Children's engagement in an interactive storytelling tool

Arturas Kulsinskas
Aalborg University Esbjerg
akulsi11@student.aau.dk

Povilas Konopackas
Aalborg University Esbjerg
pkonop11@student.aau.dk

***Abstract*—This paper investigates engagement with the interactive storytelling system. The system is a TUI, which consist of a physical board, which recognises special platforms, and application running on the tablet connected via bluetooth to the board. The tablet is running a software that shows the users a story, and uses the board with the special platforms as input. A qualitative study with repeated measures was conducted at Esbjerg International School to test whether the system is engaging for Year 3 students, which elements of it cause engagement or disengagement, and compare it to an application with touchscreen input. The results show that the system is indeed engaging and is preferred by a majority to an application with traditional input. The most engaging elements of the system were the novelty input mechanic and auditory feedback based on input.**

Keywords—engagement; TUI; children; LEGO; interactive; storytelling.

I. INTRODUCTION

Over the past several years LEGO bricks have been augmented with digital components in a variety of projects [1] [2] [3] in order to create new types of interactions with traditional bricks. Tangible user interfaces (TUIs) developed with such bricks offer several advantages over traditional bricks, such as adding programmable elements (sensors and motors) or having elements not available in traditional bricks (sound generators, for example).

Previous work [2] has explored the possibility of digitally augmented bricks in early school education, specifically to help children improve their reading skills. The goal of that research was to design a TUI, explore its acceptance and viability, and provide design guidelines for a fully functioning prototype. The research was undertaken with groups of children aged 6-7 years old. The TUI used in the project involved a LEGO platform, and children should build models of the characters they would encounter in the story. As the story started, the children would read each page out loud, one child at a time, and place the model of a character which is performing some action in the current page to advance the story. The results of the research show that the children were accepting the technology and found certain elements of it (input mechanic and feedback) to be engaging. The children used the system the way it was expected – they all read one at a time and helped each other when they had problems. However, the viability of the technology was not fully tested as text comprehension was not investigated and the analysis of user engagement was lacking.

This research builds on top of the previous work in terms of design of the system and engagement model [2] with the intention of investigating user engagement in more detail as it was found to be a crucial element in such an educational system, since high engagement leads to more frequent use [4], which, in turn, leads to quicker mastery of skills [5]. While investigating engagement, only a few attributes (part of aesthetics, control, feedback and motivation) [6] were used and more attributes could describe the experience with more richness. Students working together with the system to achieve a mutual goal was also found [2] to be an important aspect of the experience and as such, the system should accommodate that. Another goal is to develop a fully working prototype of the system as previously such a system was only simulated for the users using the Wizard of Oz method [2]. Individual elements of the system also need to be investigated in order to find out, which of them have positive and negative effects on the experience. Finally, while TUIs generally are more engaging than their fully physical or digital counterparts [5] an additional goal of this research is to compare the proposed system with a similar storytelling application that uses a traditional (touchscreen) interface.

II. ENGAGEMENT

User engagement is an important element to consider when developing new technologies as it separates successful technologies from those that are just usable [7]. Several definitions of engagement exist in various fields of the scientific community, however, none of them are prevalent [8]. The focus in this research is on engagement as highly engaged users are more likely to keep using the system and have positive learning outcomes [9]. In a broad sense, engagement is considered to be a positive user experience while being involved in some activity to such an extent that the user wants to continue performing it [4] [10] [7] [11]. Schoenau-Fog [4] claims that continuation desire is the most important outcome of the engagement and looks into engagement in video games through four fundamental components: objectives, activities, accomplishments and affects. However, since this theory is designed with video games in mind, it is uncertain, if it would be applicable to an educational tool such as the proposed system. Laurel [10] placed the highest value on playfulness and sensory integration when trying to create something engaging. She defined engagement as a "desirable – even essential – human response to computer mediated activities" [10, p. 112]. Laurel applies principles of drama to engagement and bases her description of engagement on the connection between emotions and aesthetics. Laurel [10] also shows that continuation desire is an outcome of an engaging system, however, her focus is on user engagement in virtual worlds, while the proposed system should not subtract the users from the real world. Similarly to Laurel [10], Jacques' [12] description of engagement has roots in drama. Jacques [12] focuses on the affective state of interaction (intrinsic motivation and attention) and concluded that attention was a key factor leading to engagement when using educational multimedia. Attfield et al. [11] defined engagement as "the emotional, cognitive and behavioral connection that exists, at any point in time and possibly over time, between a user and a resource" [11, p. 2]. Attafield [11] proposed a model on how to measure engagement in web applications by separating experience into different attributes. Similar to Attafield [11], O'Brien & Toms [6] also suggest looking at engagement through several distinct attributes that together compose the state of engagement with a given technology. While the two models are rather similar, O'Brien & Toms's [6] model is slightly more detailed and has already been tested [13], while Attafield's et al. [11] has not been validated yet. O'Brien & Toms' [6] comprehensive model of engagement will serve as the main theory when investigating engagement in this project. While this description of engagement was not designed particularly for children, it was already successfully used when designing and testing interfaces to support children [13] [14]. The authors provide a list of attributes that should be addressed during the engaging process [6]:

