Beliefs, Knowledge scores during the Voice-over phase as compared to the Baseline, and all participants increased their scores during the VAD phase as compared to Baseline (see **Figure 1**). The visual analysis of Desires, Beliefs, Knowledge data demonstrated no evidence of the functional relation between Voice-overs training and improvements in Desires, Beliefs, Knowledge skills (Kratochwill et al., 2021) and moderate evidence of the functional relation between VAD treatment and improvements in Desires, Beliefs, Knowledge skills development. Mean Non-overlap of All Pairs (NAP) across participants was calculated to be 0.66 for Voice-overs phase and 0.87 for VAD phase. Individual Desires, Beliefs, Knowledge results for both treatments are described subsequently.

Mia

In all six Baseline sessions, Mia's data demonstrated a low level ($M = 1.50$, SD = 0.54), accelerating trend, and moderate levels of variability (see **Figure 1**). Upon introduction of the Voice-overs training, Mia showed a small increase in level (from Baseline $M = 1.50$, SD = 0.54 to Voice-overs Training $M = 1.8$, SD = 0.84), no immediacy of effect, a flat trend, and high variability of data (**Figure 1**). NAP for Mia's Desires, Beliefs, Knowledge data was calculated to be 0.60 from Baseline to Voice-overs phase. In response to the VAD phase, Mia's scores increased from Baseline ($M = 1.50$, SD = 0.55) to VAD ($M = 2.33$, SD = 0.71), with almost half of the answers at the ceiling level (**Figure 1**). Desires, Beliefs, Knowledge data in the VAD phase had an upward trend, high variability, and no immediacy of effect. There was also an increase in level and change in trend: from the Voice-overs phase level ($M = 1.8$, SD = 0.83) to VAD phase level ($M = 2.33$, SD = 0.70), and from the flat trend in Voice-overs to an upward trend in the VAD. NAP for Mia was calculated to be 0.81 from Baseline to VAD phase.

Isabella

Across the six Baseline sessions, Isabella had mid-range scores ($M = 1.83$, SD = 0.4), with a flat trend and low variability of data. With the implementation of the Voice-overs training, Isabella's Desires, Beliefs, Knowledge data showed a small change in level from Baseline ($M = 1.83$, SD = 0.41) to Voice-overs phase ($M = 2$, SD = 0), no immediacy of effect, flat trend, and absence of variability (**Figure 2**). A NAP of 0.58 was calculated from the Baseline to Voice-overs treatment phase.

In the VAD phase, Isabella showed above-baseline performance with level increase from Baseline ($M = 1.83$, SD = 0.4) to VAD ($M = 2.22$, SD = 0.66), an accelerating trend,
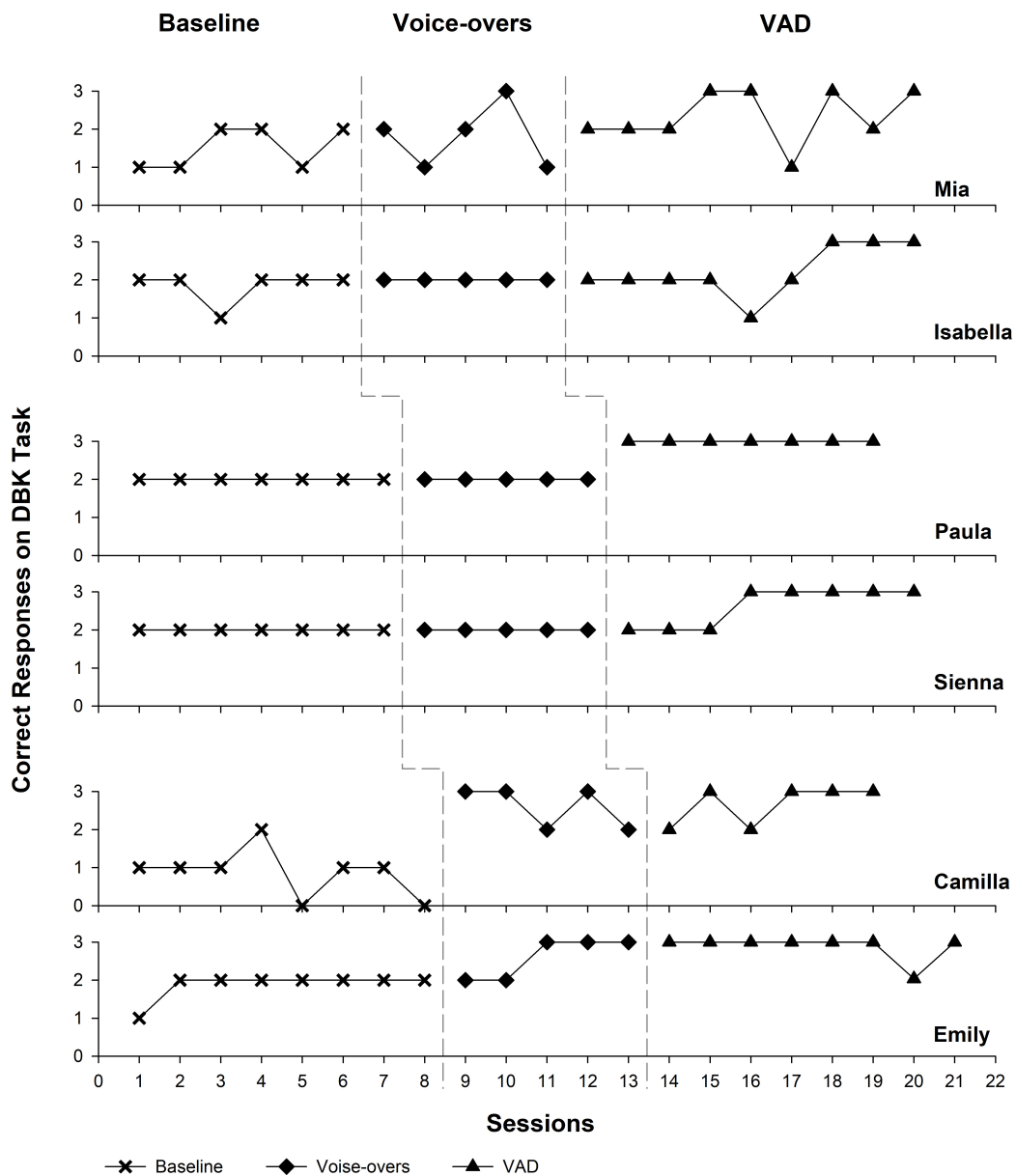
FIGURE 1 | Accuracy of responses to the three tasks in Desires, Beliefs, Knowledge measure by participants across the research phases.

and moderate variability of data (**Figure 2**). Isabella's level increased only slightly from the Voice-overs phase (from Voice-overs phase $M = 2.00$, SD = 0 to VAD phase $M = 2.33$, SD = 0.66) and showed no immediacy of effect. The greatest amount of change was observed in the trend direction that improved from being flat in the Voice-overs phase to accelerating in the VAD phase. A NAP of 0.67 from the Baseline to the VAD treatment phase was calculated for Isabella.

Paula

In seven Baseline sessions, Paula consistently solved two, Diverse Desires and Diverse Beliefs, of the three Desires, Beliefs, Knowledge questions correctly (see **Figure 1**), but not the

Knowledge Access questions. Her data showed medium level ($M = 2.00$, SD = 0), no variability, and flat trend. There was no change in Paula's response to the implementation of the Voice-overs phase: level remained the same ($M = 2.00$, SD = 0), no immediacy of effect was observed, there was the absence of variability and a flat trend also remained. A Desires, Beliefs, Knowledge NAP of 0.50 from Baseline to Voice-overs treatment phase was calculated.

