Exploring Kid Space in the wild: a preliminary study of multimodal and immersive collaborative play-based learning experiences

Sinem Aslan¹ · Ankur Agrawal¹ · Nese Alyuz¹ · Rebecca Chierichetti¹ ·
Lenitra M. Durham¹ · Ramesh Manuvinakurike¹ · Eda Okur¹ · Saurav Sahay¹ ·
Sangita Sharma¹ · John Sherry¹ · Giuseppe Raffa¹ · Lama Nachman¹

Accepted: 18 November 2021 / Published online: 8 January 2022
© Association for Educational Communications and Technology 2021

Abstract

Parents recognize the potential benefits of technology for their young children but are wary of too much screen time and its potential deficits in terms of social engagement and physical activity. To address these concerns, related literature suggests technology usages with a blend of digital and physical learning experiences. Towards this end, we developed Kid Space, incorporating immersive computing experiences designed to engage children more actively in physical movement and social collaboration during play-based learning. The technology features an animated peer learner, Oscar, who aims to understand and respond to children's actions and utterances using extensive multimodal sensing and sensemaking technologies. To investigate student engagement during Kid Space learning experiences, an exploratory case study was designed using a formative research method with eight first-grade students. Multimodal data (audio and video) along with observational, interview, and questionnaire data were collected and analyzed. The results show that the students demonstrated high levels of engagement, less attention focused on the screen (projected wall), and more physical activity. In addition to these promising results, the study also enabled us to understand actionable insights to improve Kid Space for future deployments (e.g., the need for real-time personalization). We plan to incorporate the lessons learned from this preliminary study and deploy Kid Space with real-time personalization features for longer periods with more students.

Keywords Game-based learning · Mixed reality · Collaborative learning · Intelligent tutoring systems (ITSs) · Student engagement

✉ Sinem Aslan
sinem.aslan@intel.com

¹ Research Scientist, Intel Corporation, 2111 NE, 25th Avenue, Hillsboro, OR 97124, USA

Introduction

The ongoing debate over technology-mediated learning environments in early childhood education (Blackwell et al., 2014) has primarily focused on extended screen time resulting in a lack of physical activity and traditional play with potential negative consequences on children's physical, emotional, and social development (Wood et al., 2008; Ahearne et al., 2016; American Academy of Pediatrics, 2016). Supporting these concerns, our previous ethnographic research indicated that the parents were primarily worried about the screen time of their children during technology-mediated learning and wanted to minimize the time their kids were sitting stationary in front their devices and staring at the screen (Anderson et al., 2015).

However, as an outcome of the 21st century, children grow up in technology-rich environments (American Academy of Pediatrics, 2016) and screen time has become a part of their daily activities, beginning in early childhood (Shapiro, 2015; Zabatiero et al., 2018). Therefore, it is important to investigate usages which provide developmentally appropriate technology use for younger children. Towards this end, Kid Space is designed in a way that empowers children to use their whole body and immerse themselves into the computing experiences with less screen time, but with more physical activity and interactivity through hands-on and social learning. In this preliminary study, we evaluated how students in early childhood education interacted with these experiences in the wild (i.e., in a school), using metrics for behavioral, emotional, and social engagement.

Related work

Children start to develop foundations for their college and career success as early as their early childhood years (Guilfoyle, 2013). During these years, play-based activities are critical for children's learning and development (The National Association for the Education of Young Children, 2020) as they promote engagement with cognitive, physical, social, and emotional benefits (Singer et al., 2006; Ginsburg, 2007; Arrow, 2019). Use of technology has created new opportunities for play-based learning through digital games on computers and interactive surfaces (e.g., tablets, smartphones, and multi-touch tables) (Nacher et al., 2016). However, these opportunities have drawbacks.

One major concern with educational games on computers is their potential negative impact on children's social development because they are solitary interfaces (Karno & Hatcher, 2020). Although use of interactive surfaces, specifically multi-touch tables, enables multiple children to collaborate on learning tasks (Karno & Hatcher, 2020), they still require extensive screen time and limited physical movement. To address these drawbacks, robots and smart toys have recently gained popularity (Nacher et al., 2016). However, these technologies still lack immersive, play-based learning environments which can "augment the real space but not replace the natural and real-world activities" (Nacher et al., 2016, p. 26).

Towards this end, there is some promising recent research in related literature targeting a blend of digital and physical learning experiences through visual, auditory, and tactile interactions. Leveraging speech, facial features, body gestures, and spatial location, Zhao et al. (2018) developed a Cognitive Immersive Room (CIR) for students learning a second language. The researchers implemented a user study with 16 students learning Mandarin

Chinese in the CIR environment and reported positive impact on student experience (Zhao et al., 2018). Magika is another exemplary multisensory technology supporting playful interventions for children with special learning needs (Gelsomini et al., 2019). The technology incorporates visual content projected on walls and on floors, ambient sound, smart physical objects, connected appliances, and smart lights. The system reacts to children's tangible interactions and body movements. Similarly, Rensselaer Mandarin is another recent research project enabling students to use speech and gestures to interact with an immersive, mixed reality game experience. Students interact with artificial intelligence (AI) agents through a human-scale, 360° panoramic screen to improve understanding, pronunciation, and vocabulary in Mandarin Chinese (Allen et al., 2019).

These research efforts imply two major design principles for enabling multimodal learning experiences. The first is immersion with multi-sensory interactions, where students are expected to use their whole body (e.g., speech, gesture, etc.) and immerse into novel interactions blending the physical and digital world seamlessly “through multimedia and visualization at human scale” (Zhao et al., 2018, p. 2). The second is incorporation of multimodal sensing and sensemaking technologies through AI to understand students and their context (e.g., activities, speech, behaviors, emotions, etc.) and respond to them naturally. These two principles also guided the design of Kid Space learning experiences as described in detail in the following section.

Kid Space learning experiences

Building on the related literature, our own ethnographic research at dozens of households showed that the parents shared many of the same ideas regarding the potential benefits and pitfalls of technology in the lives of their early school-age children (Anderson et al., 2015). The majority of the parents interviewed recognized the utility of technology-supported learning but were wary of too much screen time and its potential deficits in terms of social engagement and physical activity (Anderson et al., 2015). To address these pain points, we developed Kid Space (Anderson et al., 2018; Sahay et al., 2019).

Kid Space incorporates immersive computing experiences designed to engage children more actively in physical movement and social collaboration during play-based learning. Kid Space features an animated peer learner, Oscar (see Fig. 1), designed to understand and respond to children's actions and utterances during learning experiences. To enable

Fig. 1 Evolution of oscar. **A** Design sketches of Oscar. **B** Early prototype. **C** Final prototype

developmentally appropriate educational play, the technology incorporates a set of math games designed in collaboration with two teachers in early childhood education as subject matter experts. These games enable Oscar and groups of children to play and learn together.

To help Oscar seem physically present in the space, Kid Space uses extensive multimodal sensing and sensemaking technologies as well as projection. Such technologies include face, pose, and gesture recognition, location tracking, ambient audio classification, automatic speech recognition, and natural language understanding and dialogue management. These capabilities empower Oscar to respond to children's actions and speech in a more natural and personalized manner by integrating multimodal inputs for robust inference and incorporating a game engine to bring Oscar and other game elements to life via full wall projection.