- *Aesthetics* - visual and auditory appeal of the prototype.
- *Affect* - user's emotions towards a system.
- *Attention* - concentration on a task.
- C*hallenge* - just right amount of difficulty.
- *Control* - feel of being "in charge" while interacting with a system.
- *Feedback* - appropriate and on-time response after interaction with an artefact.
- *Interest* - exceptional interest in particular subject/artefact.
- *Motivation* - internal desire to begin and continue with an activity.
- *Novelty* - new or unexpected features of the system.
- *Perceived time* - awareness of time when doing a certain task.

These attributes cover specific areas of the prototype and its usage in the context of this research. *Aesthetics* attribute is especially important for children, according to an interview with a teacher (H. Wilkins, personal communication, 2015). *Attention* relates to how much time the children spend on-task in comparison to off-task time and is also required for learning new skills or mastering existing ones [5] [12]. *Challenge* describes the difficulty of the stories provided by the prototype for children. *Control* and *feedback* help investigate if the design of the prototype and its implementation are adequate. *Interest* and *motivation* were also chosen as foci for the study as these aspects show the acceptance of such technology. *Novelty*, in case of the proposed system, would show if the users had any prior experience with an interface that uses user-made models as input, and if that played any part in their experience. *Perceived time* allows investigating how involved in the task the children were. The attribute of *affect* - user's emotional response towards interaction [7] - is too vaguely defined for investigation as the authors do not provide an explanation of what specific emotions to look at when investigating engagement. Moreover, it is difficult to look into children's feelings towards the prototype using interviews or questionnaires due to their limited vocabulary [15].

O'Brien & Toms [7] show that the process of engagement consists of four phases: point of engagement, a period of sustained engagement, disengagement and reengagement. Each phase has several attributes that are relevant in a particular phase. The point of engagement starts with *aesthetic* appeal or *novelty* of the interface. Motivation, interest and feeling that there is enough time for a task are the intrinsic factors that start the period of sustained engagement. Sustained engagement occurs when the users reach a high level on the majority of attributes: the *attention* and *interest* are kept following appropriate feedback and controls from the application, perception of time is either lost or the participants are willing to lose it. Disengagement takes place when something malfunctions, some of the attributes lose their intensity, for example *attention* or *challenge*, or the experience is interrupted by other factors (such as distractions outside the experience). For the phase of reengagement same attributes apply as in the point of engagement, the difference is that reengagement will occur when previously maintained engagement is broken.

A. Peer learning

The social component of the experience of using the system was another factor that needed to be taken into account in relation to the design of the study and the prototype, as groupwork between children was part of the requirement, established in the previous work [2] by external collaborators, for using the system. Research shows a link between children working together and their engagement [16] [17] [18] which makes the investigation of the social component of using the system important. Children who work on tasks in groups show higher engagement in the task, specifically with respect to the *attention* attribute of engagement, in comparison to working alone. Since children work together on the same problem and help each other while doing it, peer learning (PL) theories are applicable here. PL is defined as "the acquisition of knowledge and skill through active helping and supporting among status equals or matched companions" [19]. PL has also been linked to students' motivation and interest [20] [19], both of which are also elements in the chosen engagement theory [7].

