

SmallTalk: Using Tangible Interactions to Gather Feedback from Children

Sarah Gallacher¹, Connie Golsteijn¹, Yvonne Rogers¹, Licia Capra¹, Sophie Eustace²

¹University College London, London, WC1E 6BT, UK

²Fevered Sleep, 119 Farringdon Road, London, EC1R 3DA, UK

{s.gallacher, c.golsteijn, y.rogers, l.capra}@ucl.ac.uk, sophie@feveredsleep.co.uk

ABSTRACT

Gathering opinions from young children is challenging and different methods have been explored. In this paper we investigated how tangible devices can be used to gather feedback from children in the context of a theater performance. We introduce SmallTalk, a tangible survey system designed for use within a theater space to capture what children, aged 4 to 9, thought of a live performance they had just seen. We describe how the system was designed to build on previous feedback methods that had been tried; while at the same time meeting the constraints of the challenging theater context. We present results from seven deployments of SmallTalk and based on these we briefly discuss its value as a method for evaluating the theater performance. We then look at how the results validated the system design and present several design implications that more generally relate to tangible feedback systems for children.

Author Keywords

Tangible interaction; Opinion gathering; Design for children; Design research; Feedback systems

ACM Classification Keywords

H.5: Information interfaces and presentation (e.g., HCI):
H.5.2. User Interfaces; H.5.m. Miscellaneous.

INTRODUCTION

When designing for young children – be it a new product, service, or experience – it is important to evaluate and gather feedback in age-appropriate ways [12]. However, this can be a challenging task, as evaluation methods that are successful with adults, such as surveys and interviews, are not easily transferred to research with children. Recurring challenges have been identified, such as satisficing (when a child gives reasonable answers but does not carefully consider the question and answer), suggestibility (the impact of the researcher on the child), and how to format questions

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

Request permissions from Permissions@acm.org.

TEI '16, February 14 - 17, 2016, Eindhoven, Netherlands

Copyright is held by the owner/author(s). Publication rights licensed to ACM.

ACM 978-1-4503-3582-9/16/02...\$15.00

DOI: <http://dx.doi.org/10.1145/2839462.2839481>

(children tend to say “yes” no matter what the question is) [20]. Furthermore, since evaluations with children are often carried out one-to-one with a researcher or in pairs or small groups, it can be time-consuming if feedback from many children is required.

Based on our previous work on physical feedback systems for festivals and events [10], we were approached by a London production company, called Fevered Sleep, who wanted to investigate how young children (aged 4 to 9 years) could be involved more directly in evaluating live performances that they had just seen at the theater. This included, looking at what children remembered about a performance, and how they reflected on it. In the past, the company had experimented with different methods including survey forms, image elicitation, and book-making activities, noting challenges of maintaining the children’s attention or gathering answers to open-ended questions. Furthermore, these methods required significant mediation making them difficult to do in situ immediately after the performance, due to the strict schedules of the theater.

In collaboration with Fevered Sleep, we designed a tangible survey system called SmallTalk, based on requirements derived from challenges encountered with other methods, and the demanding theater context. SmallTalk was designed as a set of five interactive question boxes with which children interact in turn. The boxes ask questions about the performance and the theater experience. The children provide their answers through tangible interactions, such as buttons, dials, and spinners. A study of their use at seven deployments in a theater space captured a wide range of answers from children. The results suggest several design implications for obtaining feedback from young children including, continuity of interactions, atmosphere of the wider environment and open-ended questioning.

BACKGROUND

The work presented in this paper brings together two strands of research that have not been combined in this way before: 1) evaluation methods for research with children, and 2) tangible interaction for children. Research has investigated how to evaluate interactive products – including novel tangible interaction designs – with children. In parallel, it has studied the benefits of tangible interaction when designing for children. We combine these two important strands of work in exploring *how tangible interaction can be used in a*

novel evaluation system that can gather feedback and opinions from children.

Evaluation Methods for Research with Children

Within HCI, gathering feedback from children is most common in usability research with related works typically comparing usability methods to see which work best in different contexts. Many of the classic usability methods are not as effective when used with children – who can find it difficult to answer open-ended questions or use rating scales for paper-based or online surveys. Indeed, one study showed that ranking designs and discussing rationale were more effective than rating designs and asking open-ended questions [11]. Others have compared methods for eliciting verbal comments from children, finding that the best approach was to prompt a child with questions while they were performing tasks with the product [15]. In contrast, a later replication of the study with more participants found that peer tutoring was best [9]. While such ‘think aloud’ methods have had success with older children [4], they are often difficult for younger children [7]. Another approach has been to elicit answers through drawing [19]. Observation methods have also been explored, but can risk interpretation bias [18].