To create multimodal learning experiences for Kid Space, we followed the five major steps of the ADDIE model guiding the instructional design process: Analysis, Design, Development, Implementation, and Evaluation (Molenda, 2003). During this process, we closely collaborated with two first-grade teachers as subject matter experts. Seven semi-structured interviews with the teachers were conducted to understand major pain points for the first-grade curriculum, learn the needs and characteristics of the students, and gather their iterative feedback on our design prototypes.

The teachers indicated *Number and Operations in Base Ten* as an area which the students struggled the most with in the first grade. Therefore, we scoped our instructional design process for learning outcomes targeting this area. For iterative implementation, evaluation, and improvement of the design prototypes, user experience (UX) tests were conducted with 29 first-grade students. In these tests, the students (mostly in pairs) interacted with Oscar and experienced learning activities in a controlled lab setting (see Fig. 2

(A) *Planting Flowers* Game (Student View)

(B) *Planting Flowers* Game (Screen View)

(C) *Watering Flowers* Game (Student View)

(D) *Watering Flowers* Game (Screen View)

Fig. 2 Overview of the games. **A** *Planting Flowers* game (Student view). **B** *Planting Flowers* game (Screen view). **C** *Watering Flowers* game (Student view). **D** *Watering Flowers* game (Screen view)

for a sample view from the UX tests, where a wizard (i.e., a research team member) controlled Oscar's animations and its verbal interactions with children). Multiple UX researchers and designers, in a separate observation room, monitored the students in real time with video and audio feeds. Based on their observations, they offered revision suggestions for the next iteration of the prototypes.

After these iteration cycles, the design and development of a core set of learning games was finalized (see Fig. 2 for an overview of the games). The experience begins with a short introduction, where students knock on a door projected on the wall and Oscar emerges from the door. Oscar greets students and asks for their names. Oscar addresses students by their names throughout the game. The introduction follows a brief *Warm-up* game (red light, green light) to create a social rapport between Oscar and the students as well as to energize them through physical activities (e.g., jumping when Oscar says, "green light" and standing still when Oscar says, "red light"). Upon completion of the *Warm-up* game, Oscar introduces students to an overarching goal of creating a lively meadow for him. Oscar then asks students how to make his meadow more appealing and guides them towards the idea of flowers. The *Planting Flowers* game begins. During this game, Oscar suggests the number of flowers that students should grow by placing the correct number of physical flowerpots against the projection wall (see Fig. 2A and B). Children have two types of pots to select from: small pots (1 s pots) that grow a plant with a single digital flower and larger pots (10 s pots) that grow a plant with ten flowers. Students submit their answer to Oscar by placing the flowerpots next to the projected wall. If correct, flower buds grow in the location of the pots (physical to digital experiences transformation, see Fig. 2A). If not correct, Oscar provides basic scaffolding for the next trial (e.g., "Looks like we don't have enough pots yet").

Upon completion of the *Planting Flowers* game, Oscar informs students that the flowers are still buds, and they need help to continue growing. Oscar prompts students to suggest what is needed for flowers to bloom. Students respond with suggestions and when they say "water," Oscar brings a watering can. The *Watering Flowers* game begins. In this game, students are presented with a series of math problems (presented as clues) targeting the relevant learning outcomes (e.g., "Find the number that is 5 more than 10") and asked to locate correct answers on a projected number grid by tapping the right number (see Fig. 2C and D). Each correct answer puts a little more water into Oscar's watering can. After completing four clues successfully, the can fills up and Oscar waters the flowers, which triggers the buds to bloom.

Student engagement in Kid Space

Despite different definitions of student engagement for different learning settings (Fredericks et al., 2004; Reeve & Tseng, 2011; Pekrun & Linnenbrink-Garcia, 2012), there is a common understanding of the relationship between learning and engagement in the literature: "A student who is engaged is primed to learn; a student who is disengaged is not" (D'Mello, 2021, p. 80). In other words, engagement is a prerequisite for learning (Chiu, 2021). Therefore, improving and sustaining student engagement is critical in the design of any learning technologies—especially for young learners who "... have short attention spans and a lot of physical energy" (Shin, 2006, p. 3).

In Kid Space, we adopt a multi-componential perspective for engagement and define it in three different pillars: (1) behavioral engagement, (2) emotional engagement, and (3) social engagement. Behavioral engagement refers to “effort and persistence [in learning tasks], with an emphasis on the amount or quantity of engagement rather than its quality (Fredricks et al., 2004; Pintrich, 2000)” (as cited in Pekrun & Linnenbrink-Garcia, 2012, p. 266). Within behavioral engagement, there are three sub-pillars including task engagement (whether a student is on or off task), focus of attention (where the student’s attention is focused, as a measure for screen time), and physical activity (whether the student is physically active). Emotional engagement is described as the student’s learning-related affective states during the learning tasks based on each quadrant of the circumplex model (Russell, 1980), including being excited, satisfied, bored, or confused. As adopted from Pekrun and Linnenbrink-Garcia (2012), social engagement is defined as verbal or non-verbal interactions between collaborating peers. With social engagement, the aim is to understand the level of interactions between student–student and student–Oscar (as a peer learner).

In this current study, using a functional prototype of Kid Space, we aim to investigate these three pillars of engagement on a small sample of students and understand students’ experiences with Kid Space to iteratively improve its technology and usage for upcoming studies.

Methods

Research design and questions

Kid Space learning experiences empower young children to use their whole body and immerse themselves into computing experiences with less screen time, but with more physical activity and interactivity through hands-on and social learning. With the level of interactivity, physicality, and social learning, we hypothesized that we would observe positive behavioral, emotional, and social engagement of students during Kid Space learning experiences. To investigate this hypothesis, an exploratory case study (Merriam, 1991; Yin, 1994) was designed using a formative research method (Reigeluth & Frick, 1999; Reigeluth & An, 2009) with the first-grade students from two classrooms in an elementary school in the northwestern United States.

Out of the 37 students, 26 agreed to participate in the study with signed consent forms from their parents. However, we were able to schedule learning sessions with only ten of these students from one classroom as the school was later closed due to the COVID-19 pandemic. Using this limited number of students in this case study, we aimed to address the following research questions:

1. To what extent were the students on task, attending to the projected wall versus other physical elements, and physically active during Kid Space learning experiences?
2. To what extent were the students excited, satisfied, confused, or bored during Kid Space learning experiences?
3. To what extent were the students directly interacting with their peers and Oscar during Kid Space learning experiences?

These preliminary engagement insights and key learnings from this case study will help us iterate on the design for a larger scale longitudinal study, when the school gets back to its normal schedule.

Research setup

Layout of interaction space and sensing hardware

Prior to the study, the physical setup of Kid Space was deployed in a classroom in the school. Due to the complexity of the setup, the process took several weeks involving rigorous testing and debugging sessions for seamless operations. Figure 3 shows a sample layout of interaction space and sensing hardware with the labels 1–8 denoting their locations in the physical space: (1) and (2) cameras for pose and gesture detection as well as location tracking; (3) and (4) cameras for face identification; (5) LiDAR for touch detection on the wall; (6) camera for detecting students' interactions with objects close to the projection wall (e.g., locations of flowerpots used in the *Planting Flowers* game); (7) projector to display 3D scenes on the wall; and (8) microphone array for ambient audio detection. Additionally, all participants in the space wore wireless microphones to capture their speech during the learning experiences.