PL, applied in the classroom, has an effect on the speed of skill mastery, better comprehension of material, as well as benefits for the children's social skills [20]. These advantages over solo learning are more apparent in younger (grades 1-3) children than older children [21]. Different methods of PL application exist and can be described using two attributes: 1) equality – no clear leader, everyone's opinion is valued and 2) mutuality (extensive communication) [22] [23]. High mutuality is always desirable, but equality can be low, depending on a specific method without making the method invalid.

The most appropriate method of PL for this research is collaborative learning [16] [22] [19]. When using this method, the children are presented with a large problem, which they could solve individually, given enough time, and are encouraged to divide it into smaller parts, work on them, and present results to the rest of the group. Through presentations and feedback children gain new knowledge and fix their mistakes. A teacher moderator can facilitate discussions to increase mutuality in the group. Equality in the group depends on individual personalities of children and pre-established dynamics, but is generally expected to be high [22].

III. METHODS

This section contains details about the methods used to achieve the results of this research. Firstly, the prototype and its development are described. Next, the design of the study and data collection during the study are explained. Finally, the methods of data analysis used in this research are listed.

A. Prototype

A prototype was developed in order to help explore children's engagement while interacting with an interactive storytelling tool that uses a TUI. The users of the prototype build physical models using LEGO bricks and use them as input while reading through a story displayed on a tablet. The prototype consists of two parts: hardware and software (Figure 1). The hardware was responsible for recognising models, while the software received information about the models and performed appropriate actions (played sounds and advanced the story).

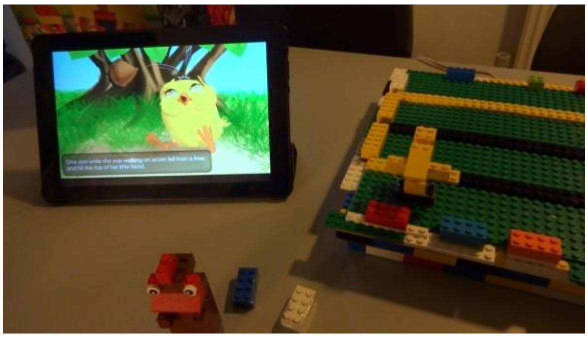

Fig. 1. The software part (left) runs on a tablet, while the hardware (right) recognises the custom-built models.

Hardware. A new piece of hardware had to be created and programmed in order to recognise custom-built models in real time and send this information to a tablet. A LEGO board is used as a surface on which the users place their models. The models must be built on specially marked bricks, to which we refer as platforms, that have resistors inside. The wires running through the surface board are connected to an Arduino microprocessor, on which an ohmmeter has been built. Using the ohmmeter, Arduino can distinguish between different platforms because they have resistors of different values. The microprocessor then sends the resistance of the platform to the tablet using a bluetooth connection.

Software. The software runs on an Android tablet and was written using ActionScript 3 and Adobe AIR programming languages. The software has a main menu, where the users can select stories by tapping their title page. After selecting the story, the users are presented with a screen that contains all the characters appearing in the story and prompts the users to build models of the characters on specific platforms. Tapping the screen again takes the users to the first page of the story they selected. Each page consists of an illustration and text. If the bluetooth connection is enabled, the users have to place the model of a character which is performing an action in the current page. If the connection is disabled, the users can tap the screen to advance to the next page. Using Android's native "BACK" button, the users can return to previous page. This design was already tested during previous work with a mock-up simulation of the prototype [2] and was well accepted and understood by the participants.

The illustrations and text in the application come from "Tales with GiGi" [24] videos, and are used in the prototype with a written permission of the original creators. Two stories ("Chicken Little" and "The Wolf and the Crane") created by the "Tales with GiGi" team are used in the prototype.

The application also has an auditory element. While the application is on, a cheery ambient music is played. If the users are experiencing a story with bluetooth connection enabled, placing a model on the surface will also produce a sound that such a character would make, for example, placing a wolf would produce howling sound effect.

Designing for learning. The goal of the proposed system is to help the user improve their existing reading skills through repetition and feedback while using it together with other users. The users are expected to build the models and read the stories with other members of their group. This was designed by keeping the peer learning theories in mind.

Collaborative learning method of peer learning is best suited for the proposed system and used in class with other students or as an extracurricular activity since children would be close to each other and improving their reading skills in a group can be a substitution for individual reading assignments. Children from the same class are very likely to have similar reading skills, which would make the group's equality high.