Upon the implementation of VAD training, Paula demonstrated a rise in level from Baseline ($M = 2.00$, SD = 0) to VAD ($M = 3.00$, SD = 0), immediacy of effect, flat trend, and no variability of data. In other words, Paula immediately reached the ceiling in her responses and remained there for all seven VAD
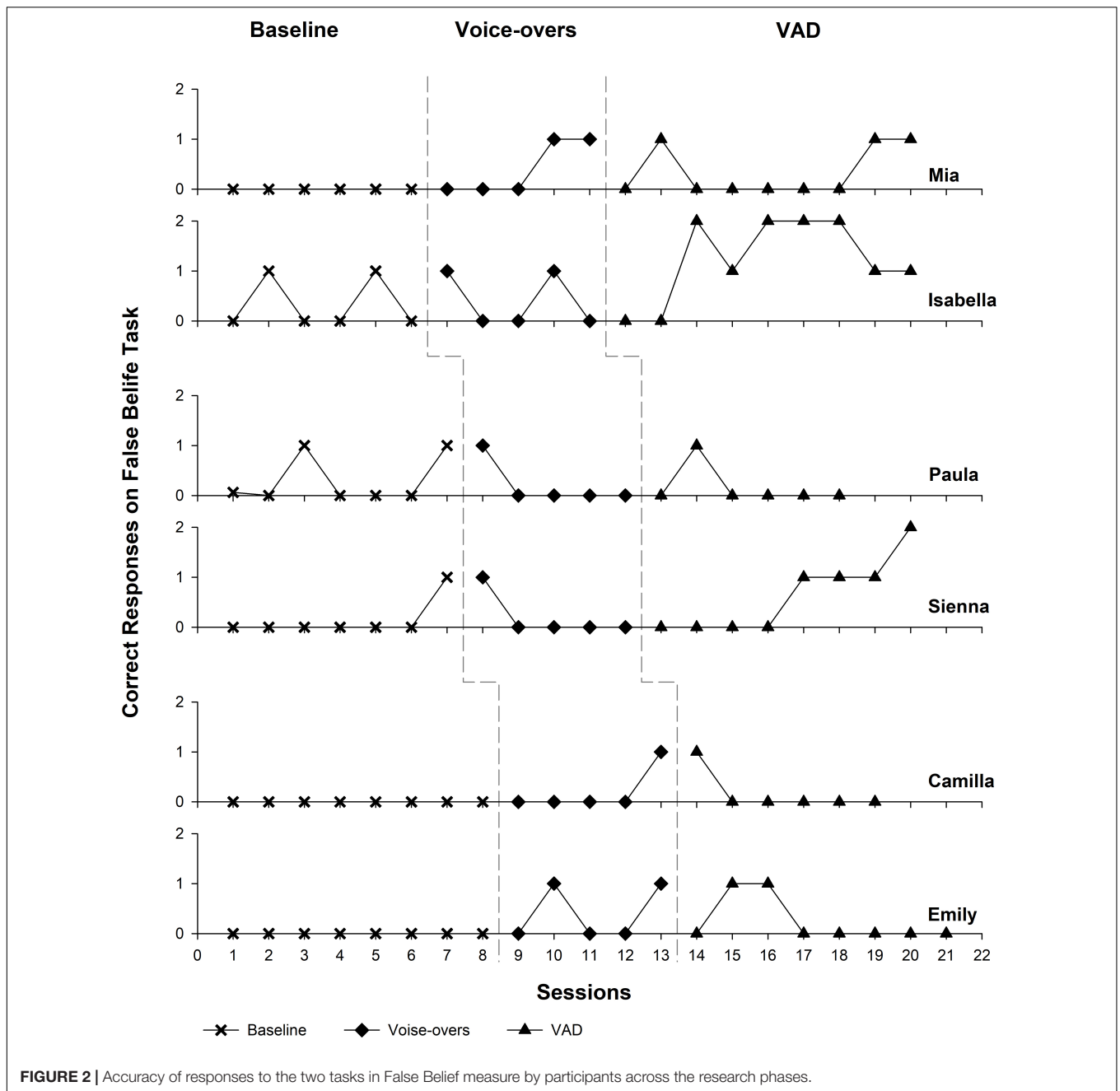
FIGURE 2 | Accuracy of responses to the two tasks in False Belief measure by participants across the research phases.

training sessions. A NAP of 1.00 from Baseline to VAD phase was calculated for Paula's Desires, Beliefs, Knowledge data.

Sienna

During seven Baseline sessions, Sienna consistently answered two questions correctly, Diverse Desires and Diverse Beliefs (see **Figure 1**) showing mid-level scores ($M = 2.00$, SD $= 0$), flat trend, and no variability. Sienna did not answer the Knowledge Access questions correctly, and her Desires, Beliefs, Knowledge showed no change in response during implementation of the Voice-overs phase; level remained the same ($M = 2$, SD $= 0$), as did the absence of variability and a flat trend. A NAP of 0.50 from the Baseline to

Voice-overs treatment phase was calculated for Desires, Beliefs, Knowledge data.

In the VAD treatment phase, Sienna's data (**Figure 1**) had a increase in level from Baseline ($M = 2$, SD $= 0$) to VAD ($M = 2.63$, SD $= 0.51$), no immediacy of effect, a steep accelerating trend, and a moderate variability of data. Since Sienna's performance on the Desires, Beliefs, Knowledge measure was identical during the Baseline and Voice-overs phases, the same changes in level, trend, and data variability were observed from Baseline to VAD phases and from Voice-overs to VAD phases. A NAP 0.81 from Baseline to VAD phase was calculated on Sienna's Desires, Beliefs, Knowledge measure.

Camilla

Throughout eight Baseline sessions, Camilla had low scores ($M = 0.88$, SD $= 0.64$), with data showing a downward trend and moderate variability. Upon the introduction of the Voice-overs phase, Camilla's data (**Figure 1**) demonstrated an increase in level: from Baseline ($M = 0.88$, SD $= 0.64$) to Voice-overs Training ($M = 2.60$, SD $= 0.55$), immediacy of effect, downward trend, and moderate variability. NAP was 0.98 for Desires, Beliefs, Knowledge data from Baseline phase to Voice-overs treatment phase.

In response to the VAD training, Camilla performed at above-baseline levels [level changed from Baseline ($M = 0.88$, SD $= 0.64$) to VAD phase ($M = 2.67$, SD $= 0.52$)], showing an upward trend and moderate variability (**Figure 1**). In most of the VAD phase sessions, Camilla performed at ceiling levels on the Desires, Beliefs, Knowledge measure. Only minor changes in Camilla's VAD data were observed in comparison to the Voice-overs phase. There was almost no increase in level (from Voice-overs phase $M = 2.6$, SD $= 0.54$ to VAD phase $M = 2.67$, SD $= 0.52$), no immediacy of effect, a change in the trend from downward to upward, and less data variability. NAP from Baseline to VAD phase was 0.98.

Emily

During eight Baseline sessions, Emily's Desires, Beliefs, Knowledge data (**Figure 1**) were consistently mid-level ($M = 1.88$, SD $= 0.35$), demonstrating low variability and a slight upward trend. Specifically, Emily consistently responded correctly to two questions on Diverse Desires and Diverse Beliefs, but not the Knowledge Access task. Upon introduction of Voice-overs phase, Emily's Desires, Beliefs, Knowledge data (**Figure 1**) demonstrated increase in level: from Baseline ($M = 1.87$, SD $= 0.35$) to Voice-overs Training ($M = 2.60$, SD $= 0.55$), no immediacy of effect, steep accelerating trend, and moderate variability of data. NAP of 0.83 was calculated on Desires, Beliefs, Knowledge measure from Baseline to Voice-overs Training.

During the VAD phase, Emily's Desires, Beliefs, Knowledge data (**Figure 1**) demonstrated a rise in level from Baseline ($M = 1.87$, SD $= 0.35$) to VAD ($M = 2.87$, SD $= 0.35$), no immediacy of effect, a slightly downward trend driven by an outlier, and low variability. Emily almost always responded correctly to all three questions, except for one session. There was an increase in level from the Voice-over phase ($M = 2.60$, SD $= 0.54$) to the VAD phase ($M = 2.87$, SD $= 0.35$), less variability of data in the VAD phase, and change in the trend from steep upward to slightly downward. NAP of 0.95 was calculated on Desires, Beliefs, Knowledge measure from Baseline to VAD.

False Belief

False belief was assessed by two false belief tasks per session from a False Belief measure; each scored as either correct (1) or not (0). None of the six participants demonstrated improvement in false belief understanding in the Voice-overs phase. Visual analysis of False Belief data found no evidence of the functional relation between voice-overs training and children's false belief skills, and the mean NAP across participants was 0.63 for the Voice-overs phase. Only two participants showed improvement in the VAD phase and the mean NAP across participants was 0.59 for the VAD phase. Since fewer than three demonstrations of an effect were found by the visual analysis, per Kratochwill et al. (2021), we concluded no evidence of the functional relation between VAD treatment and early ToM skills development. Individual False Belief results for both treatment phases are presented in **Figure 2** and described below.