Wizard interface for Oscar

We ultimately aim for a fully automated system that recognizes children's activities and speech to drive the game logic, animations, and dialog outputs. However, given the state of development during the deployment, we incorporated a human (i.e., a wizard) to facilitate dialogue interactions. All other multimodal recognition capabilities were automated. The wizard was co-located in the space where learning experiences took place, with room dividers in between. The wizard monitored student activity and speech via live feeds from

Fig. 3 Layout of interaction space and sensing hardware

1. Current Position Indicator
2. Current Sub-activity Dialogues
3. Upcoming Sub-activity Dialogues
4. Additional Dialogues Independent of Flow
5. Custom Dialogue Entries
6. Current Answer and Its Evaluation

Fig. 4 Wizard interface

the cameras with audio. To trigger relevant responses from the wizard, a custom-built “Wizard Interface” was used (see Fig. 4).

In this interface, gameplay was organized in a series of labeled sub-activities, and each of these was further divided into pairs of observed actions and dialogues. Upon understanding that a specific action had occurred, the wizard would click the corresponding dialog button to trigger speech and actions from the game logic. For example, when the children completed an action requested by Oscar, the wizard could trigger the response, “Thank you!” along with an animation. Additionally, when it was appropriate in the game logic, sub-activities advanced automatically on button clicks. There were also buttons for out-of-context utterances (i.e., “Yes”, “No”, “Please repeat yourself”, etc.) as well as a text box for custom dialog entries. These were needed occasionally to compensate for a technical error or discrepancy in the pre-scripted flow of the games. However, the wizard was generally able to rely on the scripted game flow to complete a session.

Learning sessions

As the learning experiences required groups of students working together, the classroom teacher was asked to form groups of two students for this deployment. The teacher used math ability level (level 1: easy content for numbers 1–30; level 2: moderate-level content for numbers 31–50; level 3: difficult content for numbers 51–100) and social dynamics of the students (the extent to which pairs could seamlessly work as a team) as metrics to assign the participating ten students into five different groups. Based on this assignment, there were two groups in level 1, one group in level 2, and two groups in level 3. To facilitate the learning sessions, the school employed an instructional assistant dedicated for the study. The assistant facilitated some critical aspects of learning, especially when human intelligence was needed (human-AI collaboration). These aspects included advanced pedagogical interventions, including scaffolding for students who struggled answering questions. Prior to the learning sessions, we conducted a detailed training session with the assistant and provided multiple opportunities to practice her role in the real setup.

Before each learning session, the researchers conducted technical tests and a dry run with the instructional assistant. The assistant then accompanied each group of students from their classroom to the Kid Space classroom. The assistant completed pre-session study materials as a part of the research metrics for the students (see Fig. 5). The research

Fig. 5 Summary of research procedures

Fig. 6 A sample picture from the learning sessions

team then helped the students to wear their color-coded vests (red and blue) and their head-worn microphones to facilitate data capture and the identification of each child in the space. During the sessions, the assistant also wore a microphone, as her utterances were also incorporated into our natural language understanding system. The actual learning session for each group took around 20 min (see Fig. 6 for a sample picture from the session). At the end of each session, the research team helped the students remove all equipment and completed after-session study instruments as a part of the research metrics for the students.

Data collection and analysis

Data collection

The data collection was completed in three different days during one week at the school (see Fig. 7 for an overview of the data collection). During the data collection week, two of the students (S1 and S8) were absent and did not participate. The wizard customized

Content Difficulty	Difficult Level				Moderate Level		Easy Level			
Student IDs and Groups	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10
Attendance ✓: Attended X: Absent	X	✓	✓	✓	✓	✓	✓	X	✓	✓
Timelines Day and Time	Day 1: 9:40–10:10		Day 2: 9:40–10:10		Day 2: 10:10–10:40		Day 3: 9:40–10:10		Day 3: 9:40–10:10	
Warm-Up Game # of rounds played	1		1		1		1		1	
Planting Flowers Game # of rounds played	2		2		3		2		2	
Watering Flowers Game # of rounds played	4		3		2		2		1	

Fig. 7 Overview of data collection

the number of rounds for the *Planting Flowers* and *Watering Flowers* games based on the students' willingness to do another round or the time remaining for the session (which was highly impacted by the students in level 1 as they spent more time answering the questions). To address the research questions of the study, multiple sources of data were collected (see Table 1 for an overview of data collection instruments).

Data analysis

Observational, interview, and questionnaire data

All observational, interview, and questionnaire data were consolidated and organized for an in-depth analysis. For the observational data, color-coded graphs were created to summarize the results for each participant. Similarly, the interview results were outlined with direct quotes whenever applicable. Note that before utilizing the questionnaires, we tested their validity on five children in the first grade (different children from the study participants) through the “think-aloud protocol” (Groves et al., 2009, p. 264) (i.e., asking children to explain what they understood from each question and how they would answer) and refined the questions accordingly. We also had expert reviews by two first-grade teachers to finalize these questions. After the administration of these questionnaires with the students and the teacher involved in the study, we calculated internal consistency metrics (Cronbach's Alpha) for both Math Attitude (as completed by the students) and Social Behavior (as completed by the teacher) and got 0.65 and 0.87 accordingly. Note that Cronbach's Alpha ≥ 0.6 means a questionnaire is accurately measuring the variable of interest (Griethuijsen, 2014).

Multimodal data (audio and video)

The Human Expert Labeling Process (HELP) (Aslan et al., 2017) was incorporated to train three annotators in labeling the audio and multi-view video data to quantify various dimensions of student engagement during the learning sessions. Third-party annotators were utilized to rigorously watch the session videos with audio and label the data by marking the time boundaries (start and end times of annotated labels) and the relevant tags for the

Table 1 Summary of data collection instruments

Research constructs	Instrument	Reference	Completed by	Frequency
<i>Research metrics for teacher profile</i>				
Teaching experience, perceptions about technology, and teaching style	Semi-Structured Interview Protocol		Researcher	Once
Teaching style, classroom dynamics, learning environment	Unstructured observation protocol for math classes		Researcher	Three times
<i>Research metrics for instructional assistant profile</i>				
Teaching experience and perceptions about technology	Semi-structured interview protocol		Researcher	Once
<i>Research metrics for student profile</i>				
Math attitude	Adopted items from math attitude scale	Krinzinger et al. (2009)	Student	Once
Social behavior	Adopted items from the temperament of early childhood scale	Rowe and Plomin (1977)	Teacher	Once
<i>Research metrics for student engagement</i>				
Students' emotions (Before/After)	Adopted items from smileyometer scale	Alhussayen et al. (2015)	Student	Every session
	Structured observation protocol for learning sessions		Instructional assistant	Every session
Students' behavioral, emotional, and social engagement	Semi-structured observation protocol for learning sessions		Researcher	Every session
	Structured observation protocol for learning sessions		Instructional assistant	Every session
	Semi-structured interview protocol for learning sessions		Instructional assistant	Once
	Multimodal data (Audio and Video)		Data collector	Throughout all sessions

labels on the data (labeling state changes, e.g., switching from “On Task” to “Off Task” for task engagement). In addition to providing the labels, the annotators were also required to explain the rationale for each label they provided. This ensured that the annotators were cognitively engaged during the labeling process.