Each member of the group would be responsible for building at least one model and reading pages of the story in an order, determined by the group. As a result, every member would contribute to the task equally. Reading the story out loud would result in instant feedback and correction by the other members of the group, which would make mutuality of the group high as well.

B. Design of the study

The main goal of this study was to investigate user engagement while using a TUI with custom-made models as the input mechanic. In order to achieve this, a qualitative study was designed and carried out. To help better understand the experience, the was also compared to a traditional input method - touchscreen (control condition). Repeated measures design was chosen as it allowed comparing the two conditions with less participants. This was important due to limited access to students due to their busy schedules and their limited number. The individual features of the systems were also in focus to learn, which features caused engagement, and which broke it. Qualitative data was reinforced with quantitative data where appropriate. The participants of the test were Year 3 (ages 7-8) students at Esbjerg International School (EIS).

The study was carried out at EIS, which was a familiar environment for the participants, making them more comfortable [15]. The test started with students coming into the test room in groups of three and being introduced to the research and what they would need to do. Each group experienced both conditions in different orders. Two groups started with experimental condition and the other two started with control condition. The order of the stories was kept the same amongst all groups. After experiencing both conditions, the participants were interviewed.

Data collection and analysis. Measuring the engagement of children is difficult, as a large number of tests require the ability of reading complicated questionnaires, self-evaluation, verbal or written self-expression [25] [15] [26] in order to provide the researchers with reliable data. While interviews are considered a suitable source for investigating a specific experience, as they ask the participants to directly explain their experience [27], special care must be taken while designing the interviews, since the chosen target group has a smaller vocabulary than adults (which limits their expression) and they might not understand complicated questions [15]. The open-ended questions were designed to cover all the chosen engagement attributes [7]. The social component of the experience (peer learning) was also taken into account when designing the questions, in order to explore what kind of effect it can have on the *feedback* attribute of engagement, as the children provided feedback to each other on their reading. Additionally, on-task time, which at high level can indicate engaging experience [5] [12], was measured to reinforce the *attention* attribute interview data with qualitative data.

In addition to interviews, video recordings were also made. The recordings were used for interview transcriptions and precise measurement of on-task time. They also produced a different type of data in order to allow investigation of the application features. Law et al. [28] claim that the user experience must be gauged during the interaction with an artefact rather than asking afterwards. Instead of questionnaires, which can be confusing to children and, thus, lead to lower quality data [29] [30], Hanna et al. [15] proposes to use indicators of engagement and disengagement, such as smiles and frowns. Read et al. [25] improved on this method and propose to observe positive (laughing, smiling, excitable bouncing, with fingers in mouth or having tongue out being additional signs of concentration) and negative (frowns, shrugs, negative vocalizations and signs of boredom like ear playing or fiddling) markers to evaluate engagement with a product.

IV. RESULTS

This section contains the results obtained from the prototype test, conducted at the Esbjerg International School. The results are divided into three sections: interviews, engagement markers, and on-task time.

Interview questions targeting participants' perception of the aesthetics of the system revealed that all children liked the illustrations in the stories, describing them as “nice” and “funny”, and the sounds the animals made when a model was placed on the board, saying that they were “funny” and “matched the sounds [the animals] really make”. Only one child noticed the background music and suggested including a variety of tracks, as having just one track was “boring”. In two cases, the children said they did not like the sounds of wolf and fox, because of the fact that they were the bad guys in their respective stories.

Six children said that their attention was occasionally distracted by the construction work happening outside. Additionally, three children felt distracted by their own models, when it was not their turn to read, with one participant noting that “they were talking to me”.

When asked about the difficulty of the text in the stories, four participants described them as “medium” in difficulty, while the other seven felt it was easy.

Four participants had some experience with systems that bring their physical LEGO toys to life. The systems mentioned were LEGO Mindstorms, where children “can build a robot and then put [commands] on computer”, and LEGO Dimensions, a video game that allows the users to “build something and bring it to your computer”.

While the children said that they can tell time, they had difficulty evaluating the time it took them to test the system. The actual test took about 30 minutes for each group, but the children’s answers varied from 23 minutes, to two hours.

The majority of the children felt that the system worked fine. Four children felt that there were problems with the system. Two of them were disturbed by the system not recognising the models on the first try, one felt that the models were “difficult to remove” from the board, another disliked the part where they had to build the models.