In response to these difficulties and conflicting findings, new methods have been developed or adapted especially for children. An example is the Fun Toolkit that enables children to distinguish between different constructs of fun [21]. Another is a pairwise comparison method called ‘This or That’, which lets children select one product over another for a number of questions. It was found to be more consistent than the use of multi-dimensional scales and rating activities [24]. Problem Identification Picture Cards (PIPC) combine thinking-aloud with picture cards that children place in a box when they identify a problem. They were found to help children identify more problems with a product over simply being asked to verbalize them [6].

While these new methods show promising results, more work is needed to validate them and understand when they can be used [18]. Moreover, with the exception of the Fun Toolkit, there is a strong emphasis on usability. When carrying out more open-ended evaluations that go beyond whether something is usable or fun, there is a need for more explorative, engaging, and open-ended tools. To this end, tangible interaction offers much potential.

Tangible Interaction for Children

One of the aims of using tangible interaction is to design user experiences that are embodied, natural, and fun [8, 4]. It is assumed that physical interfaces are easier to use, support learning and development, and can facilitate collaboration [e.g., 5, 13, 23, 25]. For example, researchers have used children’s development theory to provide guidelines on how to design tangibles for children [1]; have compared how children interact with tangible versus graphical user interfaces [3, 22]; and have developed tangible learning systems [5, 13]. While these studies show benefits of

tangible interaction in the anticipated areas of fun, engagement, and learning, there is little theoretical or empirical evidence for the claimed benefits, or validated design guidelines on tangible interaction for children [2, 17, 25]. More empirical work is needed that investigates the value of using tangible interaction for eliciting feedback from children. The aim of our study was to investigate the efficacy of using tangible interaction *in situ*, to gather feedback from children after a theater production.

DESIGN REQUIREMENTS

SmallTalk was used to evaluate a performance, called Dusk (devised and directed by David Harradine and Samantha Butler). Dusk is a form of immersive theater that juxtaposes real life performances with watching a film. The narrative running throughout is about a hybrid creature, called an ‘It’, who loses his family and friends one day and then goes on a journey to find them. On this journey he travels through different places, such as the beach and a mountain, and encounters different creatures, such as an owl and a girl. The film appears on a large screen in the theater space while an actor, another It, is present and interacts with the children when the film is interrupted. The actor and the audience are all wearing life-like fluffy tails, to look like Its, and it becomes clear that the It in the film is trying to find his way to the physical theater space where his family and friends are (the audience). Audience interaction consists of dancing and howling in the hope that the lost It can hear them. This unusual experience – part interactive cinema, part theater - reaches its climax when the lost It in the film goes through a red door on screen and physically emerges through an identical, real, red door that exists in the theater space where the audience are. The performance concludes with spontaneous dancing and howling as the lost It is reunited with his community and the line between fantasy and reality for the audience is shattered. Dusk continues the company’s fascination in the child’s eye view of the natural world, exploring themes of home, loss, community, change and reunion.

Getting children as young as four to reflect on their phenomenological experience is a challenge. The production company was keen to understand more on how the children remember and reflect on the performance; what stuck in their minds and what emotions it evoked in them. This kind of feedback can help them to develop further age-appropriate productions. To evaluate previous productions, the company had experimented with a variety of methods with the most recent comparing tailor-made iPad and paper-based book-making activities, in which children selected images from the performance and theater experience to create their own ‘stories’. Based on the findings of previous methods, we collaboratively developed a set of requirements for the design of a new, tangible approach.

1) Engaging and Physical Experience: Although both iPad and paper-based approaches resulted in valuable feedback on what children remembered about the performance and what

it meant to them, children showed more interest in making their own physical storybooks than in assembling images in the story app. In line with this attraction to the physical, a first requirement was to explore a physical means for gathering feedback that was engaging and capable of maintaining the children's attention. It should be an exciting activity in itself, rather than a task that children have to do after a fun theater experience.

2) *In situ Interaction*: Previous theater performance evaluations took place at the children's school two days after the performance happened. Evaluating in the theater space immediately after the performance can be beneficial as children may remember more detail and are still 'in the moment'. A second requirement was thus to create a system that could be used in the theater space after the performance. Additional challenges come into play, such as mediation and time required.

