Computers & Education 182 (2022) 104434

Contents lists available at ScienceDirect

Computers & Education

journal homepage: www.elsevier.com/locate/compedu

Child-driven, machine-guided: Automatic scaffolding of constructionist-inspired early literacy play[1]

Ivan Sysoev [a,*], James H. Gray [a], Susan Fine [b], Sneha Priscilla Makini [a], Deb Roy [a]

[a] *MIT Media Lab, USA*
[b] *Northeastern University, USA*

ARTICLE INFO

Keywords:
Early years education
Games
Human-computer interface
Informal learning
Mobile learning

ABSTRACT

Child-driven approaches to learning, such as constructionism, can greatly contribute to children's positive relationship with the subject via personally meaningful and grounded learning experiences. However, these approaches typically need scaffolding to ensure learners' progress. Providing scaffolding is nontrivial and time-consuming, typically requiring real-time, one-on-one involvement of the scaffolder. Can scaffolding procedures be at least partially automated? We explored this question in the special case of a constructionist-inspired early literacy app designed for 4- to 6-year-old children. We created scaffolding mechanisms for word building while attempting to preserve the open-ended and child-driven nature of interaction. The system was evaluated during an 11-week-long design study in kindergarten classrooms. We found that scaffolding mechanisms facilitated creative expression and literacy-related social interactions between children, as well as enabled highly autonomous play for some of them. However, despite the scaffolding aid, children with low executive functioning (EF) and phonological awareness (PA) were prone to engage with the app in an impulsive and unsystematic manner, hindering their learning. We discuss possible strategies to mitigate the negative effects of low PA and EF. Despite these challenges, the child-driven, machine-guided approach appears to be promising.

1. Introduction

Several educational approaches, such as constructionism (Papert, 1980), rely on child-driven learning activities. These activities can be intrinsically motivating, supportive of learners' senses of agency and self-efficacy, and meaningfully connected to the learners' lives — all of which is beneficial for learning. However, to be effective, child-driven learning requires scaffolding (i.e. guidance). With learners working on their own projects, scaffolding often needs to be highly individualized, and thus labor-intensive. Progress in artificial intelligence (AI) poses an intriguing, and currently open (Kahn & Winters, 2021), question on whether scaffolding can to some extent be automated.

Early literacy learning is a valuable domain for examining this question. The role played by reading and writing makes it important that children not only master the technical skills involved, but also form positive attitude to literacy. The senses of fun and empowerment associated with child-driven approaches might help in this task. However, it is also well-established that some early literacy

[1] This work is derived from the first author's PhD dissertation (Sysoev, 2020). The source code of the app can be found at https://github.com/mitmedialab/speechblocks-ii

* Corresponding author.
E-mail address: isysoev@alum.mit.edu (I. Sysoev).

https://doi.org/10.1016/j.compedu.2022.104434
Received 6 July 2021; Received in revised form 27 December 2021; Accepted 6 January 2022
Available online 17 January 2022

skills (e.g. phonics) require explicit and systematic instruction (Castles et al., 2018) or, in child-driven case, guidance. Previous experiments with child-driven phonics software showed that such guidance can be very labor-intensive (Sysoev, 2020, section 5.6). Additionally, classroom teachers are not necessarily well-versed in phonics instruction (Castles et al., 2018). These factors present a compelling case to explore a child-driven, automatically scaffolded approach.

To do so, we implemented an early literacy app SpeechBlocks II, aimed at facilitating learning of encoding and decoding. In SpeechBlocks II, children spell words of their own choice and compose pictures out of sprites associated with the words. The scaffolding system learns which words children would like to make through several input mechanisms, then guides them through the spelling process. We explored SpeechBlocks in an 11-week-long classroom-based study utilizing Design-Based Research method.

Following Sandoval's (2014), our study was guided by several conjectures. Our core conjecture was that the scaffolding mechanisms would support autonomous play with the app while maintaining the senses of playfulness, expressive freedom and agency. We expected that there will be a relatively small need for adult involvement, qualitatively different from what we saw in our earlier studies with a non-scaffolded child-driven literacy app, SpeechBlocks (Sysoev, 2020, section 5.6). We further conjectured that play with the system would facilitate the learning of phonological awareness (PA). We also anticipated that automatic scaffolding may affect literacy-related social interactions, which were prominent with SpeechBlocks (Sysoev et al., 2017). Finally, we anticipated individual differences in how children engage with the app. We looked at two variables potentially relevant to these differences: children's initial literacy skills and their self-regulation capacity, viewed through the concept of executive functioning (EF).

We found that for some children, automatic scaffolding enabled highly engaged, nearly autonomous play, which was rich in expression and led to literacy-related social interactions. However, we also observed a large number of impulsive, short-term-reward-oriented behaviors. They were more likely to emerge with children who had low PA and EF and might have hindered learning from the app. We see several potential pathways for mitigating this issue. First, more sophisticated mechanisms for automatically aligning the difficulty with the child's skills could make word building more appealing for struggling learners. Second, other forms of scaffolding, e.g., ones oriented towards the emotional support of the learner, could be necessary. Third, a combination of instructionist and constructionist approaches might be beneficial. Despite these challenges, we believe that the child-driven, machine-guided approach holds significant promise since it combines learners' motivation and empowerment, Vygotskian learning and increased learner autonomy.

2. Background

2.1. Motivation for child-driven designs

A large volume of research shows that increased learners' control over learning activities results in increased motivation, effort, perceived competence, task performance, learning outcomes (Patall et al., 2008; Iyengar & Lepper, 1999) and children's willingness to participate in the activity later on their own volition (Swann & Pittman, 1977). Choice can be particularly valuable when children can align learning activities with their interests and personally meaningful subjects. For instance, incorporating children's teachers, friends, hobbies, favorite foods, etc. into math tasks led to significantly improved performance (Anand & Ross, 1987; Cordova & Lepper, 1996). Interest significantly affects the comprehension and recall of texts (Estes & Vaughan, 1973, Renninger, 1992). Children writing about a topic of interest were observed to effortlessly use much more complex sentences than in their regular schoolwork (Kelly & Safford, 2009).

Another important factor contributing to learning is the sense of empowerment. Children's perception of whether they are in control of their lives contributes to their academic achievement (Findley & Cooper, 1983, Stipek, 1980). Desires to "leave a mark in the world", to designate ownership of things, and to be like adults are considered to be key motivators for early writing (Strickland & Morrow, 1989).

Such learning approaches as Montessori (A. S. Lillard, 2017), Waldorf (Clouder & Rawson, 1998), Reggio Emilia (Edwards et al., 1998) and constructionism (Papert, 1980; Resnick, 2014, 2017) build upon child-driven, personally meaningful, empowering activities. In recent years, there was a significant rise in interest to the intersection of constructionism and AI (Kahn & Winters, 2021). However, the focus is on using constructionist methodology to teach *about* AI (e.g. Lane, 2021; Druga et al., 2018; Kahn et al., 2018; Ali et al., 2019), or on using AI techniques to analyze constructionist play (Berland et al., 2014; Tissenbaum, 2020). In contrast with that, in this work, we examine how an intelligent guidance system can directly support constructionist-like learning activities.

2.2. Child-driven designs for early literacy

Montessori classrooms utilize an open-ended, child-driven, guided approach to early literacy. For instance, using a material called the Moveable Alphabet, children arrange words of their choosing with the help of a teacher (P. P. Lillard, 1972). Children also learn literacy through expressive writing in Tools of the Mind curriculum (Bodrova and Leong, 2006). Other educators developed activities stimulating meaningful, intrinsically motivated writing, such as classroom newspapers and invitation letters to adults (Strickland & Morrow, 1989).

Digital systems bring additional affordances to the child-driven learning setup by providing immediate feedback and additional expressive capabilities. Early systems of that type include the Talking Typewriter (Moore, 1966) and Talking Blocks (Falbel, 1985). Modern systems include Word Wizard (Abel, 2020) and SpeechBlocks (Sysoev et al., 2017), PictureBlocks (Makini et al., 2020), StoryBlocks (described by Woolf (2020)) and Arthur's Comic Creator. In studies involving such apps, children were observed to create a variety of expressions related to their interests and their everyday lives, such as spelling names of themselves and people they know, items related to their hobbies, favorite fictional characters, and messages to other people; engaging in rhyming and word play; creating fun nonsense words; and building imaginative scenes and stories (Makini et al., 2020; Sysoev, 2020, section 5.2). High levels of engagement, manifestations of agency and self-efficacy were observed (Sysoev et al., 2017).