All children liked working together with others. In the future, only one of them would like to work alone instead of with others, saying that he would only need others because he doesn't "want to read all of [the text] by myself".

Despite all the children saying that they would like to use the system again, one participant said that he would prefer traditional books over the system, saying that he "[likes] reading books" and "has a normal book at home".

Finally, when asked to retell the stories, all participants were able to correctly recall the characters, the events and even the message the stories were sending.

Overall, interview data shows that while there were several instances when the engagement was broken, the users felt that the experience of using the system was engaging as most of the attributes were at a high level.

Engagement markers (Table 1) show that most engagement was caused by the sounds the system produced as well as the input mechanic (placing models to advance the story). Disengagement with the input mechanic was noticed when the children were fiddling with their model waiting for their turn, or when the model was not recognised by the system. Story (text) and illustrations caused engagement in both conditions. However, signs of boredom in relation to following the story were noticed in the control condition.

The additions to an experience with traditional touchscreen interface had mostly positive influence on the user engagement, while taking some of the attention away from the illustrations and making the story feel less boring.

TABLE I. ENGAGEMENT MARKER DATA

Affected attribute	Type	Condition	Instances	Markers
Sounds	Pos.	Exp.	17	Positive vocalisation, smiles
Input mechanic	Pos.	Exp.	14	Smiles, positive vocalisation
	Neg.	Exp.	7	Frown, fiddling
Illustrations	Pos.	Exp.	2	Smiles, positive vocalisation
	Pos.	Control	8	Smiles, positive vocalisation
Story	Pos.	Exp.	9	Positive vocalisation, smiles, finger in mouth
	Pos.	Control	9	Smiles, positive vocalisation
	Neg.	Control	2	Yawn, fiddling

On-task time was a percentage of how much time the children have spent doing task-related things (looking at screen, reading and placing models). The data, based on the Wilcoxon signed-ranks test, show no statistically significant difference between the versions ($p > 0.05$), which means that the condition appears to have had no effect on the attention directed towards it.

V. DISCUSSION

The aims of the study were to develop a fully functional prototype of an interactive storytelling tool that uses user-built LEGO models as input mechanic, evaluate its features, user engagement while using it, and compare it to an application that uses a traditional interface to display stories in digital format.

The study revealed an issue with model recognition that results in incorrect feedback from the system. Apart from that, all other features of the prototype function as expected - when a model is recognised, the application provides auditory feedback and, in case of a correct model being placed, turns a page in a story. Features unique to the system (auditory feedback and input mechanic) caused engagement when they functioned properly. Interviews showed that using the system was an engaging experience for the children and they liked it more than the version with traditional interface.

While overall the prototype was engaging, one of the prototype's key features - input mechanic, or more specifically, the models themselves, had caused the children to disengage from the story and the system. This might have occurred due to the fact that they are used to playing with the models they build. The distractions caused by the models were also apparent in previous work [2] and in an attempt to avoid them in this study, the number of bricks available and the size of the models were limited. This has reduced the build time and resulted in more structurally stable models. However, it appears that removing these distractions completely is impossible in an environment where children work with their peers, as they need to wait for their turn to read. During this time they might get impatient if the current reader is having difficulty. This behaviour is not limited to children, educational material or playing with LEGO. Users who are currently waiting for their turn can be referred to as "passive players" and research shows [31] that not having something to play with while waiting for their turn results in worse overall experience than when fiddling with pieces and parts belonging to the system.

Markers of boredom on story in control condition as well as a lower number of positive markers on illustrations in control condition may indicate that the other features captivated the children's focus more in the experimental condition, and since they were not present in the control condition, caused boredom. This further reinforces the above mentioned finding that "passive players" need something to do while waiting for their turn.

The interview data showed that the children liked the system, and, based on the individual attributes of engagement [6], found it to be engaging, since all the investigated attributes were at a high level required to reach engagement. When comparing the two versions, the children showed preference to the system and cited sounds and input mechanic as the main reasons for their choice. This is in line with [5], where a TUI in a collaborative environment was also compared to its GUI counterpart.