3) *Minimal Mediation*: While book-making activities and interviews with children can provide valuable feedback, they require much mediation. As there are often multiple performances in rapid succession and production staff have other duties, it is not always possible to do such evaluations immediately after a performance. Therefore, we aimed to design a system that required minimal mediation and could be mediated by anyone without influencing the results.

4) *Time-efficient Evaluation*: A related challenge is the time it takes to carry out evaluations. As the theater space had to be prepared for the next performance as soon as the previous one finished, there was a limited amount of time in which evaluations in the space could be carried out. Another requirement was thus to design a system that could be used by a group of children in a short time span.

5) *Open-ended Feedback*: A further goal was to gather feedback to open-ended questions and qualitative results. While multiple choice surveys and rating or ranking scales can be done fairly quickly, they are typically limited in the insights they can give, and risk children saying "yes" no

matter what the question is [20]. Other methods, such as interviews or image elicitation, can give more detailed feedback but take much mediation and time. We thus also aimed to gather open-ended feedback while still striving to keep mediation and time required to a minimum.

DESIGNING AND IMPLEMENTING SMALLTALK

The questions to be asked through the SmallTalk system were developed in collaboration with the production company and aimed to find out what the children most remembered and reflected on from the performance. After initial brainstorming, a large list of questions were clustered under categories and condensed into a much smaller, final set. The questions in this final set were then logically grouped together into five themes and then realized on five physical interactive boxes (one per theme). Figure 1 shows each of these boxes, including the questions (Q) that were asked on each, and the physical interactions (I) used to input answers. Boxes 1 and 2 asked simple demographic questions and about coming to the theater. Boxes 3 and 4 were directly related to the performance, each asking one question about locations and creatures in the performance. Box 5 asked open-ended questions on what the children remembered about the performance and the emotions it evoked. As time was limited in the theater context, it was decided that only box 5 would gather open-ended responses due to the potentially longer interaction times associated with this type of input.

Each box had a 'Start' button at the top and 'Finish' button at the bottom to activate the box and submit the responses respectively. Audio was used on each box to support children who could not yet read. Hence, in addition to being able to read the question and answer text, 'question audio' buttons could be pushed to hear the question and 'answer audio' allowed the child to hear the selected answer, for example when they pushed an answer button or turned a dial through the answer positions. In all audio recordings, the voice of the It from the performance was used, as he was familiar to the children.

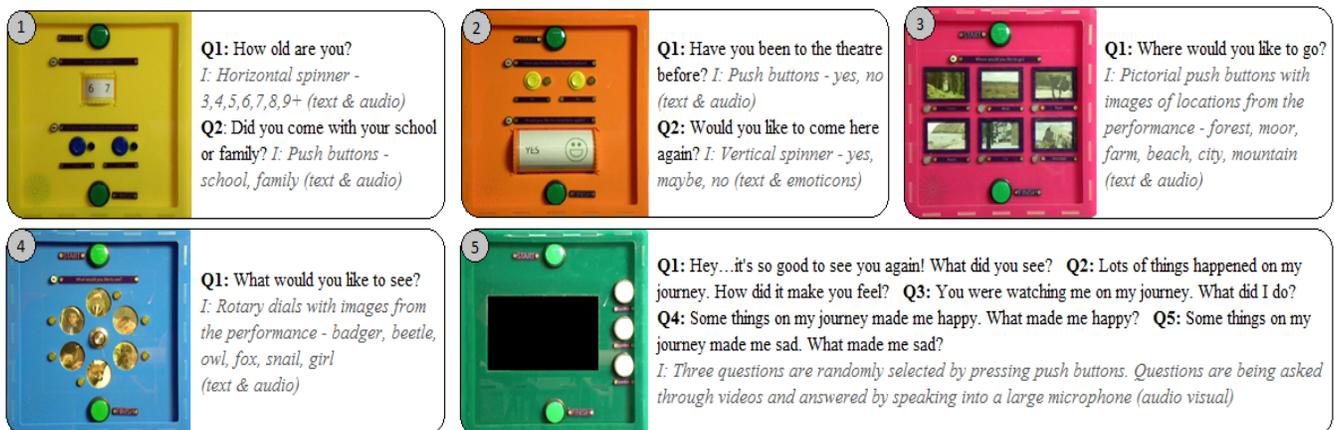


Figure 1. Overview of boxes, questions, and interactions. All questions were printed on labels and accompanied by an audio button next to this label that could be pressed to hear the question. Answers were played aloud when selected, e.g. by buttons.