2.3. Scaffolding

While child-driven designs possess important advantages, leaving children to their own devices might not be an optimal choice. Reviews of discovery learning research suggest that guided discovery is markedly more efficient than pure discovery (Kirschner et al., 2006; Mayer, 2004). Brennan (2013) argued that external structure in constructionist environments can support, rather than constrain, learners' agency. In the literacy domain, there is a need for the systematic and explicit presentation of spelling-sound patterns rather than letting children discover them (Castles et al., 2018; Beck & Beck, 2013). While literacy discourse typically centers around instruction, these practices were successfully applied in a child-driven, guided context by Montessori classrooms (Franc & Subotic, 2015). Guidance in response to children's mistakes was shown to be crucial for children's learning from literacy software, particularly for students with poor self-regulation (Kegel et al., 2009). Experience with the child-driven literacy app SpeechBlocks suggests the need for extensive guidance in order for children to express their ideas (Sysoev, 2020, section 5.6).

Building on Vygotsky's theories that learning primarily occurs in the presence of a helpful "more knowledgeable other" (Vygotsky, 1978), Wood, Bruner & Ross proposed the concept of scaffolding (Wood et al., 1976), which comprises (1) assisting the child in the context of a particular project, (2) simplifying the task when necessary, and (3) fading scaffolding as the child's skill improves (Wood & Wood, 1996). The term diluted somewhat over the years, with many scaffolding systems omitting elements such as fading or adapting to the learner's knowledge (Belland et al., 2017).

The concept of scaffolding is widely applied in computer-supported learning - for instance, in intelligent tutoring systems (e.g. Nye et al., 2014; Anderson & Gluck, 2001), including ones for literacy (Kegel & Bus, 2012; Jacovina & McNamara, 2017; Gordon & Breazeal, 2015; Reeder, Shapiro, Wakefield, & D'Silva, 2015) Computer-scaffolded systems were shown to be comparable in effectiveness to one-on-one human tutoring in a variety of fields (Belland et al., 2017; Ma et al., 2014).

In recent years, there has been a rise of interest in scaffolding of open-ended and exploratory learning activities (e.g., AlMamun et al., 2020; Kim & Lim, 2019), including scaffolding via intelligent systems (Winkler et al., 2021; Munshi et al., 2018). However, in these studies, learners still work within a context of a particular assignment or a task, rather than creating projects on their own initiative. On the other hand, in constructionist research, computer-based scaffolding is typically limited to user interface elements that help the learner structure the process or reflect on it (e.g., Chapman, 2006; Tseng, 2015). Kahn and Winters (2021) state that development of AI systems which can guide students through their own projects is still an area of future research. The present paper discusses an approach to exactly this problem.

2.4. Early literacy fundamentals

Reading and writing involve a variety of skills, but they can be arguably divided into two relatively independent groups: those pertaining to encoding/decoding (translation between written and oral forms of language) and those pertaining to linguistic comprehension (Hoover & Gough, 1990). In the early stages of literacy acquisition, encoding/decoding skills are particularly important. Research shows that simply memorizing the spelling of individual words is a very inefficient strategy for literacy learning (Castles et al., 2018), and subword patterns linking pronunciation and spelling must be acquired. This requires an understanding of the sound structure of the words - a skill called *phonological awareness*. It takes practice because sounds of speech are coarticulated (blended) together and not easily separable. Complementary to this is *graphophonemic knowledge* - internalizing the mapping between letters/letter combinations (called *graphemes*) and phonemes. While some languages have nearly one-to-one correspondence between letters and phonemes, it is highly context-dependent in English, making it nontrivial to acquire (Seymour et al., 2003). A good summary of the research on learning to read words is provided by Ehri (2005). SpeechBlocks II focuses on acquisition of phonological awareness and graphophonemic knowledge.

2.5. Executive functioning

Self-regulation skills greatly affect children's learning (see A. S. Lillard, 2017; ch. 4), including their interactions with learning software (Kegel & Bus, 2012). Abundance of choice may strain self-regulation (Patall et al., 2008). These factors make self-regulation a relevant variable to consider in this study. We viewed it through the construct of executive functioning (EF (Friedman & Miyake, 2017);), which is often used in educational research (e.g. (Blair & Razza, 2007)). Executive functioning is the ability to direct mental resources towards a certain goal (Friedman & Miyake, 2017), and is typically viewed as a combination of impulse inhibition, memory updating and task switching (ibid).

3. Design

3.1. SpeechBlocks II: an early literacy app

Our experiments with built-in scaffolding were performed using an Android literacy app called SpeechBlocks II. SpeechBlocks II is designed for tablets. The design of the app is based on SpeechBlocks (Sysoev et al., 2017) and PictureBlocks (Makini et al., 2020). Similar to SpeechBlocks, children can spell any words (real or nonsense) within the app, hear them pronounced and save them. Similar to PictureBlocks, children obtain small icons (sprites) when they spell imageable words and compose pictures out of them (Fig. 1). Previous studies demonstrate that these two expressive activities are engaging for children. The app has two main screens: the word building screen and the picture canvas for arranging the compositions. We used a set of 1711 imageable words during the study with images derived from the FlatIcon website.

The word building process was modified relative to the original SpeechBlocks. While the original design allowed for easy remixing of words, we perceived that it had limitations for constructing words block-by-block, particularly in the scaffolded mode: it was hard to see the contents of both the canvas and the keyboard at once; it was easy to clutter the canvas with pieces of words, which led to accidental snapping of blocks to wrong strings; which word the child was working on was less clear for the scaffolding procedure; and it was difficult for the scaffolding mechanisms to rearrange the blocks automatically. For these reasons, we implemented a different word building interface: *word box* (Fig. 1). It is a container into which the blocks could be dropped, after which they slide to the left to form words. The blocks can be rearranged by dragging them around.

The app utilizes onomatopoeic mnemonics to have a consistent visual representation of phonemes regardless of which graphemes code them. For instance, *[s]* is represented by a hissing snake, and *[k]* is represented by the sound of a karate kick (Fig. 2). The mnemonic characters are animated and can take the shape of various graphemes (e.g., the S in SNAKE and the C in CITY). Their design and evaluation are detailed in a separate publication (Sysoev et al., 2021). The app could operate using both letters and phonemes as building blocks; it is shown in phoneme mode on Fig. 1.

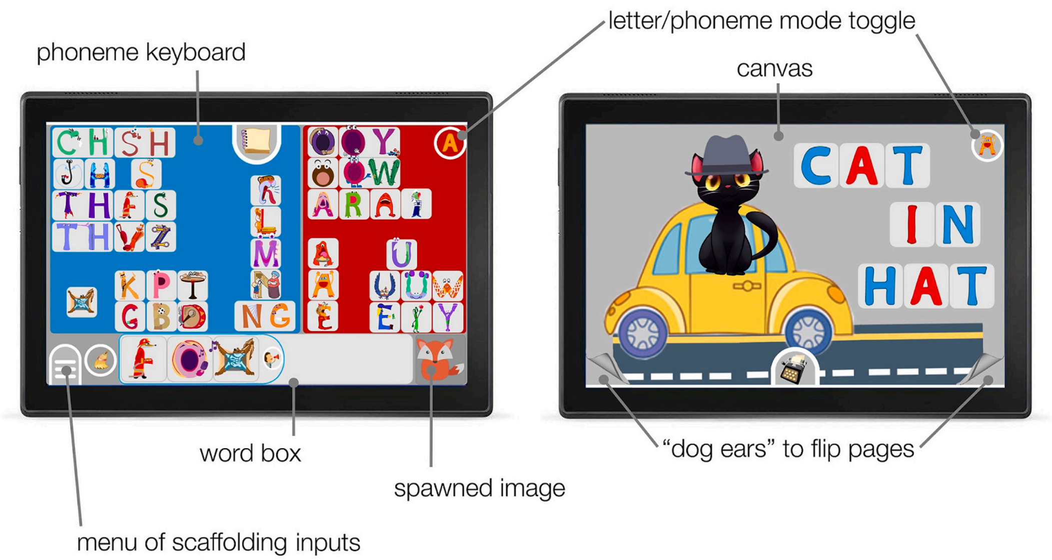

Fig. 1. SpeechBlocks II.