Finally, as [2] only investigated the acceptance of the system and not the viability, it was an important factor to look at. In this study, the viability was considered to be whether the system has a negative effect on the affected educational material. Specifically, if children understand and retain the stories they read as well with the system as they do with a traditional interface. Based on the fact that all the children could retell the stories and reason on the lessons they were trying to teach, regardless of condition, it is likely that the system and other distractions had no effect on the comprehension and retention of the material and, combined with it being more engaging, is thus a viable replacement for stories told through traditional interface (digital or paper).

A. Engagement in relation to attributes

The attribute of *aesthetics* consists of participants' thoughts on visuals and sounds. While the illustrations were liked by all the participants, engagement markers show that they were engaging to children in the control condition. As the illustrations were similar in both stories, it could be said that they were more engaging in the control condition because the children had no other things, such as sounds or models in the experimental condition, to focus on.

The background music was present in both conditions, but was only noticed by a single participant, who wanted the application to have more than one track. The volume of the music was kept low in order not to distract the participants. As such, it can be said that the background music had no effect on the user engagement, meaning that it could be changed or entirely removed without causing negative experience for the participants.

The sounds the animals made were positively received by the participants. The exceptions occurred when children's perception of the sounds was influenced by the characters' actions in the story, i.e. they did not like the sound of the fox, because it killed the other animals in "Chicken Little". The sounds were the most engaging element of the system based on the markers, showing that it is a crucial feature of such a system.

Participants' *attention* was occasionally distracted by outside factors, such as construction work outside the test venue or posters in the test room. However, on-task time data shows that, with exception of one child, everybody spent over 90% of time on-task.

The *challenge* in the stories was appropriate for all the participants, with hints of it being too easy in case of two children who displayed signs of boredom in the control condition. However, the boredom could also have been caused by the wait time, as the markers were observed when the children were waiting for their turn to read. Regardless, due to the difficulty being appropriate (based on both the participants and their teacher), it is very likely that the system had no effect on the level of challenge.

While four participants had experience with similar systems, it does not appear that the *novelty* factor had any effect on using the system, as the participants who were familiar with such type of interface did not display different behaviour or interview answers on other attributes. This means that either: 1) the novelty played no part in the children's engagement with the system; or 2) the novelty was present for all of the participants.

The participants had trouble estimating how long the test took, despite bragging that they were able to *tell time*. Four out of eleven children told approximated the time spent correctly (within a small margin of error) by themselves. Most answers appeared to be more like random guesses rather than estimations, as well as being influenced by what the peers said. As a result, it is difficult to say if the system had any impact on the children's perception of time. It is possible that either: 1) the children were engaged enough to lose the perception of time; or 2) the students still had problems understanding and expressing time.

The input mechanic was another element that was observed to cause engagement based on a large amount of markers and interview answers. Children were smiling when both building the models and using them as input for the system. However, problems with the system caused disengagement. The hardware malfunctioned in several occasions, resulting in unrecognised models. While problems with hardware can occur on any system, the teacher noted that when introducing a system for children, it should work all the time or have failsafes to ensure smooth and uninterrupted experience. As a result, the input mechanic had a mixed effect on the engagement with the system: it made the system more engaging when it worked, but it caused disengagement when it did not. Eliminating the disengagement factor should be possible by reworking the system, making the recognition of the bricks more reliable. This can be done by using, for example, a board based on copper pegs rather than wires as a surface of detection, or using completely conductive bricks to ensure that the entire surface of the brick is used for detection rather than just the points where the resistors come out of the brick.

In addition to auditory and visual *feedback* from the system (in the form of sounds and changing pages), the children also received feedback from their peers. All the children liked working with others due to it being more fun, out of convenience (not having to do all the work alone) or for learning purposes. Since working together was present in both conditions, it is likely that peer feedback did not play an important role between the two conditions. However, the peer learning theory (see section Peer learning) strongly suggests that such a system might have a positive effect on children's learning in comparison to working alone.

Since all of the children liked working with the system and most would choose it over paper books if they would have such a possibility, it is highly possible that the system created a strong *motivation* towards using it. Motivation appears to stem from the differences between two versions: sounds, interactivity, element of freedom (custom models). It is, however, unclear whether the motivation would persist over a period of prolonged use of the system, as a longitudinal study is required to answer this.