Mia

Mia scored 0 on all False Beliefs tasks in Baseline (**Figure 2**). Upon introduction of the Voice-overs treatment, Mia's data demonstrated some increase in level from $M = 0$, SD $= 0$ to $M = 0.40$, SD $= 0.55$, emergence of steep upward trend, moderate variability of data, and no immediacy of effect. NAP of 0.70 was calculated for Mia's False Belief data from Baseline to Voice-overs Phase.

Visual analysis did not indicate a considerable change in Mia's False Belief performance in the VAD phase (**Figure 2**) from the Baseline performance. There was a small increase in level as compared to Baseline phase ($M = 0$, SD $= 0$) to VAD ($M = 0.33$ SD $= 0.50$); there was no immediacy of effect, no obvious trend emerged in the VAD phase, and data showed moderate variability. In comparison to the Voice-overs phase, there was a small drop in level (from Voice-overs phase $M = 0.40$, SD $= 0.55$ to VAD phase $M = 0.33$ SD $= 0.50$), a change in the trend from upward to flat. Mia's NAP for False Belief was 0.67.

Isabella

Isabella's False Belief data (**Figure 2**) was at a low level with a Mean of 0.33 (SD $= 0.52$), showed no distinct trend, and had moderate variability. During the Voice-overs phase, Isabella's False Belief performance data remained at the low level ($M = 0.40$, SD $= 0.55$), showed no immediacy of effect, had no pronounced trend, and showed moderate variability of data. An NAP of 0.53 was calculated for Isabella's False Belief data from Baseline to Voice-overs treatment phase.

In the VAD phase, Isabella's False Belief data (**Figure 2**) showed a rise in level from Baseline ($M = 0.33$, SD $= 0.52$) to VAD ($M = 1.22$, SD $= 0.83$), no immediacy of effect, steep accelerating trend, and high variability. In a similar fashion, Isabella's False Belief data showed a rise in level from the Voice-overs phase ($M = 0.4$, SD $= 0.54$) to VAD ($M = 1.22$, SD $= 0.83$) and an emergence of upward trend. A False Belief NAP of 0.80 from Baseline to VAD phase was calculated for Isabella's False Belief data.

Paula

During the Baseline, Paula's False Belief data (**Figure 2**) was at a low level ($M = 0.29$, SD $= 0.49$) across seven sessions, and had no distinct trend, as most of Paula's False Belief scores were 0 with two spikes, when she correctly answered to one of the two False Belief tasks. In the Voice-overs treatment phase, Paula's False Belief data remained at low levels ($M = 0.20$, SD $= 0.45$), showed no immediacy of effect, exhibited a downward trend,

and had low variability. A NAP of 0.46 from Baseline to Voice-overs treatment phase was calculated for Paula's False Belief data.

Upon introduction of VAD training, Paula's False Belief data (**Figure 2**) showed no considerable change in comparison to Baseline or Voice-overs treatment phases. The data remained at low levels ($M = 0.17$, $SD = 0.41$), had no distinct trend, and variability stayed low. A NAP of 0.44 from Baseline to VAD phase was calculated for Paula's False Belief data.

Sienna

Sienna's False Belief Baseline data (**Figure 2**) showed a low level ($M = 0.14$, $SD = 0.38$), a slightly upward trend, and low variability of data across the seven sessions. Like Baseline condition, Sienna's False Belief data in the Voice-overs condition was at a low level ($M = 0.20$, $SD = 0.45$), showed no immediacy of effect, demonstrated a slightly downward trend, and exhibited low variability. An NAP of 0.53 from Baseline to Voice-overs treatment phase was calculated.

Upon the introduction of VAD training, Sienna's False Belief data (**Figure 2**) showed an increase in level when comparing Baseline ($M = 0.14$, $SD = 0.37$) to VAD ($M = 0.62$, $SD = 0.74$), no immediacy of effect, steep accelerating trend, and moderate variability of data. In comparison to the Voice-overs phase, Sienna's False Belief data also increased in level from Baseline ($M = 0.2$, $SD = 0.44$) to VAD ($M = 0.62$, $SD = 0.74$), and trend direction changed from downward to upward. An NAP of 0.66 from Baseline to VAD phase was calculated for Sienna's False Belief data.

Camilla

Camilla scored 0 on all tasks in eight sessions of the Baseline phase (**Figure 2**). During the Voice-overs phase Camilla's data slightly increased in level from Baseline ($M = 0$, $SD = 0$) to Voice-overs Training ($M = 0.20$, $SD = 0.45$), showed no immediacy of effect, had an upward trend due to one correct answer in the last session, and demonstrated low variability. NAP of 0.60 was calculated for Camilla's False Belief data from Baseline to Voice-overs Training.

Camilla's False Belief data (**Figure 2**) showed only minor changes during the VAD phase; observed were a slight increase in level from Baseline ($M = 0$, $SD = 0$) to VAD ($M = 0.17$, $SD = 0.41$), no immediacy of effect, and a downward trend that was due to one correct answer in the first session of the phase and incorrect answers in all other sessions. NAP of 0.58 from Baseline to VAD phase was calculated for Camilla's performance on False Belief.

Emily

Emily scored 0 on all False Beliefs tasks in all eight sessions of Baseline (**Figure 2**). During the Voice-overs phase Emily showed a slight increase in level from Baseline ($M = 0$, $SD = 0$) to Voice-overs Training ($M = 0.40$, $SD = 0.55$), no immediacy of effect, no distinct trend, and moderate variability of data (**Figure 2**). NAP of 0.70 was calculated for Emily from Baseline to Voice-overs Training.

Once VAD was introduced, Emily's False Belief data (**Figure 2**) showed a slight increase in level from Baseline ($M = 0$, $SD = 0$)

to VAD ($M = 0.25$, $SD = 0.46$), no immediacy of effect, and no distinct trend. In comparison to Voice-overs phase, Emily's data showed some drop in level; from Voice-overs Training ($M = 0.40$, $SD = 0.55$) to VAD ($M = 0.25$, $SD = 0.46$) the trend changed from upward to downward. NAP of 0.63 from Baseline to VAD phase was calculated for Emily's False Belief data.

DISCUSSION

The purpose of this study was to investigate whether apps for preschoolers, enhanced with ToM-promoting language, could help accelerate the development of children's ToM skills when played on their own or when paired with a with a follow-up adult-led conversation. Visual analysis showed that apps with voice-over-enhancements promoted the development of children's earlier-emerging ToM skills when the voice-over play was also followed by a discussion (VAD condition). Voice-overs without discussion were not effective in accelerating the earlier-emerging ToM skills. The study was not effective in promoting children's false belief understanding.

All six children improved on the three earlier-developing skills (diverse desires, diverse beliefs, and knowledge access (Desires, Beliefs, Knowledge) in VAD conditions indicating a positive conceptual change in social-cognitive understanding. During the baseline phase, no participant showed a conceptual understanding of knowledge access. From this, we can conclude that participants' improvements appear to be due to the conceptual insights gained from our training rather than participant maturation for two reasons. First, a multiple-baseline design allowed for the control of maturation with children starting the intervention at different points in a staggered fashion, and no improvements in ToM understanding were observed prior to the training implementation (during baseline phase) for any child. Second, in the natural course of development, typically developing (TD) children tend to master knowledge access tasks at 53.4 months of age (Wellman et al., 2011), taking on average 3–6 months to progress from understanding diverse beliefs to understanding knowledge access tasks (Rhodes and Wellman, 2013). By comparison, at the end of the study, all children were younger than 53.4 months ($M = 48.25$, ranging from 47 to 52.5 months of age), and all had advanced to knowledge access mastery in just 2–3 weeks. Thus, it is reasonable to assume that the VAD condition helped children improve earlier-developing ToM skills.