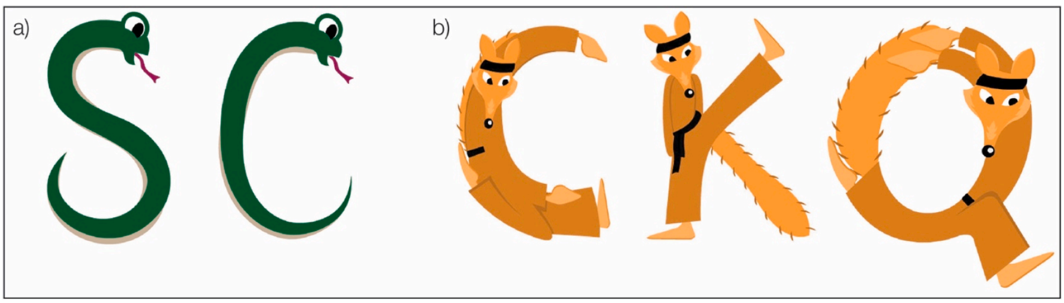

Fig. 2. Mnemonic characters for [s] (a) and [k] (b) assuming the shapes of letters S, C, K and Q.

The design described here resulted from multiple (about 20) iterations of prototyping and playtesting. It was conducted with 4- to 7-year-old children at several locations, including a university-affiliated preschool, an afterschool program, and a children's museum. Playtesting was approved by IRBs of MIT and the museum.

3.2. Scaffolding architecture and inputs

Our scaffolding system consists of two components that are nearly independent from each other. The first component is a set of input mechanisms allowing the system to determine which word the child would like to make. The second component is a procedure that guides the child through the process of constructing the word, providing help when necessary. Before proceeding to this procedure, we will briefly describe the input mechanisms. An interested reader may find more details in (Sysoev, 2020, section 3). Since our players were expected to be preliterate, words to choose from were represented as icons in all of these inputs.

Word bank. Its intent was to provide simple and reliable, though limited, means of input. It was a simple catalog of words arranged by categories. We selected its content based on the analysis of popular words in SpeechBlocks. We also chose words that could complement each other in scenes.

Speech recognition. The intent of this mechanism was to provide the greatest flexibility. We utilized the Google Cloud Speech API. We anticipated that (1) performance of ASR on children's voices would be limited (see Yeung & Alwan, 2018), and (2) that children may form their requests in phrases (e.g., "please give me a fox"). To account for these factors, we displayed multiple candidate results of recognition, and allowed the child to pick between them.

OCR. Children often are interested in environmental words (Strickland & Morrow, 1989). Previously, we observed that children frequently chose to copy them down into SpeechBlocks (Sysoev, 2020, section 5.2.2). We surmised that automatically guided spelling may be more educationally valuable (due to the emphasis on phonology), as well as easier for the child, than copying letter-by-letter. To this end, we implemented an OCR input, utilizing Google on-device OCR API.

Semantic associations. Browsing through words that are semantically related to previously built ones can be useful in constructing scenes. To support this, we utilized the association network from PictureBlocks (Makini et al., 2020), which children could enter through any sprite they built and traverse to any depth. Picking a word from the network would then invoke the scaffolding.

Invented spelling interpretation. Invented spelling (e.g., FES for "fish" and KT for "cat") is exhibited by some children and reflects their phonological knowledge (Read, 1971), and guiding it is beneficial for encoding/decoding development (Richgels, 2001; Richgels, 2001). We interpreted children's spellings using a modified version of the Wagner-Fischer algorithm (Vintsyuk, 1968; Wagner & Fischer, 1974) and allowed the child to pick from several guesses of the interpreter. If the input and the target word were sufficiently similar, the input morphed into the target word; otherwise, the guidance procedure was invoked.

3.3. Guidance procedure

Early on, we decided that the guidance procedure (Fig. 3) would direct the child phoneme-by-phoneme instead of letter-by-letter. For instance, for the word *phone*, the guidance procedure would mention that the first sound is *[f]* rather than that its first letter is P. The intent was to emphasize the phonological structure of the word. The aforementioned mnemonic characters were intended to help children locate the corresponding blocks, as well as to highlight the phoneme-graphene relationships. In practice, they appeared to be useful for some children, but not for all (Sysoev et al., 2021). The alignment between phonemes and graphemes was computed using an EM-like algorithm and further refined manually (Sysoev, 2020, section 4.1).

To simplify word construction for the child, the guidance procedure constrains the keyboard and the word box. The constrained keyboard holds blocks corresponding to the phoneme-graphene pairs in the target word plus a few distractors. The word box has

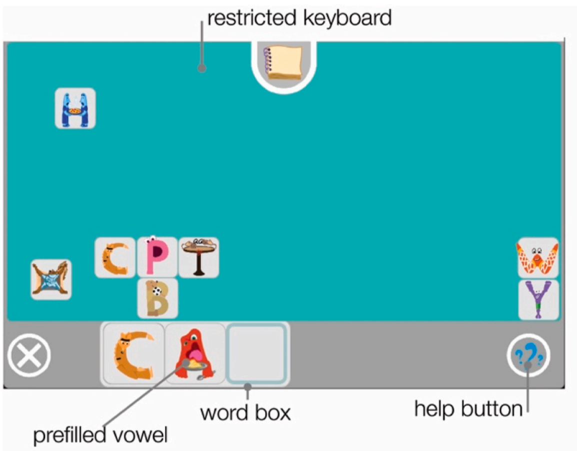

Fig. 3. Word building in guided mode.

certain slots prefilled. Varying the number of prefilled slots and the number of blocks on the keyboard in principle allows for adjustment of difficulty.

At each step in word building, the system sounds out the completed prefix of the word. It then says "next goes" and tells the child the next phoneme. Pressing the *help* button causes the system to indicate which block corresponds to the current phoneme by referring to the corresponding mnemonics (e.g., "*[s]*, like Sally the snake hissing"). We avoided directly highlighting the needed block out of the concern that children would build words mechanically by picking highlighted blocks without paying attention to phonemes.

The scaffolding system responds to erroneous actions by the child. When an incorrect block is dragged into the box, the system discards it and repeats the prompt. If the next block is also incorrect, the system additionally invokes the onomatopoeic cue. Further mistakes lead to the same response. During playtesting, we experimented with making corrections only after the word is fully assembled, to give children a chance to notice mistakes themselves. We eventually opted for immediate correction, since none of playtesting participants corrected mistakes on their own, and delayed correction was confusing and frustrating for them.

Playtesting also revealed that children occasionally (a) place the correct block in a random position within the assembled prefix or (b) place an out-of-order block (such as the last one) at its proper position within the word. In both situations, knowledge is demonstrated by the child and needs to be acknowledged. Our system accepts these actions, slightly modifying their outcome when necessary (e.g., gliding a block to the correct position).

A key aspect of the scaffolding concept is adjusting the level of difficulty to the child's current ability. Our design had provisions for it, but, aside from escalating the feedback level in response to mistakes, we did not implement such automatic adjustment. We did it in order to avoid additional design complexity that might complicate the interpretation of the outcomes. Instead, we used a fixed difficulty level: the keyboard contained 10 blocks, and the vowel slots in the words were prefilled. The latter was motivated by the observation, from the literature on spelling development (e.g., Richgels, 2001) and our own playtesting experience, that spelling vowels tends to be more difficult.

4. Method, study setup and participants

We evaluated our design in a classroom-based study, which was exploratory in its nature and conducted within the design-based research paradigm (Barab & Squire, 2004; Collins et al., 2004). We focused on qualitative observations to examine and, if necessary, modify the conjectures described in the introduction. To examine these conjectures, we focused primarily on qualitative observations in the classroom. We supplemented these observations with some exploratory quantitative analysis, presented in sections 5.2 and 5.3. To obtain a sense of the app's effect on learning, we conducted pre- and post-assessments of children's phonological awareness (PA) and compared it with a control group where SpeechBlocks weren't introduced into the classrooms. The study was approved by IRBs of MIT and Northeastern University.

The study was conducted in five kindergarten classrooms at a public charter school in Greater Boston area. The school served a predominantly low-SES population that had a high fraction of children of color. We received written consent from the families of the participating children. The participants were between 4 and 5 years old. Three classrooms with 32 participants (14 boys and 18 girls) were placed in the control condition and continued to operate as usual. Two classrooms with 25 participants (14 boys and 11 girls) were in the treatment condition and received the app. The split into the treatment and control groups at the classroom level was done to facilitate in-person observations of play. However, it introduced classroom effects in quantitative analysis. Because of the exploratory nature of the study, we found this tradeoff acceptable. One treatment participant was removed from quantitative analysis due to inability to obtain pre-PA score (we couldn't orient him to focus on the test).