The attribute of *affect*, as previously mentioned (see section Engagement), was not investigated due to several reasons. However, while it is difficult to evaluate affect, it could potentially be measured using the engagement markers [25]. However, this would require a deep understanding of expressed emotions of the chosen target group to be able to translate markers into feelings. This would provide a more reliable set of data than asking children to talk about their feelings in an interview, which would be problematic due to their limited vocabulary [15].

B. Bias

The researchers are aware that this study has its biases and that testing outcomes could have been affected by them.

Children are prone to bias [15] (T. V. Snitker, personal communication, 2014). It is difficult to get reliable answers from children since their vocabulary is smaller than that of adults (which results in simplified interview questions), they might be trying to please the interviewer, or feel peer pressure when other children are present. This was especially noticeable during the interview question about perceived time, when one participant shouted “5 hours!”, but when his peer said that “it took 1 hour and 25 minutes”, he instantly changed his mind and said “we took 2 hours! 35! Ok, then… 1 hour”. Additionally, some of the answers like “fun” or “nice” are not very descriptive and could mean different things to different children. As a result, the interpretation of such phrases is also a cause of bias.

The video camera used to film the test was placed in front of the children, and although all participants were told that they would be filmed, it still could have caused distraction and tension for the children [15]. For example, when he became aware of the camera during the interview, the participant pointed to it and asked “Is it still filming?”. He then tried to hide from the view of the camera by crawling under the table. Thus, it is possible that other children were consciously aware of the camera during the test and modified their behaviour.

The small sample of participants provided only general observations, and large scale test would be needed to provide more reliable data. Additionally, on-task time, which is part of the engagement, provided no statistically significant results. This could also be improved by having a larger sample.

The analysis of engagement marker data in relation to comparison between the two versions is also prone to bias, due to the theory [6] not taking into account possible differences of magnitude between the various markers (i.e. laughter could indicate higher engagement than smile).

Currently, the conclusions can only be drawn from the single study conducted during this research, and might change on a longitudinal study with several separate tests and the same group of children - the long-term exposure to the system might have an effect on the children’s motivation to continue using the system.

VI. CONCLUSIONS

This study shows that using a system with user-made physical models as input mechanic in a storytelling application results in an engaging experience for children, more so than using a system with traditional input method, especially due to use of novelty input mechanic and auditory feedback based on the input. Feedback from peers in a form of help and hints on reading also resulted in a positive experience. For the purposes of the study a working prototype of such a system was developed. The features that were not available in a fully digital application (sounds and novelty input mechanic) were engaging. However, malfunctions associated with these features caused disengagement in some participants, which could be avoided given further development of the prototype.

Due to the small sample and other bias, the reliability of the study is questionable. As such, at the current stage the prototype should not yet advance to final production. However, this study should serve as a starting point for a large scale, long-term study with an improved version of the prototype.

REFERENCES

[1] B. Almog, "Brixo," 2016. [Online]. Available: http://www.getbrixo.com/. [Accessed May 2016].

[2] A. Kulšinskas and P. Konopackas, "Internship report," 2015.

[3] LEGO, "LEGO Ultra Agents," 2016. [Online]. Available: http://www.lego.com/en-us/ultraagents. [Accessed May 2016].

[4] H. Schoenau-Fog, "The player engagement process--An exploration of continuation desire in digital games," in *Think Design Play: Digital Games Research Conference*, 2011.

[5] L. Xie, A. N. Antle and N. Motamedi, "Are tangibles more fun?: comparing children's enjoyment and engagement using physical, graphical and tangible user interfaces," in *Proceedings of the 2nd international conference on Tangible and embedded interaction*, 2008.

[6] H. L. O'Brien and E. G. Toms, "The development and evaluation of a survey to measure user engagement," *Journal of the American Society for Information Science and Technology*, vol. 61, no. 1, pp. 50-69, 2010.

[7] H. L. O'Brien and E. G. Toms, "What is user engagement? A conceptual framework for defining user

engagement with technology," *Journal of the American Society for Information Science and Technology,* vol. 59, no. 6, pp. 938-955, 2008.

[8] E. A. Boyle, T. M. Connolly, T. Hainey and J. M. Boyle, "Engagement in digital entertainment games: A systematic review," *Computers in Human Behavior,* pp. 771-780, 2012.

[9] B. A. Greene, "Measuring cognitive engagement with self-report scales: Reflections from over 20 years of research," *Educational Psychologist,* vol. 50, no. 1, pp. 14-30, 2015.