No children showed improvement in false belief higher-order ToM tasks during any phase. These findings are not surprising: ToM development is a sequential progression of conceptual achievements and requires children to master less sophisticated concepts first to achieve more complex social cognitive understanding later (Wellman and Liu, 2004; Rhodes and Wellman, 2013). Our findings align with previous studies that found pre-test performance on knowledge access tasks to correlate with children's improvement on false belief training (Benson et al., 2013; Rhodes and Wellman, 2013). Gola (2012), whose ToM video training was effective at enhancing false belief

tasks, but not tasks related to earlier-emerging diverse desires, beliefs, and knowledge (Desires, Beliefs, Knowledge), reasoned that because the participants performed near mastery levels on the Desires, Beliefs, Knowledge baseline assessment; they already had the necessary foundation and built upon it to achieve false belief understanding. In contrast, our participants scored low during the baseline testing and thus showed improvement on diverse desires, beliefs, and knowledge access that precede false belief skills.

Although all children benefited from the combination of Voice-overs in the games and follow-up discussions for Desires, Beliefs, Knowledge, only two showed improvements under Voice-overs only condition. Several explanations for these findings are possible. Children may have individual needs in terms of language development required for ToM progress: some may have required minimal support for ToM skills (Desires, Beliefs, Knowledge only) development and thus demonstrated improved performance after being exposed to mental-state vocabulary in games. In contrast, others may need more support to achieve the same results and thus benefitted from game-based discussions. Possibly, children who improved in the VAD condition built their understanding over time, benefiting from both being exposed to voice-overs and then to VAD. Finally, VAD effectiveness in comparison findings underscores the benefit of a conversational partner in cultivating these skills, and the high need for interactivity.

Interactive language-based traditional apps in our study were not successful in promoting ToM skills without a follow-up discussion. It could be that the interactivity in the games was insufficient for the ToM development. More research is needed on the types and levels of interactivity that could act as "digital adult" in supporting ToM development. It is also possible that contingent interactivity may be sufficient to promote other social-emotional skills, such as emotion recognition and social skill literacy (Craig et al., 2016; Peebles et al., 2018), emotion regulation (Craig et al., 2016; Rasmussen et al., 2019), prosocial behaviors (Shoshani et al., 2022) and social self-efficacy (Craig et al., 2016), but not ToM, given the importance of active use of the mental state language for ToM development found in some studies (Grazzani and Ornaghi, 2011; Ornaghi et al., 2011; Guajardo et al., 2013).

Extant research on preschool learning from digital devices frequently finds a greater benefit when adults support digital use than use alone by children (Reich et al., 2016; Neumann, 2020). For instance, studies of eBook reading find greater learning from these devices when facilitated by an adult (e.g., Neumann and Neumann, 2014). Further, joint media engagement, involving adults and young children, tends to increase learning (Dore and Zimmermann, 2020). Similarly, we found that voice-over app play was effective in supporting early ToM development when it was coupled with adult conversation.

Designing for Parent-Child Co-engagement

The results of this study add to the mounting evidence of the benefits of joint media engagement with digital games

for children's learning, specifically, the conversations that happen during and after the gameplay (e.g., Sobel et al., 2019; Eisen and Lillard, 2020; Musick et al., 2021). Given that conversations during gameplay are not always possible, some researchers propose that rather than expecting parents to join in digital play, it may be more practical to design games that would encourage parents to initiate post-play discussion and foster discussion of the experience during game play (Farber, 2021; Musick et al., 2021). Among the ideas to promote and improve the quality of the conversations are including conversational prompts about the games as often done in children's TV shows, providing conversation-starter guides, and designing games to support parents to act as the cheerleaders and spectators (Musick et al., 2021). Further, in mystery-solving games or treasure hunts, children can find clues by figuring out and explaining characters' false beliefs to their adult partners. Apps with "social" settings or adventure games like LEGO® DUPLO® used in this study could allow players to record and modify voice-overs to narrate the story or role-play the characters using mental state verbs and ToM-enhancing sentence structures.

A growing number of designers and researchers are building technology to enable parental support of children's learning. For example, the work by Stuckelman et al. (2021) demonstrates how an interactive app can model and encourage parents' dialogic reading and discussions. A team of Harvard researchers, in collaboration with a public media producer and educational media developer, have created a series of early literacy apps to encourage child-parent conversations and interactions and, as a result, promote children's vocabulary development and literacy skills (Rowe et al., 2021; Harvard Graduate School of Education, 2022). Lastly, newly emerging platforms, such as Amazon Glow, are being built specifically with co-engagement in mind (Amazon, 2021).

Multiple-device games could be designed for adult-child strategy building that requires mental state verbs and post-game online celebration of the wins to reminisce about the experience. Further, ToM-enhancing language embedded in game content could allow the parents to draw on and learn from specific language. This could also include discussion prompts to help children transfer and further improve ToM skills beyond the gameplay context. Finally, in addition to promoting child-parent media co-engagement, future studies should continue focusing on digital "knowledgeable others" to combat SES-related disparities in child language skills and ToM understanding. This might take the form of AI-conversation partners, such as in the study by Xu et al. (2021) on "dialogic reading" but be programmed to promote the use of mentalistic language present a promising avenue for future research. As the development and accessibility of AI, interactive digital platforms, intelligent agents, and multiplayer devices grow, so do the opportunities to use them in helping to shape children's social-linguistic environments and influencing ToM skills.

Limitations

Although the functional relation between children's understanding that people have different desires, beliefs, and knowledge sources and the use of games enriched with ToM-promoting language is followed by an adult-led conversation about the games was established, these results cannot be generalized to larger populations due to the nature of single-case research (Ledford and Gast, 2018). Additionally, the current study's design did not allow for detecting and quantifying unique contributions of different linguistic elements in the voice-overs, and individual contributions of different interactive components. Finally, whereas single-case design does not require a control group because of a baseline condition for each participant, the study could have benefited from participants who did not undergo any training or did the Voice-over or VAD phase only. Doing so would further demonstrate the absence of the maturation effect in children's performance. More research is warranted, as some of these questions would best be examined in the context of a group study.

Additional limitations concern the sample of the study. First, our sample consisted only of girls, which could be a limitation as there is evidence that girls develop Theory of Mind skills (Blijd-Hoogewys and van Geert, 2017) and some language skills (Bornstein et al., 2004) earlier than boys do. Second, all the participants in the study were exposed to a language other than English at home. Previous research suggests a positive effect of bilingualism on the rate of ToM development (see Schroeder, 2018 for a review). This study did not control for participants' mastery of a second language, and we can't say whether it had contributed to the outcome.

Further, time constraints did not allow for the implementation of the maintenance phase, which is limiting because the maintenance of participant's knowledge gains remains unknown. Moreover, even though each phase of the study met the single-subject evidence standards (Kratochwill et al., 2021), a more extended data collection period would have allowed for more sessions that could potentially allow children to improve ToM skills through repeated exposure to voice-over enriched games or VAD. A longer data collection period could have also allowed children to have more control of the procedures, such as choosing which games to play or how many to play per session. Letting children control some procedures would more closely resemble real-life gameplay.

Lastly, this study did not directly compare Voice-over and VAD methods with other strategies for facilitating ToM development, which limits our ability to conclude whether there are advantages of using games to promote ToM skills. Instead, it can be established that VAD could be an effective option available for parents and teachers to promote mental state understanding in children or content designers developing educational games to teach ToM skills. The ability to generalize the effectiveness of these results requires further investigation.

CONCLUSION

This study is the first to our knowledge to explore how educational digital apps can support children's ToM development—skills that underlie, among others, perspective taking, prosocial behaviors, and academic achievement. Our findings indicate that ToM-promoting language that is effective in face-to-face settings can be successfully implemented in digital games, especially if an adult-led conversation follows. Gameplay coupled with an adult-led conversation resulted in ToM learning, unlike gameplay alone. Although all typically developing children master ToM skills with time, there are advantages to achieving conceptual understanding sooner than later. Our findings suggest that embedding ToM language within a digital game is associated with quicker development of early ToM skills. Such results are promising, as digital games, including well-designed games are popular and may meaningfully improve children's Theory of Mind skills. To help translate the study results into practice, we have provided suggestions on how to leverage mobile apps for preschoolers to create socio-linguistic environments that promote ToM development.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Institutional Review Board (IRB) at George Mason University (GMU). The IRB at GMU reviewed all the methods and procedures to ensure the rights and welfare of the study participants. Permission was also obtained from the principal of the preschool. Written informed consent to participate in this study was provided by the participants' legal guardians/next of kin.