Due to the subjective nature of behavioral, emotional, and social engagement, the same data were annotated by three different annotators. However, focus of attention and physical activity labeling was more objective in nature, hence required only a single annotator. Operational definitions of the labeling dimensions along with the specific labels used are outlined in Table 2. Note that in addition to these labels, we also included “Cannot Decide” and “Not Available” labels, which were used by the annotators if they could not decide which label to assign or when the data were not available for labeling (e.g., due to technical issues). In addition to these labels, one annotator also labeled the start and end times of the three games so that the results could be outlined with regards to the different games: *Warm-up*, *Planting Flowers*, and *Watering Flowers*.

Before conducting analysis on the labeled data, first the interrater agreement was computed as a sanity check among the annotators for the labels requiring multiple-annotator inputs. Since the annotation included identification of segments in lieu of using pre-defined segments with fixed durations, it was necessary to align the multiple annotations. We divided the labels into 2-s windows with 1-s overlaps and performed analysis using Gwet’s AC1 as a metric (Gwet, 2014). The window size was determined based on the preliminary analysis of the data, considering uninterrupted duration of each label.

Table 2 Engagement labeling dimensions and operational definitions of labels

Behavioral engagement	
Task engagement	Whether the student is (1) On Task (e.g., jumping during <i>Warm-up</i> game, working with pots/number grid, listening to the assistant’s instructions/interventions, asking questions about learning task, etc.) or (2) Off Task
Focus of attention	Whether the student’s focus of attention (i.e., based on where the student is looking at using head pose as a proxy) is on (1) Projected Wall (Oscar/content), (2) Instructional Assistant, (3) Flowerpots, (4) Peer, or (5) Other (i.e., something else)
Physical activity	Whether the student is (1) Stationary (e.g., standing still, sitting, etc.) or (2) Physically Active (e.g., jumping, walking, etc.)
Emotional engagement	
Emotional states	Whether the student is (1) Satisfied (if the student is showing neutral or positive emotions during the learning task [except for excitement which is represented in another category]), (2) Excited (if the student is excited about the learning task, (3) Confused (if the student is getting confused during the learning task—in some cases, this state might include other negative states, such as frustration), (4) Bored (if the student is getting bored during the learning task), or (5) Other (if another emotional state is observed different than the ones listed)
Social engagement	
Social engagement with oscar	Whether the student shows verbal/non-verbal cues indicating a (1) Direct Interaction with Oscar (e.g., talking, physically interacting [e.g., touch], maintaining eye contact, etc.) or (2) No Direct Interaction with Oscar
Social engagement with peer	Whether the student shows verbal/non-verbal cues indicating a (1) Direct Interaction with Peer (e.g., talking/listening, physically interacting [e.g., touch], maintaining eye contact, etc.) or (2) No Direct Interaction with Peer

Computed AC1 scores showed that there was a high agreement among the annotators across all engagement dimensions with coefficient values ranging from 0.73 to 0.89 with standard error ranging from 0.003 to 0.005. As there was high agreement among the annotators, majority voting (labels that at least two out of three annotators agreed upon) was used to come up with final labels for each engagement dimension and the data was analyzed accordingly. The data points with no agreement (6.2% of the entire data) were removed from further analysis.

Internal validity, reliability, and external validity

As suggested by Merriam (1991), this study used mixed methods for data collection and analysis from multiple sources and multiple participants for triangulation to ensure internal validity (i.e., trust value) and reliability (i.e., consistency). Due to the small sample size of the study, how to interpret external validity (i.e., transferability) of the results is critical for our readers. Towards this end, we provided a rich description of the research results and established the typicality of the sample by providing specific information about profiles of the teacher, instructional assistant, and students so that the readers have a baseline information for making the judgement for how the results could be generalizable to other cases, as suggested by Merriam.

Results

Meet the teacher, instructional assistant, and students

The initial interview with Melissa revealed the following details about her role as a teacher (see Fig. 8A). Although she reported a positive attitude towards educational technology, she believed that it could result in some challenges for monitoring student engagement.

Fig. 8 Profile of **A** Teacher Melissa and **B** Instructional assistant Mary

Similarly, the initial interview with Mary outlined her background as an instructional assistant (see Fig. 8B). She indicated that she used a variety of educational technologies for teaching and learning. Additionally, as she previously worked in the school, she was very familiar with the school culture and structure.

In addition to the teacher and the instructional assistant, we investigated the important characteristics of the students who were involved in the study as these characteristics could have potential impact on the results. Math Attitude Scale as self-reported by the students (see Fig. 9A) and Social Behavior Scale as reported by the teacher for all students (see Fig. 9B) outlined important details about the students' characteristics.

As these graphs illustrate, some students scored higher (positive) in the Math Attitude Scale and lower (negative) in the Social Behavior Scale. For instance, as shown in Fig. 9, S2 had a very positive attitude towards math but he had relatively low ratings for social

Fig. 9 Profile of the students based on **A** Math attitude scale and **B** social behavior scale. *Note.* A white square on the graph indicates that there was no answer provided. Additionally, in the original scale, the last two questions were measuring the same construct with one positive (i.e., feeling fine) and one negative (i.e., feeling worried) statement. They were rephrased in the figure to be able to align the visual representation

behaviors. Aligned with this result, our in-session observations revealed that he showed the best performance in solving the questions, while he was noted as one of the shiest students in the student group. On the contrary, some students scored lower in Math Attitude Scale and higher in the Social Behavior Scale (e.g., S5). Aligned with this result, S5 was identified as the student who showed minor signs of boredom during the Kid Space learning experiences. Finally, some students, like S7 and S10, scored low on both scales. Our in-session observations showed that like S2, S7 was one of the shiest students in the student group and he had a lot of difficulty answering the questions. Similarly, the instructional assistant identified S10 as the only student who showed occasional off-task behaviors in the session and had a lot of difficulty answering the questions.

RQ1: To what extent were the students on task, attending to projected wall versus other physical elements, and physically active during Kid Space learning experiences?

[With Kid Space learning experiences], [w]e do not need paper, pencils - we do not need a lot of supplies for opening up opportunities. ... Which parents would not want their kids ... physically active and engaged [during learning]? ... Parents do not like their kids to play video games, but this is an educational game, so it is entirely different. (Instructional Assistant Mary, Interview, 2020).

To understand the students' behavioral engagement during the learning sessions, we investigated three important pillars: (1) task engagement, (2) focus of attention, and (3) physical activity (refer to Table 2 for a detailed description of each). In this section, the results for each of these pillars will be outlined.

The first pillar, task engagement, provided insights about the extent to which the students stayed on task throughout the Kid Space learning experiences. Two independent data sources were used to evaluate task engagement: (1) coarse-level labels from the instructional assistant as provided at the end of each session for each student and (2) fine-grained-level labels from the annotators. The first provided us a qualitative measure for task engagement from the instructional assistant's observations during the sessions (see Fig. 10A), whereas the second provided a quantitative measure for each game as judged by the multiple annotators, supporting the qualitative data (see Fig. 10B). For quantitative analysis, we employed the final fine-grained labels for different pillars and obtained average values for each student. Using these averages, we computed the mean (M) values over all students, as well as the variability among the students by computing the standard deviations (SD).