[10] B. Laurel, Computers as theatre, Addison-Wesley, 2013, pp. 111-113.

[11] S. Attfield, G. Kazai, M. Lalmas and B. Piwowarski, "Towards a science of user engagement (position paper)," in *WSDM Workshop on User Modelling for Web Applications*, 2011.

[12] R. Jacques, "Engagement as a Design Concept for Multimedia," *Canadian Journal of Educational Communication,* vol. 24, no. 1, pp. 49-59, 1995.

[13] L. J. Corrigan, C. Peters and G. Castellano, "Identifying task engagement: Towards personalised interactions with educational robots," in *Affective Computing and Intelligent Interaction (ACII), 2013 Humaine Association Conference on*, 2013.

[14] J. P. Rowe, L. R. Shores, B. W. Mott and J. C. Lester, "Integrating learning, problem solving, and engagement in narrative-centered learning environments," *International Journal of Artificial Intelligence in Education,* pp. 115-133, 2011.

[15] L. Hanna, K. Risden and K. Alexander, "Guidelines for usability testing with children," *interactions,* vol. 4, no. 5, pp. 9-14, 1997.

[16] W. Damon, "Peer education: The untapped potential," *Journal of applied developmental psychology,* vol. 5, no. 4, pp. 331-343, 1984.

[17] S. D. Scott, R. L. Mandryk and K. M. Inkpen, "Understanding children's collaborative interactions in shared environments," *Journal of Computer Assisted Learning,* vol. 19, no. 2, pp. 220-228, 2003.

[18] K. Inkpen, K. S. Booth, S. D. Gribble and M. Klawe, "Give and take: Children collaborating on one computer," in *Conference companion on Human factors in computing systems*, 1995.

[19] K. J. Topping, "Trends in peer learning," *Educational psychology,* vol. 25, no. 6, pp. 631-645, 2005.

[20] K. R. Wentzel and D. E. Watkins, "Peer relationsips and collaborative learning as contexts for academic enablers," *School Psychology Review,* vol. 31, no. 3, pp. 366, 2002.

[21] C. A. Rohrbeck, M. D. Ginsburg-Block, J. W. Fantuzzo and T. R. Miller, "Peer-assisted learning interventions with elementary school students: A meta-analytic review.," *Journal of Educational Psychology,* vol. 95, no. 2, pp. 240, 2003.

[22] W. Damon and E. Phelps, "Critical distinctions among three approaches to peer education," *International journal of educational research,* vol. 13, no. 1, pp. 9-19, 1989.

[23] A. King, A. Staffieri and A. Adelgais, "Mutual peer tutoring: Effects of structuring tutorial interaction to scaffold peer learning.," *Journal of Educational Psychology,* vol. 90, no. 1, pp. 134, 1998.

[24] GiGi, "Tales with GiGi," 2016. [Online]. Available: http://taleswithgigi.com. [Accessed May 2016].

[25] J. Read, S. MacFarlane and C. Casey, "Endurability, engagement and expectations: Measuring children's fun," in *Interaction design and children*, 2002.

[26] D. Y. Xu, J. C. Read, G. Sim, B. McManus and P. Qualter, "Children and 'smart' technologies: can children's experiences be interpreted and coded?," in *Proceedings of the 23rd British HCI Group Annual Conference on People and Computers: Celebrating People and Technology*, 2009.

[27] J. W. Creswell, Qualitative inquiry and research design: Choosing among five approaches, Sage, 2012.

[28] E. L.-C. Law, V. Roto, M. Hassenzahl, A. Vermeeren and J. Kort, "Understanding, scoping and defining user experience: a survey approach," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems,* 2009.

[29] D. Mellor and K. A. Moore, "The use of Likert scales with children," *Journal of pediatric psychology,* vol. 39, no. 3, pp. 369-379, 2014.

[30] C. T. Chambers and C. Johnston, "Developmental differences in children's use of rating scales," *Journal of Pediatric Psychology,* vol. 27, no. 1, pp. 27-36, 2002.

[31] A. Krzywinski, W. Chen and E. Røsjø, "Digital board games: peripheral activity eludes ennui," in *Proceedings of the ACM International Conference on Interactive Tabletops and Surfaces*, 2011.