The classrooms had a portion of the daily schedule allocated for small-group activities during which the children rotated in groups of four or five between learning stations. Some of the stations' activities were literacy-oriented, such as building words with magnet letters, letter tracing and matching letter cutouts with shapes on a page. During group activities, children also practiced letter names and sounds, as well as syllable counts. Under the treatment condition, we introduced SpeechBlocks II at one of the stations. Children spent between 10 and 15 min with the app (depending on the schedule). Each participant played with the app approximately twice a week. The study lasted for 11 wk. The teachers did not participate in the app-related activities. One or two researchers were present during each session to provide technical assistance when needed. Occasionally, observers (see below) also performed this role under our supervision. Assistance included responding to questions about UI, proactively assisting children who appeared to forget its usage or experienced issues, rebooting the app when bugs occurred, helping children handle the headphones. We also introduced UI to the players. To simplify learning of rather complex interface, features were activated and introduced over the course of several weeks (see Table 1 for the sequence). When a new feature was introduced, we started a session by modeling its use in front of the group, asking

Table 1
Feature introduction sequence.

week	feature
1	mnemonic characters, keyboards, and scaffolding interface
3	word bank
4	invented spelling interpreter
5	associations network
7	OCR
8	speech recognition

their input (e.g. which words to spell) when appropriate, before handing them the tablets for individual play. Additionally, the facilitators responded to children when they initiated conversations with them and checked disruptive behaviors.

Several types of data were collected during the study. First, we gathered observations of children's behavior. We avoided video recording to preserve the privacy of nonconsenting children in the classrooms and collected these data via observer notes. Observers were the researchers and six speech-language pathology students from Northeastern University who volunteered to assist with the study as part of their clinical practice. Two at a time were present. To capture unexpected phenomena, we decided that the notes should be open-ended. However, we instructed the observers to prioritize several types of observations related to our conjectures: verbalizations of intentions (e.g. "I'm going to build BATMAN!") or literacy-related concepts (e.g. children saying letter names or sounds), requested or received literacy-related help, confusion regarding literacy concepts (e.g. confusing letter names with their sounds) or the interface (e.g. forgetting how to start building a word), indications of (dis)engagement (e.g. laughs, smiles, yawns, looking away from tablet, bored swiping on screen) and senses of agency and self-efficacy (e.g. exclamations "Look what I made!"), and social interactions of children (e.g. sharing of ideas, requesting help from peers). To train the observers, we provided examples of behaviors of interest, then simulated their work environment by presenting a video of children playing with early SpeechBlocks and asking them to make observations, and finally jointly discussed the observations that each of them made.

The app also wrote logs of play interactions in a level of detail sufficient to reconstruct every on-device aspect of the play sessions. The logs allowed us to reconstruct a number of variables (such as mistake rates) for exploratory analysis (such as in the end of section 5.2). They also allowed to recreate what was built by children and to clarify observation entries in case of unclear details. Finally, we administered two assessments before and after the study. We measured children's phonological awareness (PA) using the corresponding component of the CTOPP-2 test (R. K. Wagner et al., 2013). We utilized developmental scores for the PA component in our analysis, since it evaluates the "raw" level of a child's skill rather than their age-adjusted skill. To measure children's executive functioning, we resorted to the Hearts & Flowers test (Wright & Diamond, 2014), which was used in some prominent studies of this ability (e.g., Diamond et al., 2007) and taxes all three core executive functioning skills. We used the total number of correct responses on all three parts of the test (congruent, incongruent and mixed stimuli) as the aggregate score. The test was administered on the tablets. CTOPP-2 and Hearts & Flowers were administered by the same volunteers who conducted observations. Therefore, the testers knew which classrooms received SpeechBlocks, and thus were not masked to condition.

The data analysis was performed by the first author as part of his dissertation research. The qualitative data were analyzed inductively, by looking for themes that emerged across all sessions and children, in the records of their interactions with the app, each other, the environment and the teachers/facilitators. The quantitative (regression) analysis was performed using the *lm* function in *R* (R Core Team, 2019) and the *brms* package for Bayesian mixed-effect modeling (Bürkner, 2017).

5. Results

5.1. Engagement

Similar to earlier studies with expressive literacy media (Makini et al., 2020; Sysoev et al., 2017), multiple children exhibited notable engagement with the app. They made excited exclamations when it was their turn to play with the app and were impatient to receive the tablets and start playing. They were at times disappointed that their turn ended before they could implement their ideas, making exclamations such as "Oh, come on!" at the signal to switch stations. Teachers reported that children were disappointed when it was not their turn to play with the app. On one occasion, due to a classroom scheduling issue, a session lasted for 30 min. At the end of the session, 3 out of 4 children maintained focused engagement, and one of them was still disappointed when the play time ended. Children asked if they could take the devices home to keep playing.

One aspect of engagement noted in previous works but lacking in the current study was the fun of making nonsense words (e.g., CUPEAR and ZOOBALLBALL) through remixing. This was a consequence of the redesigned word-building interface.

5.2. Types of play

Analysis of play observations across all children and sessions revealed three main types of play with SpeechBlocks II: *word crafting*, *imaginative play* and *impulsive exploration*. For each of them, there were children who interacted with the app primarily in this way. Though it could also be said that there were three main types of players, at least 10 of the 25 children exhibited various mixtures of the three play types, making player types less distinct than play types.

Word crafting focused on the creation of words apparently for the sake of it. Similar to earlier studies (Sysoev et al., 2017; Sysoev, 2020, section 5.2.2), a very popular category of such words were names. Some players experimented with long words, such as TRANSPORTATION. Word crafters enjoyed collecting the words they created on the canvas.

Imaginative play had two primary forms: making static scenes (such as those in Fig. 4) and enactment, in which children used sprites akin to physical toys. For example, a child built several wild animals, then put a crocodile over them and enact the crocodile devouring the other animals by moving it back and forth while saying, "Chomp! Chomp!". These forms of play were often combined by building a scene and then enacting some action within it. Children explored a diverse range of themes, such as family, home life, city, jungle and space. Among their sources of inspiration were the topics they studied in the classroom and the works of their peers.

Let us examine one example of imaginative play (Fig. 5). At first, the child built two ninjas and said, "They are father and son. They are practicing". She then expressed a desire to give them weapons and used invented spelling recognition to create SOD (*sword*). Then she resorted to speech recognition to build SHIELD. Afterwards, she tapped on the sword to see the related words, picked DAGGER and

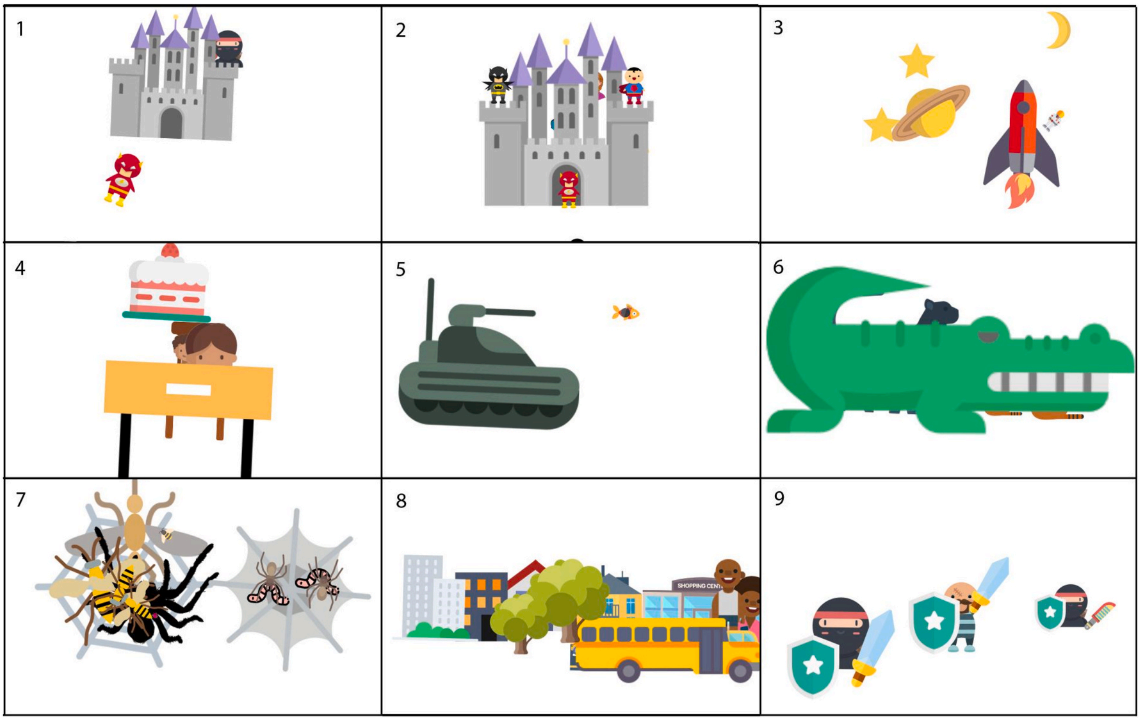

Fig. 4. Scenes built by one of the participants.