As shown in Fig. 10A, except for one student (S10 had scored low in both Math Attitude and Social Behaviors scales, as explained previously) who became distracted occasionally, the students showed on-task behaviors the majority of the time during the learning sessions. This is supported by the fine-grained-level labels provided by the annotators as well (see Fig. 10B). Despite the slight differences between games, the students were on task 99.9% ($SD=0.2\%$) of the time during the *Warm-up* game, 98.8% ($SD=1.8\%$) during the *Planting Flowers* game, and 98.9% ($SD=2.6\%$) during the *Watering Flowers* game. Aligned with the qualitative results, when we specifically checked the *Planting Flowers* and *Watering Flowers* games, the slightly lower on-task percentages were mainly caused by S10. The standard deviation values increased by 1.6% and 2.2% respectively for the two games, when S10 was included in the inter-subject variability calculations.

Fig. 10 Students' task engagement labeled by the **A** Instructional assistant, and **B** annotators

In addition to the task engagement, we also investigated the focus of attention and physical activity of the students as the other pillars of behavioral engagement. These pillars are particularly important because the Kid Space learning experiences were designed in a way that the students could learn through multimodal physical interactions using hands-on learning experiences, as opposed to sitting stationary in front of a device and staring at the screen.

Towards this end, the focus of attention pillar provided insights about what the students were mostly attending to during the Kid Space learning experiences: (1) projected wall (including Oscar), (2) flowerpots, (3) instructional assistant, (4) peer, or (5) other (something else), whereas the physical activity pillar provided insights about the extent to which the students were physically active during the Kid Space learning experiences. Using the labels from the annotators, Fig. 11 summarizes the game-specific distributions for (A) focus of attention and (B) physical activity of the students.

The results in Fig. 11A show that when only utilizing the labels where the students were labeled as on task by the annotators, the students were engaged with the projected wall 74.2% of the time, whereas 25.8% of the time their focus of attention was on some other physical elements of the Kid Space learning experiences (e.g., flowerpots, instructional assistant, peer, etc.). However, there is an observable difference in these percentages across different games as a function of their interaction design. In the *Warm-up* and *Watering Flowers* games, in which the interactions were mostly digital, the focus of attention on the projected wall was as high as 91.3% (SD=9.0%) and 88.1% (SD=14.7%), respectively. However, in the *Planting Flowers* game, in which the interactions were both digital and physical (through physical manipulatives), it is as low as 43.1% (SD=15.4%). More importantly, when investigating the patterns of the focus of attention on the projected wall, we identified that it occurred in brief bursts across all games, where the average duration of uninterrupted attention was 12 s (SD=13) for *Warm-up*, 8 s (SD=4) for *Planting Flowers*, and 11 s (SD=10) for *Watering Flowers*.

Fig. 11 Students' **A** Focus of attention and **B** physical activity as labeled by the annotators

In addition to the focus of attention, we also investigated the level of physical activity as another pillar of behavioral engagement. In her interview, the instructional assistant Mary highlighted the physicality aspect of Kid Space learning experiences as follows:

[G]etting kids physical instead of sitting on a seat in the classroom. ... They were able to start with jumping or dancing; flowerpots - standing back, moving [around], moving the pots for the next question; pushing the number grid and stepping back and waiting for the next clue - there were a lot of physical aspects [throughout Kid Space learning experiences]. (Instructional Assistant Mary, Interview, 2020).

When investigating the extent to which the students were physically active during the Kid Space learning experiences, we found out that more than half of the time (53.9%), the students were physically active (e.g., jumping, walking, etc.) (see Fig. 11B). More importantly, there were only slight differences in terms of physical activity across the games: 43.7% (SD=24.9%) in *Warm-up*, 61.8% (SD=8.0%) in *Planting Flowers*, and 56.2% (SD=26.6%) in *Watering Flowers*.

RQ2: To what extent were the students excited, satisfied, confused, and bored during Kid Space learning experiences?

The students are associating learning [as] boring and not motivating; but when interacting with Oscar, they forgot learning and they were thinking they were just playing a game. Oscar getting out from the door ... Oscar saying their names ... Seeing him go and get the watering can, ... [after] ... putting the pot on the wall, ... seeing the

flowers bloom: A plastic flowerpot can grow a digital flower. (Instructional Assistant Mary, Interview, 2020).

To understand the emotional engagement of the students during the Kid Space learning experiences, two independent data sources were utilized: (1) coarse-level labels from the instructional assistant and the students for overall student emotions before and after the learning sessions and (2) fine-grained-level labels from the annotators for the emotional states of the students during the learning sessions. The first data source provided us a qualitative measure for overall emotional engagement from the instructional assistant's observations and the students' self-reports, whereas the second one provided a quantitative measure for each game as judged by the multiple annotators, supporting the qualitative data. Similar to the behavioral engagement, we employed the final fine-grained labels to obtain student averages, and then computed mean and standard deviations over all students.

For the first data source, prior to and after each learning session, the instructional assistant asked the individual students to report how they were feeling (self-reports). She also noted down her observed emotions for each student independently. Figure 12A shows whether reported or observed emotions were positive, somewhat positive, or negative before and after each session. As the figure demonstrates, there was no negative change

Fig. 12 Students' emotional engagement, as labeled by the **A** Instructional assistant and students, and **B** annotators

of student emotions (from positive to negative) when comparing before and after data. In fact, the students either kept their somewhat positive/positive emotions (e.g., S2, S7) or improved from somewhat positive to positive emotions (e.g., S10) after the learning sessions.

To further investigate the reasoning behind the students' emotions, the instructional assistant asked the students what made them feel that way after each session. The students reported (1) having fun in both the *Planting Flowers* and *Watering Flowers* games (S2, S3, S5, S6, S7, S10), (2) enjoying talking and playing with Oscar (S4, S9), and (3) having fun when playing red light/green light during the *Warm-up* game (S9). Supporting these results, when asked if they wanted to come back for another session, all of the students indicated that they wanted to play with Oscar again, as shown in Fig. 12A.

In addition to the coarse-level labels from the instructional assistant and the students, the fine-grained-level labels from the annotators were analyzed for emotional engagement: how much the students were excited, satisfied, bored, or confused during the learning sessions (see Table 2 for a detailed description). As shown in Fig. 12B, the results indicate that with some slight differences across the games, the students were feeling satisfied the majority of the time ($M=88.9\%$, $SD=2.0\%$), whereas they felt excited ($M=6.1\%$, $SD=0.6\%$), confused ($M=2.3\%$, $SD=0.2\%$), and bored ($M=0.3\%$, $SD=0.07\%$) for a small portion of the time. In addition to these emotional states, the annotators also provided some other emotions they observed during the sessions in the "Other" label (2.4% of the time). These emotions included surprised, disappointed, shy, stressed, annoyed, and upset.

As the results in Fig. 12B demonstrate, the most common negative emotion was confusion. When further investigating what resulted in the students' confusion during the learning sessions, our observational notes showed that some of the students struggled when solving the questions as the question difficulty did not match their ability levels. In these cases, Oscar was unable to adjust the content in real time based on the ability level of the students or provide necessary scaffolding to the students to solve these questions. This is due to the currently limited capabilities of the system. Instead, the instructional assistant often intervened with these students to provide the necessary scaffolds.

RQ3: To what extent were the students directly interacting with their peers and Oscar during Kid Space learning experiences?

[Kid Space] ... allowed [the] students to learn in a different way. It allows students [to] stand up and work with a partner - fun, imaginative, [and] social learning, versus sitting in a classroom and listening to a teacher only. ... Kids started getting very into it: high fives, looking at each other and smiling, counting together, laughter together, [and] running to push the numbers. It was a great interaction with the kids. (Instructional Assistant Mary, Interview, 2020).