AUTHOR CONTRIBUTIONS

MN, SR, AE, and KC contributed to the conception and design of the study. AE suggested the methodology for the study and consulted on data analysis. MN, AE, and KC developed data collection procedures. MN collected the data. KC helped with game selection and prototype development. MB reviewed and consulted on the language for voice-over and solicited teachers' feedback. MN and SR wrote the first draft of the manuscript. All authors reviewed the manuscript, read, and approved the submitted version.

REFERENCES

AAP Council on Communications and Media (2016). Media and young minds. *Pediatrics* 138:e20162591.

Amazon (2021). *Introducing Amazon Glow—A Whole New Way to Bring Families Together for Fun and Learning [Press Release]*. Available online at: https://press.aboutamazon.com/news-releases/news-release-details/introducing-amazon-glow-whole-new-way-bring-families-together [Accessed September 28, 2021].

Amsterlaw, J., and Wellman, H. M. (2006). Theories of mind in transition: a microgenetic study of the development of false belief understanding. *J. Cogn. Dev.* 7, 139–172. doi: 10.1207/s15327647jcd0702_1

Apple (2021). *Apps for Education*. Available online at: https://www.apple.com/education/products/#learning-with-ipad [Accessed January 7, 2021].

Arnold, D. H., Chary, M., Gair, S. L., Helm, A. F., Herman, R., Kang, S., et al. (2021). A randomized controlled trial of an educational app to improve preschoolers' emergent literacy skills. *J. Child. Media* 15, 457–475.

Astington, J. W. (2003). "Sometimes necessary, never sufficient: false-belief understanding and social competence," in *Individual Differences In Theory Of Mind: Implications For Typical And Atypical Development*, eds B. Repacholi and V. Slaughter (New York, NY: Psychology Press), 13–38.

Astington, J. W., and Baird, J. (2005). "Introduction: why language matters," in *Why Language Matters For Theory Of Mind*, eds J. W. Astington and J. Baird (New York, NY: Oxford University Press), 3–25.

Benson, J. E., Sabbagh, M. A., Carlson, S. M., and Zelazo, P. D. (2013). Individual differences in executive functioning predict preschoolers' improvement from theory-of-mind training. *Dev. Psychol.* 49, 1615–1627. doi: 10.1037/a0031056

Berrett, A. N., and Carter, N. J. (2018). Imagine math facts improves multiplication fact fluency in third-grade students. *J. Behav. Educ.* 27, 223–239.

Bers, M. U. (2020). *Coding as a Playground: Programming and Computational Thinking in the Early Childhood Classroom*. New York, NY: Routledge, doi: 10.4324/9781003022602

Bindman, S. W., Sobel, K., Behbakht, S. M., Divanji, R. A., Shahn, E., and Scanlon, M. (2021). ""Let's draw chocolate milk together": supports and challenges for families' joint engagement with artistic tablet apps," in *Proceedings of the Interaction Design and Children (IDC'21)*, Athens, 183–195.

Blackwell, C. K., Lauricella, A. R., Wartella, E., Robb, M., and Schomburg, R. (2013). Adoption and use of technology in early education: the interplay of extrinsic barriers and teacher attitudes. *Comput. Educ.* 69, 310–319. doi: 10.1016/j.compedu.2013.07.024

Blijd-Hoogewys, E., and van Geert, P. L. (2017). Non-linearities in theory-of-mind development. *Front. Psychol.* 7:1970. doi: 10.3389/fpsyg.2016.01970

Bornstein, M. H., Hahn, C. S., and Haynes, O. M. (2004). Specific and general language performance across early childhood: Stability and gender considerations. *First Lang.* 24, 267–304.

Bosacki, S., Moreira, F. P., Sitnik, V., Andrews, K., and Talwar, V. (2020). Theory of mind, self-knowledge, and perceptions of loneliness in emerging adolescents. *J. Genet. Psychol.* 181, 14–31. doi: 10.1080/00221325.2019.1687418

Callaghan, M. N., and Reich, S. M. (2018). Are educational preschool apps designed to teach? An analysis of the app market. *Learn. Media Technol.* 43, 280–293.

Callaghan, M. N., and Reich, S. M. (2022). "Applying a developmental lens to educational game designs for preschoolers," in *Research Anthology on Developments in Gamification and Game-Based Learning* (Hershey, PA: IGI Global), 471–487.

Caputi, M., Lecce, S., Pagnin, A., and Banerjee, R. (2012). Longitudinal effects of theory of mind on later peer relations: the role of prosocial behavior. *Dev. Psychol.* 48, 257–270. doi: 10.1037/a0025402

Charness, G., List, J. A., Rustichini, A., Samek, A., and Van De Ven, J. (2019). Theory of mind among disadvantaged children: evidence from a field experiment. *J. Econ. Behav. Org.* 166, 174–194.

Cingel, D. P., Sumter, S. R., Stoeten, E., and Mann, S. (2020). Can television help to decrease stigmatization among young children? The role of theory of mind and general and explicit inserts. *Media Psychol.* 23, 342–364.

Clements, W. A., Rustin, C. L., and McCallum, S. (2000). Promoting the transition from implicit to explicit understanding: a training study of false belief. *Dev. Sci.* 3, 81–92. doi: 10.1111/1467-7687.00102

Craig, A. B., Brown, E. R., Upright, J., and DeRosier, M. E. (2016). Enhancing children's social emotional functioning through virtual game-based delivery of social skills training. *J. Child Fam. Stud.* 25, 959–968.

Cutting, A. L., and Dunn, J. (1999). Theory of mind, emotion understanding, language, and family background: Individual differences and interrelations. *Child Dev.* 70, 853–865. doi: 10.1111/1467-8624.00061

De Villiers, J. G., and De Villiers, P. A. (2000). "Linguistic determinism and the understanding of false," in *Children's Reasoning And The Mind*, eds P. Mitchell and J. Riggs (Hove: Psychology Press), 191–228. doi: 10.1111/1460-6984.12317

De Villiers, J. G., and De Villiers, P. A. (2014). The role of language in theory of mind development. *Topics Lang. Disord.* 34, 313–328.

Dessen, M. A., and de Hollanda Souza, D. (2014). Theory-of-mind development in Brazilian low-income children. *Int. Educ. Res.* 2, 41–52. doi: 10.12735/ier.v2i3p41

Devine, R. T., and Hughes, C. (2018). Family correlates of false belief understanding in early childhood: a meta-analysis. *Child Dev.* 89, 971–987. doi: 10.1111/cdev.12682

Devine, R. T., White, N., Ensor, R., and Hughes, C. (2016). Theory of mind in middle childhood: longitudinal associations with executive function and social competence. *Dev. Psychol.* 52, 758–771. doi: 10.1037/dev0000105

Dore, R. A., Amendum, S. J., Golinkoff, R. M., and Hirsh-Pasek, K. (2018). Theory of mind: a hidden factor in reading comprehension? *Educ. Psychol. Rev.* 30, 1067–1089.

Dore, R. A., and Zimmermann, L. (2020). "Coviewing, scaffolding, and children's media comprehension," in *The International Encyclopedia Of Media Psychology*, eds J. van den Bulck and D. R. Roskos-Ewoldsen (Hoboken, NJ: Wiley), 1–8.

Dore, R. A., Shirilla, M., Hopkins, E., Collins, M., Scott, M., Schatz, J., et al. (2019). Education in the app store: using a mobile game to support U.S. preschoolers' vocabulary learning. *J. Child. Media* 13, 452–471.

Ebert, S., Lehrl, S., and Weinert, S. (2020). Differential effects of the home language and literacy environment on child language and theory of mind and their relation to socioeconomic background. *Front. Psychol.* 11:2321. doi: 10.3389/fpsyg.2020.555654

Eisen, S., and Lillard, A. S. (2020). Learning from apps and objects: the human touch. *Mind Brain Educ.* 14, 16–23. doi: 10.3389/fpsyg.2021.621372

Engle, P. L., Fernald, L. C., Alderman, H., Behrman, J., O'Gara, C., Yousafzai, A., et al. (2011). Strategies for reducing inequalities and improving developmental outcomes for young children in low-income and middle-income countries. *Lancet* 378, 1339–1353. doi: 10.1016/S0140-6736(11)60889-1

Esteban, M., Sidera, F., Serrano, J., Amadó, A., and Rostan, C. (2010). Improving social understanding of preschool children: evaluation of a training program. *Electron. J. Res. Educ. Psychol.* 8, 841–860.