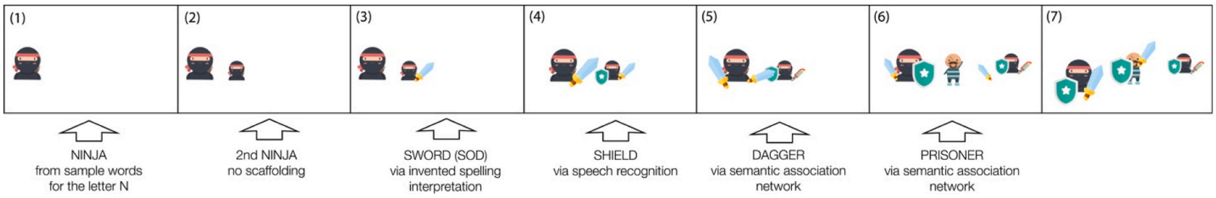

Fig. 5. A timeline of one scene construction.

gave it to the small ninja. This was followed by a long exploration of the semantic association network, until she stumbled upon the word PRISONER. This discovery prompted her to exclaim, "I'm going to make a villain to fight them!", which led to the complete scene.

The child's utterances indicate that she had a narrative in mind that she was eager to flesh out and share with others. The narrative evolved throughout the session, shifting from a training scene to a fight with a villain. In implementing it, she was supported by various scaffolding mechanisms.

Scaffolding appeared to be a vital enabler of imaginative play, which required active word building: up to 15 real words in some scenes. Children at this age generally struggle with building words independently, and it is difficult for adults to scaffold multiple children at such a rate. In a comparable study without built-in scaffolding (Sysoev et al., 2017), there were just 10 real words made by all children throughout the study, aside from their names and copies of words from the facilitating materials. By the end of our study, active imaginative players learned to use scaffolding almost completely autonomously. The teachers noted this autonomy and appreciated the resulting reduction in their workload.

Despite the presence of scaffolding, active word building for imaginative play is still demanding, requiring long sequences of coordinated actions and focused attention to phonetic prompts. We identified six children as particularly avid imaginative players. Each of them achieved a nearly maximum score on EF test, and none of them scored below the median on the PA composite.

Impulsive exploration was characterized by a lack of systematicity: while long-term plans were voiced by the players, they were rarely followed through. Instead, children focused on short-term rewards - emotional, social and cognitive - that could be gained through interaction with the system. Such play was often unproductive from the literacy standpoint.

Below is an example of such play. A child's attention was drawn by a peer saying that he is going to make ten copies of BATMAN. "I wanna make ten Batmans [too]!" exclaimed the child. He was instantly distracted from that goal by something else but later returned to it. However, he did not pay any attention to the scaffolding prompts and just randomly dragged blocks into slots. The scaffolding system rejected his choices. The child attributed it to a bug in the app and gave up. When his peer asked him about building BATMAN,

Table 2
Descriptive statistics for variables in Table 3.

var (z-score)	min	Q1	Median	Q4	max	skew	kurtosis
pre-PA	−1.38	−0.72	−0.28	0.48	2.82	1.04	0.80
pre-EF	−1.92	−0.99	0.62	0.72	0.92	−0.82	−1.07
scale entropy	−1.74	−0.74	0.01	0.63	1.96	0.19	−0.82
touch per sess.	−1.72	−0.51	−0.09	0.42	2.60	0.66	0.60
touch speed	−2.10	−0.77	0.04	0.60	1.82	−0.09	−0.58
log touch acc.	−1.83	−0.65	0.04	0.45	2.09	0.31	−0.25

Table 3
Measures reflecting impulsive interactions with the app, in association with PA and EF.

		estimate	low-95%	hi-95%
behavior: misuse of OCR	PA	−0.30	−0.89	0.27
outcome var.: calling OCR but not starting spelling	EF	−0.21	−0.77	0.37
model: logistic regression				
with random intercepts for children, **n** = 453				
behavior: sprite scaling	PA	−0.10	−0.53	0.32
outcome variable: z-score of sprite scales entropy	**EF**	**−0.47**	**−0.90**	**−0.06**
model: OLS, **n** = 24				
behavior: quick taps	**PA**	**−0.48**	**−0.89**	**−0.07**
outcome variable: z-score of mean touches per session	EF	−0.20	−0.60	0.20
model: OLS, **n** = 24				
behavior: quick swipes	PA	−0.06	−0.51	0.39
outcome variable: z-score of mean touch speed	EF	−0.36	−0.81	0.09
model: OLS, **n** = 24				
behavior: "jerky" swipes	**PA**	**−0.42**	**−0.84**	**0**
outcome variable: z-score of (log) mean touch acceleration	EF	−0.20	−0.62	0.21
model: OLS, **n** = 24				

he said, "Batman is not working". The peer responded, "You just gotta spell it! Can you hear it?" The peer then started to demonstrate how to use the scaffolding to build the word. The impulsive explorer, however, did not listen to him and was looking away. Therefore, the friend recruited his attention by asking him to pick the next block. Together, they eventually managed to complete the word.

Some attributes of this example were common among impulsive explorers. First, there was a passionately expressed desire to make something ambitious. Second, these children quickly moved away from these plans or were easily discouraged by the challenges. Third, in contrast with word crafters, impulsive explorers were interested in the outcome, but not the process, of word building - to the extent that they sometimes wanted their peers or the researchers to build words for them. Fourth, there was a lack of attention to directions - be it our demos, the feedback of the scaffolding system, or advice given by a peer or a facilitator - even when these directions were aimed at helping the children achieve their own goals. Fifth, there were attempts to "brute-force" word constructions by trying random actions.

Impulsive explorers also showed a tendency to use the app interface in unintended ways. They randomly tapped and swiped on the keyboard screen (as if playing a piano or tickling it), tinkered with the hardware buttons (e.g., power off), tried to intentionally cause bugs, spent entire sessions making a single sprite big and small (e.g., "Look, I made a gigantic egg! Now I'm going to make it teeny-tiny! Now it's huge-huge-huge-big!") and used the OCR interface to "take pictures" of each other.

Qualitatively, we observed that impulsive exploration occurred primarily with children with low EF and PA scores. To explore whether this can be observed quantitatively, we used logs to derive several measures corresponding to observed features of impulsive interactions. The descriptive statistics for the involved measures are given in Table 2. We then tried to predict these measures using z-scores of initial PA and EF. To capture misuse of OCR, we used mixed effects logistic regression to predict whether a given OCR call would result in no word building started. Random intercepts for children were used to account for the dependencies between observations. There were 453 OCR calls in total. This model was fitted using *brms*. Models for other measures used ordinary least squares estimations. To capture the "scaling" behavior, we calculated the entropy[1] of sprite sizes while being manipulated by children (up-down-scaling corresponds to higher entropy). To capture quick tapping, we measured the mean number of touches per session. In our observations of chaotic swiping, we noted (1) quick and imprecise finger motions, and (2) rapid wiggling of the finger on the screen, which would correspond to high tangential and centripetal acceleration of the touch point. Thus, we measured mean speed and acceleration of finger movement. The distribution of mean accelerations was skewed to the right. To correct this, we applied log to this measure. Table 3 shows the estimates of coefficients associated with PA and EF in the above-described models. One can see that increase in PA and EF is associated with decrease on all measures of impulsive interactions. In some cases (bolded), the association is strong enough to be statistically significant. The association of impulsive interactions with low PA and EF is plausible, since word

[1] Using the 20-sample-spacing estimator (Vasicek, 1976).