To evaluate the students' social engagement during the learning sessions, we utilized two data sources using the labels from the annotators: students' direct interactions with (1) each other and (2) with Oscar as a peer learner (refer to Table 2 for a detailed description). For both data sources, we used fine-grained-level labels and obtained student averages. Then we ran quantitative analysis to compute mean and standard deviation over the students. For the first data source, we investigated the extent to which the students showed verbal/non-verbal cues indicating direct interaction with their peers during the learning sessions (e.g., talking, physically interacting [e.g., high fives], maintaining eye contact, etc.). As shown in Fig. 13A, the levels of

Fig. 13 Students' social engagement, as labeled by the annotators with **A** Peers, and **B** Oscar

direct interactions with peers varied across the games: 1% of the time in the *Warm-up* game (as expected based on the interaction design of the game), 5.9% in the *Planting Flowers* game, and 21% in the *Watering Flowers* game.

For the second data source, the level of direct interactions of the students with Oscar was investigated. The results showed that the trend was the opposite of the first data source, across the games. As illustrated in Fig. 13B, during the *Warm-up* game, the level of direct interactions with Oscar was the highest ($M=90.4\%$, $SD=2.9\%$) as the students directly played together with Oscar in this game (i.e., Oscar was the content itself), whereas in the *Watering Flowers* and *Planting Flowers* games, direct interactions with Oscar were 27.0% ($SD=8.7\%$) and 30.0% ($SD=6.5\%$) of the time, respectively.

To further understand the extent to which the students perceived Oscar as a conversational peer learner, the observational notes from the annotators were used to investigate the ways the students interacted with Oscar. These notes showed that during the learning sessions, the students were (1) asking intellectual questions to Oscar (e.g., “What is a meadow, Oscar?” [S5, Learning Session, 2020]), (2) trying to get confirmation from Oscar (e.g., “Oscar, can you hear me?” [S4, Learning Session, 2020]), (3) referring to Oscar’s digital space (e.g., “What do you have there, Oscar?” [S9, Learning Session, 2020]), and (4) showing emotional attachment to Oscar (e.g., “Do not leave us, Oscar!” [S6, Learning Session, 2020]).

Implications

Physically active, engaging with less screen time, and social learning experiences

The results show that the students were on task 99.2% of the time. However, when investigating the times that students were on task, their screen time (i.e., focus of attention on the projected wall) was limited to 74.2% of the time with major differences across the games based on the game-specific interaction design. In the *Warm-up* and *Watering Flowers* games, the interactions were mostly digital, and the screen time was as high as 91.3% and 88.1%, respectively. In the *Planting Flowers* game, the interactions were both digital and physical (through physical manipulatives), so the screen time was as low as 43.1%. This signifies the importance of the use of physical manipulatives in the game design for future iterations of the experience. Moreover, when further investigating the patterns of screen time, we found that there were short periods of focus on the projected wall throughout the learning sessions (about 12 s on average) instead of the students constantly staring at the screen without any interruptions. Additionally, when the students were on task, more than half of the time (53.9%) they were physically active (e.g., jumping, walking, etc.). These preliminary results are promising because in traditional technology-mediated learning environments with a device (e.g., PC, tablet, etc.), one would expect almost 100% screen engagement while the students are sitting stationary in front of their computing devices and labeled as on task (Aslan et al., 2019).

From an emotional engagement perspective, the majority of the time (95.0%) the students demonstrated positive emotions (satisfied/excited) during the learning experiences. More importantly, the results indicated that there were several episodes of excitement (corresponding to 6.1% of the time), which is a relatively rare emotion to observe in learning settings (D'Mello & Graesser, 2013; Aslan et al., 2017). These positive emotional engagement results are in line with the previous research reporting positive student experience when engaged with multimodal interactions in immersive learning settings (Zhao et al., 2018).

From a social engagement perspective, a huge difference was observed when the students were attending the sessions on their own versus with a peer. When the students were working with their peers, our observational results showed that the experiences were enriched with social interactions, including giving high fives to nurture team spirit and motivation, providing scaffolding to each other whenever needed, and having more fun together with their peers. These results reinforce the critical aspect of collaborative game design for future implementations, since teamwork and fun were previously identified as the most frequent themes of motivational engagement triggers in immersive game-based learning (Duncan, 2020).

In addition to the direct interactions between the students, we also evaluated the extent to which the students interacted with Oscar as a peer learner throughout the experiences (e.g., talking, physically interacting [e.g., giving high fives], maintaining eye contact, etc.). During the *Planting Flowers* and *Watering Flowers* games, 28.5% of the time the students directly interacted with Oscar, which is about twice the time they directly interacted with each other. Even though Oscar had relatively predefined utterances and actions when interacting with the students as triggered by the human wizard, the students' interactions with Oscar showed that they perceived Oscar as a conversational peer learner. The students asked intellectual questions of Oscar, tried to get confirmation from Oscar, referred to Oscar's digital space, and showed emotional attachment to Oscar during the learning

sessions. These results are aligned with the previous research suggesting that “children respond to CAs [Conversational Agents] socially and treat CAs as companions or guides” (Nilsen, 2019; Pantoja et al., 2019; Vogt et al., 2017, as cited in Xu & Warschauer, 2020, p. 1). In this deployment, Oscar’s verbal interactions were triggered by the wizard. For future deployments, we plan to use these results as a baseline and aim to meet this baseline with Oscar’s autonomous dialogue interactions.

Need for real-time personalization

Although the students were feeling positive the majority of the time during the learning sessions, they were labeled as confused 2.3% of the time. An optimal amount of confusion in learning is beneficial to provide enough challenge for students to improve performance (Rodrigo et al., 2010; D’Mello et al., 2014), however, higher levels of unaddressed confusion could potentially result in giving up and showing off-task behaviors during learning. When specifically investigating the confusion instances to understand the root causes, the observational results showed that some of the students struggled because the questions were much harder to solve than their ability level. Even though we created a simple level of personalization by assigning the students to specific levels and using numbers that presumably could work for these levels, a need for performance-based, real-time personalization emerged. Additionally, for cases where the students struggled, the instructional assistant had to intervene often to provide necessary scaffolding. This signifies a need for Oscar to provide real-time personalized scaffolding whenever needed for future deployments, since prior research also show that personalized scaffolding significantly increases learner engagement (Winkler et al., 2020). This might require Oscar to understand common errors that students make when addressing different types of questions and craft relevant scaffolding methods to address these errors in real time.

In addition to confusion, another negative emotion, boredom, was observed extremely rarely (0.3%) during the learning sessions. Although this small percentage of boredom is not necessarily an important issue to address for this deployment, we suspect it could become a larger issue when students participate in multiple sessions in future deployments. Currently, there is only one warm-up game and two learning games. The wizard was able to adjust the number of rounds in each game based on the remaining time and the students’ willingness to participate. However, during the sessions, some of the students asked for different objects to be included in the meadow in addition to the flowers and expressed interest in playing another game. Diverse content is important for sustained engagement in the longer run for successfully personalizing the experiences. If the students are getting bored with a certain game, Oscar should be able to bring in other games based on the students’ interests.