SmallTalk was thus designed to provide an *engaging and physical experience* by having a novel physical form factor and by using a variety of different physical interactions. The first four boxes asked multiple-choice questions, and different buttons, spinners and dials were used to keep the interactions interesting throughout. Secondly, It was also designed for deployment in the theater space so that the evaluation could take place *in situ after the performance*. Each box was placed on a sturdy stand so that it was at the right height for a child. The boxes were hidden behind a curtain during the performance and were quickly uncovered afterwards and evenly positioned across the theater space. *Minimal mediation* was met by using uncluttered interfaces, simple interactions and audio support to make it possible for children to complete the boxes autonomously. Using five separate boxes also enabled a *steady flow of children to interact in a time-efficient manner*. We envisioned children using the boxes one by one in parallel, thus limiting the total time needed to gather feedback from a group.

The fifth box was designed to gather richer feedback from children through *open-ended and qualitative results*. In previous work around standalone opinion systems a telephone handset had been used to ask and answer open-ended questions [10]. However, findings showed that although this worked well for adults, children were often taken aback by a disembodied stranger's voice on the other end of the line. We therefore designed box 5 to use video clips instead, and children could view a maximum of three open-ended questions (chosen at random from a set of five) by pressing corresponding buttons on the side of the screen. We filmed the actor, who played the It in the performance, asking questions while looking at the camera as if talking to the viewer. After each question the It paused, looked into the camera, and tilted his head as if listening for an answer. In this way a friendly and familiar character was asking the questions, and a large fluffy microphone was placed on the side of the box for the children to answer into.

STUDY METHOD

We worked with the production company to develop a schedule of what performances we could evaluate with SmallTalk. The final schedule consisted of 7 performances (both matinee and evening) over a two-week period. Three of the performances were open to the general public and the other four were for school groups.

During each evaluation, the five SmallTalk boxes were set up across the theater space. After the performance parents, teachers and children relaxed in the theater cafe for 5-10 minutes while the researchers made sure everything was set up for the evaluation. When a group entered the theater space, a researcher welcomed them and introduced them to the five boxes. She explained that the children should go to each box in turn and answer the questions. The children then formed a queue at the first box and proceeded along the boxes in sequence. Three researchers were in the space during each evaluation. One researcher would always be

present at the first box to help the children and explain what to do, after which they continued to the other boxes by themselves. Another researcher was typically on hand nearby box 5, (video box) in case the children did not understand what to do, as it was different from the other four boxes. The researchers gave the children space to find out what to do at each box and how to interact, but were on hand to give simple guidance if a child did get stuck.

Three video cameras were deployed in the space to capture different angles of the action. Two video cameras were statically positioned at each end of the row of boxes and the third was a hand-held, operated by a third researcher who observed discretely from a distance. The video footage also allowed researchers to confirm the order in which the children interacted with the boxes, as inevitably the children would not remain in the same order.

Data Collection and Analysis

A variety of data collection methods were used during each evaluation. Firstly, data was automatically recorded from each box, showing what answers each child provided to the questions. In the video box this included audio recordings of the children's spoken answers to the video questions. Secondly, video recordings were captured during each evaluation for analysis afterwards to identify how the children interacted with the boxes. Finally, researchers logged observations immediately after each evaluation by making written notes and highlighting interesting events to be investigated further during data and video analysis.

Video recordings were analyzed through open coding – the clustering of similar instances in the data without a predetermined coding scheme or hypothesis [16] – to identify interesting findings on how children interacted with the boxes, such as whether they read or listened to the questions, how they understood the affordances of the different physical inputs and how they explored the answer space and provided their responses. This form of data collection and analysis allows for the evaluation of both the performance (Fevered Sleep aims) and the use of SmallTalk (researcher aims). We report on both, giving emphasis to the use of SmallTalk and how this data can inform future design of tangible feedback systems for children.

RESULTS

A total of 63 children, of which 40% were male and 60% were female, used the SmallTalk boxes over seven days of theater performances. Here we present the results from the study, firstly detailing the answers and interactions with the multiple-choice questions at boxes 1 to 4, and secondly detailing the answers and interactions with the open-ended questions at box number 5.

Multiple-Choice Question Responses

The answers logged at box 1 showed that the most common age of participants was 5 (38%) and 6 (25%) with only 9% being older than 7 and none being younger than 4. The majority of children came to the theater with their school

(57%). Box 2 showed that a large number of the children had been to the theater before (78%) and the majority (86%) would consider coming back again. Figure 2(a) shows the responses at boxes 3 and 4 where the questions were more reflective, and directly related to elements of the theater performance. Most children wanted to visit the city (25%) and meet the fox (27%). Children spent 12 seconds on average at boxes 1, 2 and 4, and 10 seconds at box 3.