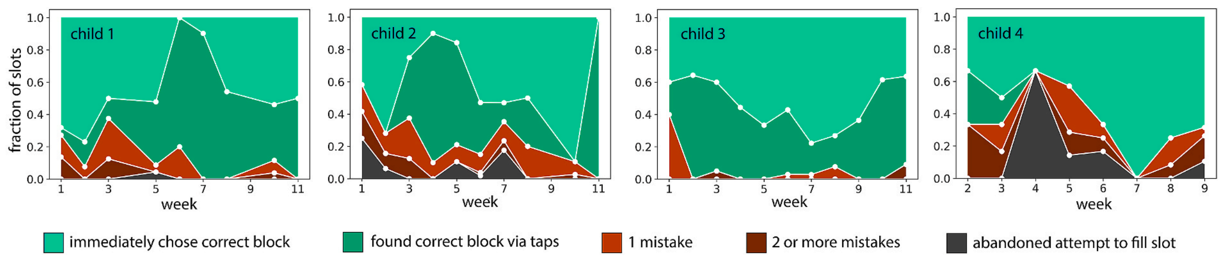

Fig. 6. Slot filling by impulsive explorers, by week.

building demands patience and impulse suppression – and lower PA skills make these demands higher.

Encouragingly, some impulsive explorers developed a much more focused play style towards the end of the study. Fig. 6 shows how slots in the constructed words were filled in different weeks by the four children whom, based on observations, we considered the most prominent impulsive explorers. Early on, many slots were filled after some mistakes or attempted and abandoned. Observationally, it corresponded to random dragging. Towards the end of the study, the frequency of this behavior was reduced, and players either learned to immediately locate the needed blocks or searched for them via taps on the keyboard.

5.3. *Exploratory analysis of learning*

To analyze effect of the app on PA scores, we utilized Bayesian mixed effect models. This allowed us to use random effects to model differences between classrooms. We utilized default priors of *brms* package, which are not expected to significantly bias the results. In all regressions, all continuous variables were standardized (converted to z-scores), which allows to see how the effects compare to the standard deviation of post-scores. In all models, n = 24 treatment +32 control = 56 participants. The descriptive statistics for the involved variables are given in Table 4.

To estimate the overall effect of the app, we fit a model with final PA as the outcome variable, fixed effects of initial PA and treatment, and random intercept for each classroom. Table 5 summarizes the estimated effects and their credible intervals. The estimated effect of treatment is positive, but the difference is not significant.

The distracted behaviors of children with low PA and EF made us suspect that these variables might moderate learning gains. To evaluate this, we fit a model with fixed effects of initial PA, condition and EF, fixed interaction effects of PA/EF and condition, and random intercepts for the classrooms. Table 6 summarizes the results. One can see that there is a significant interaction between PA and condition. There also appears to be an interaction between EF and condition, but it doesn't achieve significance. This suggests that higher initial skills might have led to more gains from the app.

Finally, we wanted to see if this phenomenon can be attributed to classroom differences. To this end, we added random slopes for

Table 4
Descriptive statistics for variables used In learning analysis.

var (z-score)	min	Q1	median	Q4	max	skew	kurtosis
pre-PA	−1.53	−0.77	−0.28	0.54	3.23	1.00	0.84
pre-EF	−1.68	−0.90	0.18	0.94	1.24	−0.28	−1.52
post-PA	−1.52	−0.7	−0.19	0.57	3.19	0.99	0.58

Table 5
Results from the simple model of treatment.

	estimate	low-90%	high-90%	low-95%	high-95%
intercept	−0.05	−0.28	0.19	−0.34	0.25
treatment	0.12	−0.25	0.50	−0.35	0.61
pre-PA	0.82	0.69	0.96	0.66	0.98

Table 6
Results from the model including interaction effects.

	estimate	low-90%	high-90%	low-95%	high-95%
intercept	−0.09	−0.34	0.16	−0.43	0.23
treatment	0.11	−0.29	0.52	−0.42	0.65
pre-PA	0.63	0.44	0.82	0.41	0.86
pre-EF	−0.13	−0.33	0.06	−0.36	0.10
pre-PA:treatment	**0.32**	**0.05**	**0.59**	**0**	**0.64**
pre-EF:treatment	0.27	−0.01	0.55	−0.06	0.61

Table 7
Results from the model including interaction effects and random slopes for classrooms.

	estimate	low-90%	high-90%	low-95%	high-95%
intercept	−0.09	−0.38	0.22	−0.48	0.32
treatment	0.11	−0.39	0.61	−0.60	0.76
pre-PA	0.63	0.38	0.88	0.31	0.95
pre-EF	−0.13	−0.44	0.18	−0.54	0.30
pre-PA:treatment	0.32	−0.08	0.73	−0.20	0.86
pre-EF:treatment	0.27	−0.20	0.76	−0.32	0.96

initial PA and EF for the classrooms. This corresponds to the assumption that different classrooms might favor students with higher or lower initial skills differently, due to variations in teaching practice. After fitting this model, we saw that interactions in question became not significant even at 90% significance level (Table 7). This shows that our data is not sufficient to reliably disentangle effects of the app and the classrooms. Nevertheless, the qualitative observations and the literature suggest that this phenomenon might indeed be present, and is worth investigating in future studies.

5.4. Agency and self-efficacy

Similar to previous work (Sysoev et al., 2017), we noted that children exhibited senses of agency, self-efficacy and authorship, which manifested through their excited comments regarding what they made, frequent and proud displays of their work to both adults and peers, plans regarding what to construct, and desire to keep their creations and bring them home. Some children kicked, stomped, jumped in excitement and exclaimed, "Yay!" after finishing words they intended to build. Some were also proud to gather many objects on one page as evidence of their hard work. E.g., a child said, "Look! This is all my stuff! This is my page! This is my book!", apparently seeing himself as a book writer.

One particularly interesting manifestation of self-efficacy was children voluntarily challenging themselves to spell difficult words. Sometimes, after mastering making a certain word in scaffolded mode, children proceeded to spell the same words without scaffolding. E.g., one child put his headphones down to avoid hearing the prompts. After completing the word, he proudly announced, "I did it all by myself!"

5.5. Social interactions

Similar to early SpeechBlocks (Sysoev et al., 2017), children eagerly engaged in social interactions centered on their play with the app, despite the introduction of headphones. These interactions played three potentially beneficial roles: *inspiring each other's ideas*, *maintaining mutual engagement*, and *directly helping each other*.

Active borrowing of ideas from peers accompanied imaginative play. In one group, which had four avid imaginative players, traces of idea borrowing could be seen in fourteen out of fifteen play sessions. For instance, a girl made a scene with panthers encroaching on tigers and showed it to a friend who reciprocated with a panther attacking a giraffe. The girl then sought to outdo it and added a crocodile to devour both panthers and tigers. When we brought printouts of the scenes made by players into the classroom, children also copied them.

There were other ways in which children supported each other's engagement. First, they acted as an audience to one another. Second, they sometimes engaged in joint play, deciding on the rules together. For instance, one child proposed to a peer, "I'm going to build your name, and you build my name. Then, let's both build Abigail's name".

However, perhaps the most valuable social interaction was peer learning. Children often came to assist their struggling peers. For instance, in the Batman example from section 5.2, the helper showed his friend how to pick the right block and made sure that he learned the principle by asking him to find the next block himself. We also observed several "leader-follower" pairs in which a more skilled child built some words and another one repeated them after her/him. These pairs talked about what they made, how they made it, and what they should do next. If the "follower" experienced any difficulties in the spelling process, the "leader" assisted him/her.

Automatic scaffolding appears to support these interactions. By fostering children's autonomy, it allowed them to fluently respond to emerging ideas. The scaffolding machinery also made it relatively easy to explain to peers what to do by referring to the machine prompts.

6. Discussion

We created a system to facilitate early literacy learning though a creative, child-driven activity in which automatic scaffolding supports learners in achieving their own goals. The key difference between our work and scaffolding systems in other research/commercial systems is that we strived to provide expressive freedom for the player, ideally allowing for open-ended choice of words to spell. Our evaluation suggests that the scaffolded system preserves and further develops the strong sides of earlier construction-based literacy media - such as empowering the learners with the senses of agency and self-efficacy and inciting supportive social interactions - while increasing the autonomy of their interactions with the software. However, the current scaffolding design was insufficient to prevent a large fraction of children (typically with low executive functioning and phonological awareness) from engaging in chaotic interactions with the app, which apparently inhibited their learning from it. This observation is consistent with earlier results on the effects of low self-regulation on learning literacy from software (Kegel & Bus, 2012).

There are several possibilities for why this effect occurred and how to mitigate it. First, the current design incorporated only minimal adaptation to the child's skill level, possibly making spelling tasks too difficult for some children and leading them to entertain themselves in unproductive ways instead. We plan to investigate adaptive scaffolding designs in future work. Second, we only implemented scaffolding for word building but not for engaging learners' attention, modeling, providing emotional support and checking distractions, as Wood et al. (1976) envisioned. It may be possible for a machine to perform some of those functions, utilizing analysis of productive/non-productive behaviors similar to the one described by Tissenbaum (2020). Alternatively, these roles may be best facilitated by people. This suggests hybrid approaches in which machines handle the of phonological awareness instruction routine while parents, teachers / coaches (Hershman et al., 2018) and peers provide other forms of support. Third, it is possible that some children first need to gain experience with non-open-ended tasks (e.g., puzzles) before moving to constructionist play.