Limitations

Due to the COVID-19 pandemic, we were able to complete the school deployment with a set of the students who attended only one session. Therefore, novelty effect (Hamari et al., 2014) could have played an important role in the results. There is a need to validate these results using a larger number of students participating in multiple sessions, as part of a longitudinal study, once the school returns to its normal operations. Having multiple sessions will also enable us to understand the impact of the Kid Space learning experiences on

student performance, which is shown to be positively correlated with engagement (Henrie et al., 2015).

Conclusion and future work

The results of this exploratory case study showed that the students demonstrated high levels of engagement with less focus of attention on screen and more physical activity during the Kid Space learning experiences. These positive engagement results were accompanied by several episodes of student excitement—a rare emotion to observe in learning settings. In addition to these promising results, the study enabled the discovery of actionable insights to improve Kid Space for future deployments (e.g., the need for real-time personalization). As a future direction, leveraging the data collected from this study, we would like to advance our sensemaking and dialogue technologies so that we can transition from human-wizard-driven experiences (particularly for verbal interactions) to more autonomous experiences incorporating human-AI collaboration. With these advancements in technological interfaces and improvements in the experience design, we plan to deploy Kid Space with more students in multiple sessions once the school returns to its normal operations.

Acknowledgements We would like to acknowledge the tremendous support that our Intel design and development team provided to enable this study: Special thanks to Sai Prasad, David I. Gonzalez Aguirre, Hector A. Cordourier Maruri, Pete A. Denman, Willem M. Beltman, Julio C. Zamora Esquivel, Shachi H. Kumar, Rahul C. Shah, Chieh-yih Wan, Cagri Tanriover, Paulo Lopez Meyer, Glen J. Anderson, and Parul Datta for their contributions. We would like to also acknowledge Free Orchards Elementary School in the Hillsboro School District, the school principal Karen Murphy, teachers Itzia Mendoza, Molly Scott, and Mark Anderson, and other school personnel as well as the parents and the students for their support to undertake this study.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

Ethical approval Prior to the study, the researchers got approval on an extensive privacy plan which also included approved consent forms for human subjects.

Consent to participate Informed consent (including parental consent) was obtained from all individual participants included in the study.

References

- Ahearne, C., Dilworth, S., Rollings, R., Livingstone, V., & Murray, D. (2016). Touch-screen technology usage in toddlers. *Archives of Disease in Childhood*, 101(2), 181–183.
- Alhussayen, A., Alrashed, W., & Mansor, E. I. (2015). Evaluating the user experience of playful interactive learning interfaces with children. *Procedia Manufacturing*, 3, 2318–2324.
- Allen, D., Divekar, R. R., Drozdal, J., Balagyozyan, L., Zheng, S., Song, Z., Zou, H., Tyler, J., Mou, X., Zhao, R., Zhou, H., Yue, J., Kephart, J. O., & Su, H. (2019). The Rensselaer Mandarin project—a cognitive and immersive language learning environment. *Proceedings of the AAAI Conference on Artificial Intelligence*, 33, 9845–9846.
- Anderson, G. J., Chierichetti, R., Agrawal, Ankur, Shi, M. (2015). *Ethnographic and Participatory Design Research on Smart Home Applications*. Unpublished manuscript.

- Anderson, G. J., Panneer, S., Shi, M., Marshall, C. S., Agrawal, A., Chierichetti, R., & Durham, L. M. (2018). Kid space: Interactive learning in a smart environment. In *Proceedings of the Group Interaction Frontiers in Technology* (pp. 1–9).
- American Academy of Pediatrics. (2016). Media and young minds. *Pediatrics*, 138(5), 1–6. <https://doi.org/10.1542/peds.2016-2591>
- Aslan, S., Mete, S. E., Okur, E., Oktay, E., Alyuz, N., Genc, U. E., Stanhill, D., & Arslan E. A. (2017). Human expert labeling process (HELP): Towards a reliable higher-order user state labeling process and tool to assess student engagement. *Educational Technology*, 53–59.
- Aslan, S., Alyuz, N., Tanriover, C., Mete, S. E., Okur, E., D'Mello, S. K., & Arslan Esme, A. (2019). Investigating the impact of a real-time, multimodal student engagement analytics technology in authentic classrooms. *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, 1–12.
- Arrow, J. (2019). *How to use play for learning*. <https://www.edutopia.org/article/how-use-play-learning>. Accessed 15 May 2020
- Blackwell, C. K., Lauricella, A. R., & Wartella, E. (2014). Factors influencing digital technology use in early childhood education. *Computers & Education*, 77, 82–90.
- Chiu, T. K. (2021). Applying the self-determination theory (SDT) to explain student engagement in online learning during the COVID-19 pandemic. *Journal of Research on Technology in Education*, 1–17.
- D'Mello, S., & Graesser, A. (2013). AutoTutor and affective AutoTutor: Learning by talking with cognitively and emotionally intelligent computers that talk back. *ACM Transactions on Interactive Intelligent Systems (TiiS)*, 2(4), 1–39.
- D'Mello, S., Lehman, B., Pekrun, R., & Graesser, A. (2014). Confusion can be beneficial for learning. *Learning and Instruction*, 29, 153–170.
- D'Mello, S. K. (2021). Improving student engagement in and with digital learning technologies. *OECD Digital Education Outlook 2021 Pushing the Frontiers with Artificial Intelligence, Blockchain and Robots: Pushing the Frontiers with Artificial Intelligence, Blockchain and Robots*, 79.
- Duncan, K. J. (2020). Examining the effects of immersive game-based learning on student engagement and the development of collaboration, communication, creativity and critical thinking. *TechTrends*, 64(3), 514–524.
- Gelsomini, M., Cosentino, G., Spitale, M., Gianotti, M., Fisicaro, D., Leonardi, G., Riccardi, F.k Piselli, A., Beccaluva, E., Bonadies, B., Terlizzi, L., Zinzoni, M., Alberti, S., Rebourg, C., Carulli, M., Garzotto, F., Arquilla, V., Bisson, M., Del Curto, B., & Borgegoni, M. (2019). Magika, a multisensory environment for play, education and inclusion. *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems*, 1–6.
- Ginsburg, K. R. (2007). The importance of play in promoting healthy child development and maintaining strong parent-child bonds. *Pediatrics*, 119(1), 182–191.
- Griethuijsen, R. A. L. F., Eijck, M. W., Haste, H., Brok, P. J., Skinner, N. C., Mansour, N., et al. (2014). Global patterns in students' views of science and interest in science. *Research in Science Education*, 45(4), 581–603. <https://doi.org/10.1007/s1165-014-9438-6>
- Groves, R. M., Fowler, F. J., Couper, M. P., Lepkowski, J. M., & Singer, E. (2009). *Survey methodology* (2nd ed.). John Wiley & Sons.
- Guilfoyle, C. (2013). For college and career success, start with preschool. *Policy Priorities*, 19(4), 1–7.
- Gwet, K. L. (2014). *Handbook of inter-rater reliability: The definitive guide to measuring the extent of agreement among raters*. Advanced Analytics, LLC.
- Fredricks, J. A., Blumenfeld, P. C., & Paris, A. H. (2004). School engagement: Potential of the concept, state of the evidence. *Review of Educational Research*, 74(1), 59–109.
- Hamari, J., Koivisto, J., & Sarsa, H. (2014). Does gamification work?—A literature review of empirical studies on gamification. *2014 47th Hawaii International Conference on System Sciences*, 3025–3034.
- Henrie, C. R., Halverson, L. R., & Graham, C. R. (2015). Measuring student engagement in technology-mediated learning: A review. *Computers & Education*, 90, 36–53.
- Karno, D., & Hatcher, B. (2020). Building computer supported collaborative learning environments in early childhood classrooms. *Educational Technology Research and Development*, 68(1), 249–267.
- Krinzinger, H., Kaufmann, L., & Willmes, K. R. (2009). Math anxiety and math ability in early primary school years. *Journal of Psychoeducational Assessment*, 27(3), 206–225.
- Merriam, S. B. (1991). *Case study research in education: A qualitative approach*. Jossey-Bass.
- Molenda, M. (2003). In search of the elusive ADDIE model. *Performance Improvement*, 42(5), 34–37.
- Nacher, V., Garcia-Sanjuan, F., & Jaen, J. (2016). Interactive technologies for preschool game-based instruction: Experiences and future challenges. *Entertainment Computing*, 17, 19–29.
- Nilsen, J. H. (2019). *"It knows how to not understand us!" A study on what the concept robustness entails in design of conversational agents for preschool children*, Master's thesis.