Ewin, C. A., Reupert, A. E., McLean, L. A., and Ewin, C. J. (2021). The impact of joint media engagement on parent–child interactions: a systematic review. *Hum. Behav. Emerg. Technol.* 3, 230–254.

Farber, M. (2021). *Gaming SEL: Games As Transformational to Social and Emotional Learning*. New York, NY: Peter Lang Publishing Group.

Fink, E., Begeer, S., Hunt, C., and de Rosnay, M. (2014). False-belief understanding and social preference over the first 2 years of school: a longitudinal study. *Child Dev.* 85, 2389–2403. doi: 10.1111/cdev.12302

Fink, E., Begeer, S., Peterson, C. C., Slaughter, V., and Rosnay, M. (2015). Friendlessness and theory of mind: a prospective longitudinal study. *Br. J. Dev. Psychol.* 33, 1–17. doi: 10.1111/bjdp.12060

Florit, E., De Carli, P., Giunti, G., and Mason, L. (2020). Advanced theory of mind uniquely contributes to children's multiple-text comprehension. *J. Exp. Child Psychol.* 189:104708. doi: 10.1016/j.jecp.2019.104708

Flynn, R. M., Richert, R. A., and Wartella, E. (2019). Play in a digital world: how interactive digital games shape the lives of children. *Am. J. Play* 12, 54–73.

Gola, A. A. H. (2012). Mental verb input for promoting children's theory of mind: a training study. *Cogn. Dev.* 27, 64–76.

Grazzani, I., and Ornaghi, V. (2011). Emotional state talk and emotion understanding: a training study with preschool children. *J. Child Lang.* 38, 1124–1139.

Griffith, S. F., and Arnold, D. H. (2019). Home learning in the new mobile age: parent–child interactions during joint play with educational apps in the US. *J. Child. Media* 13, 1–19.

Griffith, S. F., Hagan, M. B., Heymann, P., Heflin, B. H., and Bagner, D. M. (2020). Apps as learning tools: a systematic review. *Pediatrics* 145:e20191579.

Griffith, S. F., Hanson, K. G., Rolon-Arroyo, B., and Arnold, D. H. (2019). Promoting early achievement in low-income preschoolers in the United States with educational apps. *J. Child. Media* 13, 328–344.

Guajardo, N. R., and Watson, A. C. (2002). Narrative discourse and theory of mind development. *J. Genet. Psychol.* 163, 305–325. doi: 10.1080/00221320209598686

Guajardo, N. R., Petersen, R., and Marshall, T. R. (2013). The roles of explanation and feedback in false belief understanding: a microgenetic analysis. *Journal of Genet. Psychol.* 174, 225–252. doi: 10.1080/00221325.2012.682101

Guernsey, L. (2007). *Into The Minds Of Babes: How Screen Time Affects Children From Birth To Age Five.* New York, NY: Basic Books.

Hale, C., and Tager-Flusberg, H. (2003). The influence of language on theory of mind: a training study. *Dev. Sci.* 6, 346–359. doi: 10.1111/1467-7687.00289

Harvard Graduate School of Education (2022). *Cultivating Early Literacy.* Cambridge, MA: Harvard Graduate School of Education.

Hasni, A. A., Adamson, L. B., Williamson, R. A., and Robins, D. L. (2017). Adding sound to theory of mind: Comparing children's development of mental-state understanding in the auditory and visual realms. *J. Exp. Child Psychol.* 164, 239–249. doi: 10.1016/j.jecp.2017.07.009

Herodotou, C. (2018). Young children and tablets: a systematic review of effects on learning and development. *J. Comput. Assist. Learn.* 34, 1–9.

Hirsh-Pasek, K., Zosh, J. M., Golinkoff, R. M., Gray, J. H., Robb, M. B., and Kaufman, J. (2015). Putting education in "educational" apps lessons from the science of learning. *Psychol. Sci. Public Interest* 16, 3–34. doi: 10.1177/1529100615569721

Hodgkinson, S., Godoy, L., Beers, L. S., and Lewin, A. (2017). Improving mental health access for low-income children and families in the primary care setting. *Pediatrics* 139:e20151175.

Hofmann, S. G., Doan, S. N., Sprung, M., Wilson, A., Ebesutani, C., Andrews, L. A., et al. (2016). Training children's theory-of-mind: a meta-analysis of controlled studies. *Cognition* 150, 200–212.

Holmes, H. A., Black, C., and Miller, S. A. (1996). A cross-task comparison of false belief understanding in a Head Start population. *J. Exp. Child Psychol.* 63, 263–285. doi: 10.1006/jecp.1996.0050

Horner, R. H., Carr, E. G., Halle, J., McGee, G., Odom, S., and Wolery, M. (2005). The use of single-subject research to identify evidence-based practice in special education. *Except. Child.* 71, 165–179.

Huber, B., Yeates, M., Meyer, D., Fleckhammer, L., and Kaufman, J. (2018). The effects of screen media content on young children's executive functioning. *J. Exp. Child Psychol.* 170, 72–85. doi: 10.1016/j.jecp.2018.01.006

Hughes, C., and Devine, R. T. (2015). Individual differences in theory of mind from preschool to adolescence: achievements and directions. *Child Dev. Perspect.* 9, 149–153.

Imuta, K., Henry, J. D., Slaughter, V., Selcuk, B., and Ruffman, T. (2016). Theory of mind and prosocial behavior in childhood: a meta-analytic review. *Dev. Psychol.* 52:1192. doi: 10.1037/dev0000140

Keenan, T. (2003). "The preschool years and beyond," in *Individual Differences In Theory Of Mind: Implications For Typical And Atypical Development,* eds B. Repacholi and V. Slaughter (London: Psychology Press), 122–143.

Kennedy, C. H. (2005). *Single-Case Designs For Education Research.* Boston, MA: Allyn and Bacon.

Kim, J., Gilbert, J., Yu, Q., and Gale, C. (2021). Measures matter: A meta-analysis of the effects of educational apps on preschool to grade 3 children's literacy and math skills. *AERA Open* 7:23328584211004183.

Kirsch, C. (2021). "9 Promoting multilingualism and multiliteracies through storytelling: a case study on the use of the app iteo in preschools in luxembourg," in *Multilingual Literacy,* eds E. O. Breuer, E. Lindgren, A. Stavans, and E. Van Steendam (Bristol: Multilingual Matters), 187–210.

Kratochwill, T. R., Hitchcock, J. H., Horner, R. H., Levin, J. R., Odom, S. L., Rindskopf, D. M., et al. (2013). Single-case intervention research design standards. *Remedial Spec. Educ.* 34, 26–38.

Kratochwill, T. R., Horner, R. H., Levin, J. R., Machalicek, W., Ferron, J., and Johnson, A. (2021). Single-case design standards: an update and proposed upgrades. *J. Sch. Psychol.* 89, 91–105. doi: 10.1016/j.jsp.2021.10.006

Lecce, S., Caputi, M., and Pagnin, A. (2014). Long-term effect of theory of mind on school achievement: the role of sensitivity to criticism. *Eur. J. Dev. Psychol.* 11, 305–318. doi: 10.1080/17405629.2013.821944

Lecce, S., Ronchi, L., and Devine, R. T. (2021). Mind what teacher says: teachers' propensity for mental-state language and children's theory of mind in middle childhood. *Soc. Dev.* 31, 303–318.

Ledford, J. R., and Gast, D. L. (2018). *Single Case Research Methodology.* New York, NY: Routledge.

Lillard, A. S., and Kavanaugh, R. D. (2014). The contribution of symbolic skills to the development of an explicit theory of mind. *Child Dev.* 85, 1535–1551. doi: 10.1111/cdev.12227

Lohmann, H., and Tomasello, M. (2003). The role of language in the development of false belief understanding: a training study. *Child Dev.* 74, 1130–1144. doi: 10.1111/1467-8624.00597

Mann, H. B., and Whitney, D. R. (1947). On a test of whether one of two random variables is stochastically larger than the other. *Ann. Math. Stat.* 18, 50–60.