The present study has multiple limitations. First, the statistical analysis is exploratory. We sought to reveal interesting patterns and did not calculate statistical corrections to limit the possibility of type I errors. Furthermore, in the analysis of learning gains, the data didn't allow us to reliably disentangle the effects of the app from effects of the classroom. Therefore, separate studies are needed to confirm or reject the patterns observed here. Second, we studied an app developed by us. Furthermore, the analysis was performed by the first author alone. These factors introduce possible researcher bias. Third, the efficacy of the approach has not been determined and requires further investigation. Fourth, only one quick test of EF was administered, making the measurement of this elusive skill less robust. Fifth, due to the limited duration and formative nature of the study, we were not able to observe whether children's positive experience with the app translated to greater interest in other literacy activities - a very important subject for further investigation. Sixth, while we intended to increase autonomy of children's play with the app, the study still included some support by adults. It would be interesting to see how improved versions of this technology would fare in home conditions. Would the scaffolding system be able to counter the rapid decline in play time that was observed with earlier SpeechBlocks (Sysoev, 2020, section 5.4.2)? If so, would increased engagement translate to learning gains? If the answer to those questions is yes, the present approach might provide a vehicle for early literacy learning outside of classrooms.

There is currently a significant amount of interest in the educational potential of AI technologies. Typically, AI is viewed as an instructor who gives children tasks and asks them questions. Instead, this work provides an example of a model in which AI acts as a guide in child-driven, construction-based play. We believe that this approach can be fruitful in multiple areas beyond literacy learning, particularly as related technologies rapidly develop.

Author contribution

Ivan Sysoev: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Software, Writing, James H. Gray: Conceptualization, Methodology, Investigation, Project administration, Resources, Supervision, Susan Fine: Investigation, Project administration, Resources, Supervision, Sneha Priscilla Makini: Conceptualization, Software, Deb Roy: Conceptualization, Methodology, Project administration, Supervision

Acknowledgements

We would like to express our gratitude to (1) Mitchel Resnick, Catherine Snow and Elena Bodrova – for providing valuable feedback that shaped the course of our development and study; (2) Flora Su – for being the voice of SpeechBlocks, as well as providing ideas and support; (3) Nazmus Saquib and Manuj Dhariwal – for assisting with play-testing; (4) Allie Shoff, Christine Schlaug, Kathryn Falvey, Anne Mathew, Mackenzie Russell, and Lauren Brinker – for their vital help in facilitating the sessions and collecting observations; (5) Twitter – for providing financial support for the MIT team; (6) teachers and staff at the museum and the school – for welcoming us and kindly accommodating our various requests.

References

Abel, P. (2020). *Word wizard - talking educational app for iPhone and iPad. L'Escapadou*. Retrieved September 21, 2020, from https://lescapadou.com/LEscapadou_-_Fun_and_Educational_applications_for_iPad_and_IPhone/Word_Wizard_-_Talking_Educational_App_for_iPhone_and_iPad.html.

Al Mamun, M. A., Lawrie, G., & Wright, T. (2020). Instructional design of scaffolded online learning modules for self-directed and inquiry-based learning environments. *Computers & Education, 144*, Article 103695.

Ali, S., Payne, B. H., Williams, R., Park, H. W., & Breazeal, C. (2019). Constructionism, ethics, and creativity: Developing primary and middle school artificial intelligence education. In *International workshop on education in artificial intelligence K-12 (EDUAI'19)* (pp. 1–4).

Anand, P. G., & Ross, S. M. (1987). Using computer-assisted instruction to personalize arithmetic materials for elementary school children. *Journal of Educational Psychology, 79*(1), 72.

Anderson, J. R., & Gluck, K. (2001). What role do cognitive architectures play in intelligent tutoring systems. *Cognition and Instruction: Twenty-Five Years of Progress*, 227–262.

Barab, S., & Squire, K. (2004). Design-based research: Putting a stake in the ground. *The Journal of the Learning Sciences, 13*(1), 1–14.

Beck, I. L., & Beck, M. E. (2013). *Making sense of phonics: The hows and whys*. Guilford Publications.

Belland, B. R., Walker, A. E., Kim, N. J., & Lefler, M. (2017). Synthesizing results from empirical research on computer-based scaffolding in STEM education: A meta-analysis. *Review of Educational Research, 87*(2), 309–344.

Blair, C., & Razza, R. P. (2007). Relating effortful control, executive function, and false belief understanding to emerging math and literacy ability in kindergarten. *Child Development, 78*(2), 647–663.

Bodrova, E., & Leong, D. J. (2006). *Tools of the mind: Vygotskian approach to early education*. Pearson Australia Pty Limited.

Brennan, K. A. (2013). *Best of both worlds: Issues of structure and agency in computational creation, in and out of school*. Doctoral Dissertation. Massachusetts Institute of Technology.

Bürkner, P. C. (2017). brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software, 80*(1), 1–28.

Castles, A., Rastle, K., & Nation, K. (2018). Ending the reading wars: Reading acquisition from novice to expert. *Psychological Science in the Public Interest: A Journal of the American Psychological Society, 19*(1), 5–51.

Chapman, R. N. (2006). *Pearls of wisdom : Technology for intentional reflection and learning in constructionist cooperatives*. Doctoral dissertation, Massachusetts Institute of Technology. https://dspace.mit.edu/handle/1721.1/38300?show=full.

Clouder, C., & Rawson, M. (1998). *Waldorf education*. Rudolph Steiner Press.

Collins, A., Joseph, D., & Bielaczyc, K. (2004). Design research: Theoretical and methodological issues. *The Journal of the Learning Sciences, 13*(1), 15–42.

Cordova, D. I., & Lepper, M. R. (1996). Intrinsic motivation and the process of learning: Beneficial effects of contextualization, personalization, and choice. *Journal of Educational Psychology, 88*(4), 715–730.

Diamond, A., Barnett, W. S., Thomas, J., & Munro, S. (2007). Preschool program improves cognitive control. *Science, 318*(5855), 1387–1388.

Druga, Stefania (2018). *Growing up with AI: Cognimates: From coding to teaching machines*. Doctoral dissertation, Massachusetts Institute of Technology.

Edwards, C. P., Gandini, L., & Forman, G. E. (1998). *The hundred languages of children: The Reggio Emilia approach–advanced reflections*. Greenwood Publishing Group.

Ehri, L. C. (2005). Learning to read words: Theory, findings, and issues. *Scientific Studies of Reading: The Official Journal of the Society for the Scientific Study of Reading, 9* (2), 167–188.

Estes, T. H., & Vaughan, J. L. (1973). Reading interest and comprehension: Implications. *The Reading Teacher, 27*(2), 149–153.

Falbel, A. (1985). *A second look at writing to read: A teaching system for schools becomes a medium for learning in the home* (Unpublished manuscript).

Findley, M. J., & Cooper, H. M. (1983). Locus of control and academic achievement: A literature review. *Journal of Personality and Social Psychology, 44*(2), 419–427.

Franc, B., & Subotic, V. (2015). Differences in phonological awareness of five-year-olds from Montessori and regular program preschool institutions. In *Researching paradigms of childhood and education conference book of selected papers (2nd symposium: Child language and culture)* (pp. 12–20).

Friedman, N. P., & Miyake, A. (2017). Unity and diversity of executive functions: Individual differences as a window on cognitive structure. Cortex. *A Journal Devoted to the Study of the Nervous System and Behavior, 86*, 186–204.

Gordon, G., & Breazeal, C. (2015). Bayesian active learning-based robot tutor for children's word-reading skills. *Proceedings of the AAAI Conference on Artificial Intelligence, 29*(1).

Hershman, A., Nazare, J., Sysoev, I., Buitrago, J., Soltangheis, M., Makini, S. P., & Roy, D. (2018). Family Learning Coach: Engaging Families in Children's Early Literacy Learning with Computer-Supported Tools. *Proceedings of the 26th International Conference on Computers in Education*, 637–646.

Hoover, W. A., & Gough, P. B. (1990). The simple view of reading. *Reading and Writing, 2*(2), 127–160.