- Pekrun, R., & Linnenbrink-Garcia, L. (2012). Academic emotions and student engagement. *Handbook of research on student engagement*, 259–282. Springer.
- Pintrich, P. R. (2000). The role of goal orientation in self-regulated learning. In M. Boekarts, P. R. Pintrich, R. Pekrun, L. Linnenbrink-Garcia, & M. Zeidner (Eds.), *Handbook of self-regulation: Theory, research and applications* (pp. 451–502). San Diego, CA: Academic.
- Reeve, J., & Tseng, C. M. (2011). Agency as a fourth aspect of students' engagement during learning activities. *Contemporary Educational Psychology*, 36(4), 257–267.
- Reigeluth, C. M., & Frick, T. W. (1999). Formative research: A methodology for creating and improving design theories. In C. M. Reigeluth (Ed.), *Instructional design theories and models: A new paradigm of instructional theory*, 2 (pp. 633–651). Lawrence Erlbaum.
- Reigeluth, C. M., & An, Y. (2009). Theory building. In C. M. Reigeluth & A. A. Carr-Chellman (Eds.), *Instructional design theories and models: Building a common knowledge base*, 3 (pp. 385–386). RoutledgeFalmer.
- Rodrigo, M. M. T., Baker, R. S., & Nabos, J. Q. (2010). The relationships between sequences of affective states and learner achievement. *Proceedings of the 18th International Conference on Computers in Education*, 56–60.
- Rowe, D. C., & Plomin, R. (1977). Temperament in early childhood. *Journal of Personality Assessment*, 41(2), 150–156.
- Russell, J. A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology*, 39(6), 1161.
- Sahay, S., Kumar, S. H., Okur, E., Syed, H., & Nachman, L. (2019). Modeling Intent, Dialog Policies and Response Adaptation for Goal-Oriented Interactions. Proceedings of the 23rd Workshop on the Semantics and Pragmatics of Dialogue.
- Shapiro, J. (2015). *The American Academy of Pediatrics just changed their guidelines on kids and screen time*. Forbes. <https://www.forbes.com/sites/jordanshapiro/2015/09/30/the-american-academy-of-pediatrics-just-changed-their-guidelines-on-kids-and-screen-time/>
- Shin, J. K. (2006). Ten helpful ideas for teaching English to young learners. *UMBC Student Collection*.
- Singer, D., Golinkoff, R. M., & Hirsh-Pasek, K. (2006). *Play = learning: How play motivates and enhances children's cognitive and social-emotional growth*. Oxford University Press.
- Superti Pantoja, L., Diederich, K., Crawford, L., & Hourcade, J. P. (2019). Explorations of voice user interfaces for 3 to 4 year old children. In *Extended abstracts of the 2019 CHI conference on human factors in computing systems* (pp. 1–6).
- The National Association for the Education of Young Children. (n.d.). <https://www.naeyc.org/resources/topics/play>. Accessed 15 May 2020
- Vogt, P., De Haas, M., De Jong, C., Baxter, P., & Krahmer, E. (2017). Child-robot interactions for second language tutoring to preschool children. *Frontiers in Human Neuroscience*, 11, 73.
- Winkler, R., Hobert, S., Fischer, T., Salovaara, A., Söllner, M., & Leimeister, J. M. (2020). Engaging learners in online video lectures with dynamically scaffolding conversational agents. *Association for Information Systems*.
- Wood, E., Specht, J., Willoughby, T., & Mueller, J. (2008). Integrating computer technology in early childhood education environments: Issues raised by early childhood educators. *The Alberta Journal of Educational Research*, 54(2), 210–226.
- Xu, Y., & Warschauer, M. (2020). Exploring young children's engagement in joint reading with a conversational agent. *Proceedings of the Interaction Design and Children Conference*, 216–228.
- Yin, R. K. (1994). *Case study research: Design and methods*. Sage Publications.
- Zabatiero, J., Straker, L., Mantilla, A., Edwards, S., & Danby, S. (2018). Young children and digital technology: Australian early childhood education and care sector adults' perspectives. *Australasian Journal of Early Childhood*, 43(2), 14–22.
- Zhao, R., Wang, K., Divekar, R., Rouhani, R., Su, H., & Ji, Q. (2018). An immersive system with multimodal human-computer interaction. *2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018)*, 517–524.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Sinem Aslan (Ph.D.) is a research scientist at Intel Corporation with a research focus on exploring the intersection between AI and human learning to enable personalized learning experiences.

Ankur Agrawal is a researcher at Intel Corporation with research interests on exploring programmable materials, AI, tangibles, and physical spaces to design human-machine-material interfaces.

Nese Alyuz (Ph.D.) is a research scientist at Intel Corporation with research interests on affective computing, human computer interaction, and biometrics.

Rebecca Chierichetti is a user experience researcher and designer at Intel Corporation with a specialty on exploring boundaries between digital and physical worlds.

Lenitra M. Durham (Ph.D.) is a computer scientist at Intel Corporation with 25+ patents and over 15 years of experience within Intel Corporation prototyping compelling user experiences.

Ramesh Manuvinakurike (Ph.D.) is a research scientist at Intel Corporation with a research focus on language understanding and dialogue policy in multi-modal spoken dialogue systems.

Eda Okur is an AI/ML Research Scientist at Intel Corporation with over 10 years of experience on human computer interaction, natural language understanding, and multimodal dialogue systems.

Saurav Sahay (Ph.D.) is a research manager at Intel Corporation leading a team of researchers in the area of multi-modal dialog and interaction systems.

Sangita Sharma (Ph.D.) is a research scientist at Intel Corporation with research focus on affective computing, sense making technologies, and innovative user experiences.

John Sherry is the director of the User Experience Innovation Lab at Intel Corporation. His research team explores new paradigms for human-AI collaboration.

Giuseppe Raffa (Ph.D.) is a principal engineer at Intel Corporation focusing on Artificial Intelligence and Ambient Computing technologies.

Lama Nachman is an Intel fellow and director of the Human & AI Systems Research Lab at Intel Corporation. Her research is focused on creating contextually-aware experiences that understand users through sensing and sense making, anticipate their needs, and act on their behalf.