Manolov, R., Losada, J. L., Chacón-Moscoso, S., and Sanduvete-Chaves, S. (2016). Analyzing two-phase single-case data with non-overlap and mean difference indices: illustration, software tools, and alternatives. *Front. Psychol.* 7:32. doi: 10.3389/fpsyg.2016.00032

Manolov, R., Solanas, A., Sierra, V., and Evans, J. J. (2011). Choosing among techniques for quantifying single-case intervention effectiveness. *Behav. Therapy* 42, 533–545. doi: 10.1016/j.beth.2010.12.003

Meyer, M., Zosh, J. M., McLaren, C., Robb, M., McCafferty, H., Golinkoff, R. M., et al. (2021). How educational are "educational" apps for young children? App store content analysis using the four pillars of learning framework. *J. Child. Media* 15, 526–548. doi: 10.1080/17482798.2021.1882516

Michiels, B., Heyvaert, M., and Onghena, P. (2018). The conditional power of randomization tests for single-case effect sizes in designs with randomized treatment order: a monte carlo simulation study. *Behav. Res. Methods* 50, 557–575. doi: 10.3758/s13428-017-0885-7

Mori, A., and Cigala, A. (2015). Perspective taking: training procedures in developmentally typical preschoolers: different intervention methods and their effectiveness. *Educ. Psychol. Rev.* 28, 267–294. doi: 10.1007/s10648-015-9306-6

Musick, G., Freeman, G., and McNeese, N. J. (2021). Gaming as family time: digital game co-play in modern parent-child relationships. *Proc. ACM Hum. Comput. Interact.* 5, 1–25.

Neumann, M. M. (2018). Using tablets and apps to enhance emergent literacy skills in young children. *Early Child. Res. Q.* 42, 239–246.

Neumann, M. M. (2020). Teacher scaffolding of preschoolers' shared reading with a storybook app and a printed book. *J. Res. Child. Educ.* 34, 367–384.

Neumann, M. M., and Neumann, D. L. (2014). Touch screen tablets and emergent literacy. *Early Child. Educ.* 42, 231–239. doi: 10.1007/s10643-013-0608-3

Nikolayev, M., Clark, K., and Reich, S. M. (2015). "Social-emotional learning opportunities in online games for preschoolers," in *Emotions, Technology And Games,* eds S. Tettegah and W. Huang (New York, NY: Academic Press), 211–227.

Nikolayev, M., Reich, S. M., Muskat, T., Tadjbakhsh, N., and Callaghan, M. N. (2021). Review of feedback in edutainment games for preschoolers in the USA. *J. Child.Media* 15, 358–375.

Odom, S. L., and Strain, P. S. (2002). Evidence-based practice in early intervention/early childhood special education: single-subject design research. *J. Early Interv.* 25, 151–160.

Ornaghi, V., Brockmeier, J., and Gavazzi, I. G. (2011). The role of language games in children's understanding of mental states: a training study. *J. Cogn. Dev.* 12, 239–259. doi: 10.1080/15248372.2011.563487

Papadakis, S. (2021a). The impact of coding apps to support young children in computational thinking and computational fluency. A literature review. *Front. Educ.* 6:657895. doi: 10.3389/feduc.2021.657895

Papadakis, S. (2021b). Advances in Mobile Learning Educational Research (AMLER): mobile learning as an educational reform. *Adv. Mobile Learn. Educ. Res.* 1, 1–4.

Papadakis, S. (2022). "Apps to promote computational thinking concepts and coding skills in children of preschool and pre-primary school age," in *Research Anthology on Computational Thinking, Programming, and Robotics in the Classroom,* ed. Information Resources Management Association (Hershey, PA: IGI Global), 610–630.

Papadakis, S., Alexandraki, F., and Zaranis, N. (2021). Mobile device use among preschool-aged children in Greece. *Educ. Inf. Technol.* 27, 2717–2750. doi: 10.1007/s10639-021-10718-6

Parker, R. I., and Vannest, K. (2009). An improved effect size for single-case research: nonoverlap of all pairs. *Behav. Therapy* 40, 357–367.

Parker, R. I., Vannest, K. J., and Davis, J. L. (2011). Effect size in single-case research: a review of nine nonoverlap techniques. *Behav. Modif.* 35, 303–322. doi: 10.1177/0145445511399147

Peebles, A., Bonus, J. A., and Mares, M. L. (2018). Questions+ answers+ agency: interactive touchscreens and Children's learning from a socio-emotional TV story. *Comput. Hum. Behav.* 85, 339–348.

Pempek, T. A., and Lauricella, A. R. (2017). "The effects of parent-child interaction and media use on cognitive development in infants, toddlers, and preschoolers," in *Cognitive Development In Digital Contexts*, eds F. C. Blumberg and P. J. Brooks (Cambridge, MA: Academic Press), 53–74.

Peskin, J., and Astington, J. W. (2004). The effects of adding metacognitive language to story texts. *Cogn. Dev.* 19, 253–273.

Peterson, C. C., Slaughter, V., and Wellman, H. M. (2018). Nimble negotiators: How theory of mind (ToM) interconnects with persuasion skills in children with and without ToM delay. *Dev. Psychol.* 54:494. doi: 10.1037/dev0000451

Radesky, J. S., Schumacher, J., and Zuckerman, B. (2015). Mobile and interactive media use by young children: the good, the bad, and the unknown. *Pediatrics* 135, 1–3.

Rasmussen, E. E., Strouse, G. A., Colwell, M. J., Russo Johnson, C., Holiday, S., Brady, K., et al. (2019). Promoting preschoolers' emotional competence through prosocial TV and mobile app use. *Media Psychol.* 22, 1–22.

Reich, S. M., Korobkova, K. A., Black, R. W., and Sumaroka, M. (2012). "Hey! Can you show me how to do this?" Digital games mediating family interactions," in *Children's Virtual Play Worlds: Culture, Learning And Participation*, eds A. Burke and J. March (New York, NY: Peter Lang), 133–150.

Reich, S. M., Yau, J. C., and Warschauer, M. (2016). Tablet-based ebooks for young children: what does the research say? *J. Dev. Behav. Pediatr.* 37, 585–591. doi: 10.1097/DBP.0000000000000335

Reiß, M., Krüger, M., and Krist, H. (2019). Theory of mind and the video deficit effect: video presentation impairs children's encoding and understanding of false belief. *Media Psychol.* 22, 23–38.

Renouf, A., Brendgen, M., Parent, S., Vitaro, F., David Zelazo, P., Boivin, M., et al. (2010). Relations between theory of mind and indirect and physical aggression in kindergarten: evidence of the moderating role of prosocial behaviors. *Soc. Dev.* 19, 535–555.

Rhodes, M., and Wellman, H. (2013). Constructing a new theory from old ideas and new evidence. *Cogn. Sci.* 37, 592–604. doi: 10.1111/cogs.12031

Richert, R. A., Robb, M. B., and Smith, E. I. (2011). Media as social partners: the social nature of young children's learning from screen media. *Child Dev.* 82, 82–95. doi: 10.1111/j.1467-8624.2010.01542.x

Rideout, V., and Robb, M. B. (2020). *The Common Sense Census: Media Use By Kids Age Zero To Eight, 2020*. San Francisco, CA: Common Sense Media.

Roheger, M., Hranovska, K., Martin, A. K., and Meinzer, M. (2022). A systematic review and meta-analysis of social cognition training success across the healthy lifespan. *Sci. Rep.* 12, 1–21. doi: 10.1038/s41598-022-07420-z

Rowe, M. L., Turco, R. G., and Blatt, J. H. (2021). Can interactive apps promote parent-child conversations? *J. Appl. Dev. Psychol.* 76:101326.

Ruffman, T. (2014). To belief or not belief: children's theory of mind. *Dev. Rev.* 34, 265–293. doi: 10.1016/j.dr.2014.04.001

Ruffman, T., Slade, L., and Crowe, E. (2002). The relation between children's and mothers' mental state language and theory of mind understanding. *Child Dev.* 73, 734–751. doi: 10.1111/1467-8624.00435

Russo-Johnson, C., Troseth, G., Duncan, C., and Mesghina, A. (2017). All tapped out: touchscreen interactivity and young children's word learning. *Front. Psychol.* 8:578. doi: 10.3389/fpsyg.2017.00578

San Juan, V., and Astington, J. W. (2017). Does language matter for implicit theory of mind? The effects of epistemic verb training on implicit and explicit false-belief understanding. *Cogn. Dev.* 41, 19–32.