Iyengar, S. S., & Lepper, M. R. (1999). Rethinking the value of choice: A cultural perspective on intrinsic motivation. *Journal of Personality and Social Psychology, 76*(3), 349–366.

Jacovina, M. E., & McNamara, D. S. (2017). *Intelligent tutoring systems for literacy: Existing technologies and continuing challenges*. Grantee Submission https://eric.ed.gov/?id=ED577131.

Kahn, K., Megasari, R., Piantari, E., & Junaeti, E. (2018). AI programming by children using Snap! block programming in a developing country. In *ECTEL practitioner proceedings 2018: 13th European conference on technology enhanced learning*. Leeds, UK, September 3-6, 2018.

Kahn, K., & Winters, N. (2021). Constructionism and AI: A history and possible futures. *British Journal of Educational Technology*.

Kegel, C. A. T., & Bus, A. G. (2012). Online tutoring as a pivotal quality of web-based early literacy programs. *Journal of Educational Psychology, 104*(1), 182.

Kegel, C. A. T., van der Kooy-Hofland, V. A. C., & Bus, A. G. (2009). Improving early phoneme skills with a computer program: Differential effects of regulatory skills. *Learning and Individual Differences, 19*(4), 549–554.

Kelly, A., & Safford, K. (2009). Does teaching complex sentences have to be complicated? Lessons from children's online writing. *Literacy Discussion, 43*(3), 118–122.

Kim, J. Y., & Lim, K. Y. (2019). Promoting learning in online, ill-structured problem solving: The effects of scaffolding type and metacognition level. *Computers & Education, 138*, 116–129.

Kirschner, P., Sweller, J., & Clark, R. E. (2006). Why unguided learning does not work: An analysis of the failure of discovery learning, problem-based learning, experiential learning and inquiry-based learning. *Educational Psychologist, 41*(2), 75–86.

Lane, D. (2021). *Machine learning for kids: An interactive introduction to artificial intelligence*. No Starch Press.

Lillard, P. P. (1972). *Montessori: A modern approach*. Schocken Books.

Lillard, A. S. (2017). *Montessori: The science behind the genius* (3rd ed.). Oxford University Press.

Ma, W., Adesope, O. O., Nesbit, J. C., & Liu, Q. (2014). Intelligent tutoring systems and learning outcomes: A meta-analysis. *Journal of Educational Psychology, 106*(4), 901–918.

Makini, S. P., Oguntola, I., & Roy, D. (2020). Spelling their pictures: the role of visual scaffolds in an authoring app for young children's literacy and creativity. *Proceedings of the Interaction Design and Children Conference*, 372–384. https://doi.org/10.1145/3392063.3394392

Mayer, R. E. (2004). Should there Be a three-strikes rule against pure discovery learning? *American Psychologist, 59*(1), 14–19.

Moore, O. K. (1966). Autotelic responsive environments and exceptional children. In O. J. Harvey (Ed.), *Experience structure & adaptability* (pp. 169–216). Springer Berlin Heidelberg.

Munshi, A., Rajendran, R., Ocumpaugh, J., Biswas, G., Baker, R. S., & Paquette, L. (2018). Modeling learners' cognitive and affective states to scaffold SRL in open-ended learning environments. *Proceedings of the 26th Conference on User Modeling, Adaptation and Personalization*, 131–138.

Nye, B. D., Graesser, A. C., & Hu, X. (2014). AutoTutor and family: A review of 17 Years of natural language tutoring. *International Journal of Artificial Intelligence in Education, 24*(4), 427–469.

Papert, S. (1980). *Mindstorms: Children, computers, and powerful ideas*. Basic Books. https://dl.acm.org/citation.cfm?id=1095592.

Patall, E. A., Cooper, H., & Robinson, J. C. (2008). The effects of choice on intrinsic motivation and related outcomes: A meta-analysis of research findings. *Psychological Bulletin, 134*(2), 270–300.

R Core Team. (2019). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. URL: https://www.R-project.org/.

Read, C. (1971). Pre-school children's knowledge of English phonology. *Harvard Educational Review, 41*(1), 1–34.

Reeder, K., Shapiro, J., Wakefield, J., & D'Silva, R. (2015). Speech recognition software contributes to reading development for young learners of English. *International Journal of Computer-Assisted Language Learning and Teaching, 5*(3), 60–74.

Renninger, K. A. (1992). Individual interest and development: Implications for theory and practice. *The Role of Interest in Learning and Development, 26*(3–4), 361–395.

Resnick, M. (2014). Give P'sa chance: Projects, peers, passion, play. In *Constructionism and creativity: Proceedings of the third international constructionism conference* (pp. 13–20). Vienna: Austrian Computer Society.

Resnick, M. (2017). *Lifelong kindergarten: Cultivating creativity through projects, passion, peers, and play*. MIT Press.

Richgels, D. J. (2001). Invented spelling, phonemic awareness, and reading and writing instruction. *Handbook of Early Literacy Research, 1*, 142–155.

Sandoval, W. (2014). Conjecture mapping: An approach to systematic educational design research. *The Journal of the Learning Sciences, 23*(1), 18–36.

Seymour, P. H. K., Aro, M., & Erskine, J. M. (2003). Foundation literacy acquisition in European orthographies. *British Journal of Psychology, 94*(Pt 2), 143–174.

Stipek, D. (1980). A causal analysis of the relationship between locus of control and academic achievement in first grade. *Contemporary Educational Psychology, 5*(1), 90–99.

Strickland, D. S., & Morrow, L. M. (1989). *Emerging literacy: Young children learn to read and write*. International Reading Association.

Swann, W. B., & Pittman, T. S. (1977). Initiating play activity of children: The moderating influence of verbal cues on intrinsic motivation. *Child Development, 48*(3), 1128–1132.

Sysoev, I. (2020). *Digital expressive media for supporting early literacy through child-driven, scaffolded play*. Doctoral dissertation. Massachusetts Institute of Technology.

Sysoev, I., Gray, J. H., Fine, S., & Roy, D. (2021). Designing building blocks for open-ended early literacy software. *International Journal of Child-Computer Interaction, 28*, Article 100273. https://doi.org/10.1016/j.ijcci.2021.100273

Sysoev, I., Hershman, A., Fine, S., Traweek, C., & Roy, D. (2017). SpeechBlocks: A Constructionist Learning App. *Proceedings of the 2017 Conference on Interaction Design and Children*, 248–257. https://doi.org/10.1145/3078072.3079720

Tissenbaum, M. (2020). I see what you did there! Divergent collaboration and learner transitions from unproductive to productive states in open-ended inquiry. *Computers & Education, 145*, Article 103739.

Tseng, T. (2015). Making make-throughs: Supporting young makers sharing design process. *Proc. of Fablearn.* http://web.mit.edu/ttseng/www/academics/papers/fablearn_tseng.pdf.

Vasicek, O. (1976). A test for normality based on sample entropy. *Journal of the Royal Statistical Society - Series B: Statistical Methodology, 38*(1), 54–59.

Vintsyuk, T. K. (1968). Speech discrimination by dynamic programming. *Cybernetics and Systems Analysis, 4*(1), 52–57.

Vygotsky, L. (1978). Interaction between learning and development. *Readings on the Development of Children, 23*(3), 34–41.

Wagner, R. A., & Fischer, M. J. (1974). The string-to-string correction problem. *Journal of the ACM, 21*(1), 168–173.

Wagner, R. K., Torgesen, J. K., Rashotte, C. A., & Pearson, N. A. (2013). *Comprehensive test of phonological processing (CTOPP-2)*. Pro-Ed.

Winkler, R., Söllner, M., & Leimeister, J. M. (2021). Enhancing problem-solving skills with smart personal assistant technology. *Computers & Education, 165*, Article 104148.

Wood, D., Bruner, J. S., & Ross, G. (1976). The role of tutoring in problem solving. *The Journal of Child Psychology and Psychiatry and Allied Disciplines, 17*(2), 89–100.

Wood, D., & Wood, H. (1996). Vygotsky, tutoring and learning. *Oxford Review of Education, 22*(1), 5–16.

Woolf, Anneli, R (2020). *Discovering the meaning behind the story : Creating a system for documenting and supporting children's narrative development*. Doctoral dissertation, Massachusetts Institute of Technology. http://34.201.211.163/handle/1721.1/127910.

Wright, A., & Diamond, A. (2014). An effect of inhibitory load in children while keeping working memory load constant. *Frontiers in Psychology, 5*.

Yeung, G., & Alwan, A. (2018). On the difficulties of automatic speech recognition for kindergarten-aged children. *Interspeech*, 2018.