Schroeder, S. R. (2018). Do bilinguals have an advantage in theory of mind? A meta-analysis. *Front. Commun.* 3:36. doi: 10.3389/fcomm.2018.00036

Segers, E., and Kleemans, T. (2020). The impact of the digital home environment on kindergartners' language and early literacy. *Front. Psychol.* 11:538584. doi: 10.3389/fpsyg.2020.538584

Shahaeian, A., Peterson, C. C., Slaughter, V., and Wellman, H. M. (2011). Culture and the sequence of steps in theory of mind development. *Dev. Psychol.* 47, 1239–1247. doi: 10.1037/a0023899

Shakoor, S., Jaffee, S. R., Bowes, L., Ouellet-Morin, I., Andreou, P., Happé, F., et al. (2012). A prospective longitudinal study of children's theory of mind and adolescent involvement in bullying. *J. Child Psychol. Psychiatry* 53, 254–261. doi: 10.1111/j.1469-7610.2011.02488.x

Shoshani, A., Nelke, S., and Girtler, I. (2022). Tablet applications as socializing platforms: The effects of prosocial touch screen applications on young children's prosocial behavior. *Comput. Hum. Behav.* 127:107077.

Slaughter, V., and Gopnik, A. (1996). Conceptual coherence in the child's theory of mind: training children to understand belief. *Child Dev.* 67, 2967–2988. doi: 10.2307/1131762

Slaughter, V., Dennis, M. J., and Pritchard, M. (2002). Theory of mind and peer acceptance in preschool children. *Br. J. Dev. Psychol.* 20, 545–564. doi: 10.1348/026151002760390945

Slaughter, V., Imuta, K., Peterson, C. C., and Henry, J. D. (2015). Meta-analysis of theory of mind and peer popularity in the preschool and early school years. *Child Dev.* 86, 1159–1174. doi: 10.1111/cdev.12372

Slaughter, V., Peterson, C. C., and Mackintosh, E. (2007). Mind what mother says: narrative input and theory of mind in typical children and those on the autism spectrum. *Child Dev.* 78, 839–858. doi: 10.1111/j.1467-8624.2007.01036.x

Sobel, K., Yen, K., Cheng, Y., Chen, Y., and Hiniker, A. (2019). "No touch pig! Investigating child-parent use of a system for training executive function," in *Proceedings of the 18th ACM International Conference on Interaction Design and Children*, (New York, NY: ACM), 339–351.

Spruijt, A. M., Ziermans, T. B., Dekker, M. C., and Swaab, H. (2020). Educating parents to enhance children's reasoning abilities: a focus on questioning style. *J. Appl. Dev. Psychol.* 66:101102.

Stuckelman, Z. D., Strouse, G. A., and Troseth, G. L. (2021). Value added: digital modeling of dialogic questioning promotes positive parenting during shared reading. *J. Fam. Psychol.* doi: 10.1037/fam0000932 [Epub ahead of print].

Sweeney, S. (2017). Joint parent-child app play can bolster language development: slps can show parents how to harness apps to get their children talking at home. *ASHA Leader* 22. [Epub ahead of print].

Takeuchi, L., and Levine, M. H. (2014). "Learning in a digital age: towards a new ecology of human development," in *Media And The Wellbeing Of Children And Adolescents*, eds A. Jordan and D. Romer (New York, NY: Oxford University Press), 20–43.

Takeuchi, L., and Stevens, R. (2011). *The New Coviewing: Designing For Learning Through Joint Media Engagement*. New York, NY: The Joan Ganz Cooney Center at Sesame Workshop.

Tarullo, A. R., Bruce, J., and Gunnar, M. R. (2007). False belief and emotion understanding in post-institutionalized children. *Soc. Dev.* 16, 57–78. doi: 10.1111/j.1467-9507.2007.00372.x

Taumoepeau, M., and Reese, E. (2013). Maternal reminiscing, elaborative talk, and children's theory of mind: an intervention study. *First Lang.* 33, 388–410. doi: 10.1177/0142723713493347

Teepe, R. C., Molenaar, I., and Verhoeven, L. (2017). Technology-enhanced storytelling stimulating parent–child interaction and preschool children's vocabulary knowledge. *J. Comput. Assist. Learn.* 33, 123–136.

Toh, W., and Lim, F. V. (2021). Let's play together: ways of parent–child digital co-play for learning. *Interact. Learn. Environ.* 1–11. doi: 10.1080/10494820.2021.1951768

Tompkins, V. (2015). Improving low-income preschoolers' theory of mind: a training study. *Cogn. Dev.* 36, 1–19. doi: 10.1016/j.cogdev.2015.07.001

Tompkins, V., Benigno, J. P., Kiger Lee, B., and Wright, B. M. (2018). The relation between parents' mental state talk and children's social understanding: a meta-analysis. *Soc. Dev.* 27, 223–246. doi: 10.1111/cdev.12682

Tompkins, V., Farrar, M. J., and Montgomery, D. E. (2019). Speaking your mind: Language and narrative in young children's theory of mind development. *Adv. Child Dev. Behav.* 56, 109–140. doi: 10.1016/bs.acdb.2018.11.003

Vannest, K. J., Parker, R. I., and Gonen, O. (2011). *Single Case Research: Web Based Calculators for SCR Analysis. (Version 1.0) [Web-based Application]*. College Station, TX: Texas A&M University. Available online at: http://www.singlecaseresearch.org/calculators/nap

Vissers, C., and Koolen, S. (2016). Theory of mind deficits and social emotional functioning in preschoolers with specific language impairment. *Front. Psychol.* 7:1734. doi: 10.3389/fpsyg.2016.01734

Weimer, A. A., Warnell, K. R., Ettekal, I., Cartwright, K. B., Guajardo, N. R., and Liew, J. (2021). Correlates and antecedents of theory of mind development during middle childhood and adolescence: an integrated model. *Dev. Rev.* 59:100945.

Wellman, H. M. (2018). Theory of mind: the state of the art. *Eur. J. Dev. Psychol.* 15, 728–755.

Wellman, H. M., and Liu, D. (2004). Scaling of theory-of-mind tasks. *Child Dev.* 75, 523–541. doi: 10.1111/j.1467-8624.2004.00691.x

Wellman, H. M., and Peterson, C. C. (2013). "Theory of mind development, and deafness," in *Understanding Other Minds: Perspectives From Developmental Social Neuroscience*, eds S. Baron-Cohen, M. V. Lombardo, and H. Tager-Flusberg (London: Oxford University Press), 51–71.

Wellman, H. M., Cross, D., and Watson, J. (2001). Meta-analysis of theory-of-mind development: the truth about false belief. *Child Dev.* 72, 655–684. doi: 10.1111/1467-8624.00304

Wellman, H. M., Fang, F., and Peterson, C. C. (2011). Sequential progressions in a theory-of-mind scale: longitudinal perspectives. *Child Dev.* 82, 780–792. doi: 10.1111/j.1467-8624.2011.01583.x

Wimmer, H., and Perner, J. (1983). Beliefs about beliefs: representation and the containing function of wrong beliefs in young children's understanding of deception. *Cognition* 13, 103–128. doi: 10.1016/0010-0277(83)90004-5

Wood, E., Petkovski, M., De Pasquale, D., Gottardo, A., Evans, M. A., and Savage, R. S. (2016). Parent scaffolding of young children when engaged with mobile technology. *Front. Psychol.* 7:690. doi: 10.3389/fpsyg.2016.00690

Xu, Y., Aubele, J., Vigil, V., Bustamante, A. S., Kim, Y. S., and Warschauer, M. (2021). Dialogue with a conversational agent promotes children's story comprehension via enhancing engagement. *Child Dev.* 93, e149–e167. doi: 10.1111/cdev.13708

Yelland, N., and Masters, J. (2007). Rethinking scaffolding in the information age. *Comput. Educ.* 48, 362–382.

Zosh, J. M., Kathy, H.-P., Emily, J. H., Hanne, J., Claire, L., Dave, N., et al. (2018). Accessing the inaccessible: redefining play as a spectrum. *Front. Psychol.* 9:1124. doi: 10.3389/fpsyg.2018.01124