Interactions with Multiple-Choice Question Boxes

While using the first four SmallTalk boxes, the majority of children required minimal guidance from researchers and were largely able to complete the boxes autonomously. When guidance was required it was typically to give simple instruction of what to do next (usually on the first box while the children were getting accustomed to the tangible form factor). In general, we observed the children becoming more relaxed and confident as they moved down the line of boxes and better understood how to interact with them. Drawing on the video data, five children looked for adult confirmation of an answer before they entered it, otherwise answers were the children's own and not influenced by adults. Children also helped each other to overcome a challenging point and on five occasions we observed those waiting to use the box helping the current child to progress. In two cases, children worked together to decide whether they had come to the theater with their 'school' or 'family'. Although it seems the answer should be obvious, for those who came with 'mum', the options available did not exactly fit the bill, as one child stated; *"mum is not my whole family"*. Other queries centered on whether or not they had been to the theater before to see Disney's 'Frozen'. Understandably, for some, the term 'theater' also related to a movie theater or cinema.

In terms of understanding what each question asked, the majority of children used the audio buttons to listen to the questions rather than reading them. Even older children who could read well preferred to listen to the audio question or reverted to using audio after reading the first few questions. After listening to an audio question, at least six children naturally responded by speaking their answer back to the box before realizing they had to enter it manually. On another six occasions children were observed ignoring the question, instead moving straight to the answer interactions and entering their response. This only happened on boxes 3 and 4 where the answer interactions were pictorial.

With regard to understanding how to answer each question, our observations show that children of five years and younger were generally less familiar with the affordances of the different physical interactions on each box. The dial on box 4 seemed to be the least familiar physical input with seventeen children not recognizing what it was or how to use it. Of these seventeen children, twelve were aged five or younger. Additionally, some of the younger children tried to apply 'touch-screen like' interactions to the physical inputs. For example, three children were observed lightly touching and lightly swiping the round push buttons at boxes 1 and 2,

rather than pushing them. A researcher then showed how to push the buttons until they clicked.

In terms of exploring the answer space and deciding on a response, the majority of children took some time to do this mentally (without interacting with the box), before manually entering their response. Three children were observed speaking aloud to themselves while contemplating and another three children used their finger to trace around the possible responses while thinking. The dial at box 4 encouraged the most physical exploration of the answer space. At least six children physically turned the dial through each possible answer while considering their response before moving the dial back to the chosen position. Only two children were observed exploring the answer space in this way when the interactions involved buttons or spinners. However, it was also observed that children tried to select the 'correct' answer button with their first interaction. On two occasions where a child pushed a button for an answer they didn't want, they showed disappointment, as though they had lost a game, until a researcher pointed out that the answer could be changed until the 'Finish' button was pressed.

Open-Ended Question Responses

From a total of 63 children, 43 provided a spoken answer for all three questions they were asked at box number 5, with a further 14 answering at least one of them. Only 6 children didn't answer any of the questions. These tended to be the youngest children, aged four or five. Of these, two showed signs of discomfort or voiced that they *"didn't want to"*, while the others stared silently at the screen despite gentle prompts from researchers. Figure 2(b) shows the response rates to each of the video questions (i.e. of those who viewed the question, how many answered it). The average time spent at this box was 1 minute, 5 seconds.

Regarding the first question (*"Hey...it's so good to see you again! What did you see?"*), the majority of children (35) mentioned seeing an animal and in 10 cases this was the same animal that they had chosen on box 4. Eight children mentioned seeing a location and in 3 cases this was the same location that they had chosen on box 3. The other responses were varied, and related to many different elements of the theater performance, some which were quite prominent and others that had a much more fleeting role. Answers included *"You!"*, *"I saw a lion"*, *"A log"*, *"We saw the boat"*.

For the second question (*"Lots of things happened on my journey. How did it make you feel?"*), the majority of children (20) responded with a positive emotion such as *"happy"*, *"excited"* or even *"silly"*. The rest were a mix of positive and negative emotions including *"sad"*, *"scared"* and *"worried"*.

Answers to the third question (*"You were watching me on my journey. What did I do?"*) mostly referred to the journey that the main character embarks on in the play and included responses such as, *"You travelled from the forest through places and then you ended up here"*. Others talked about how

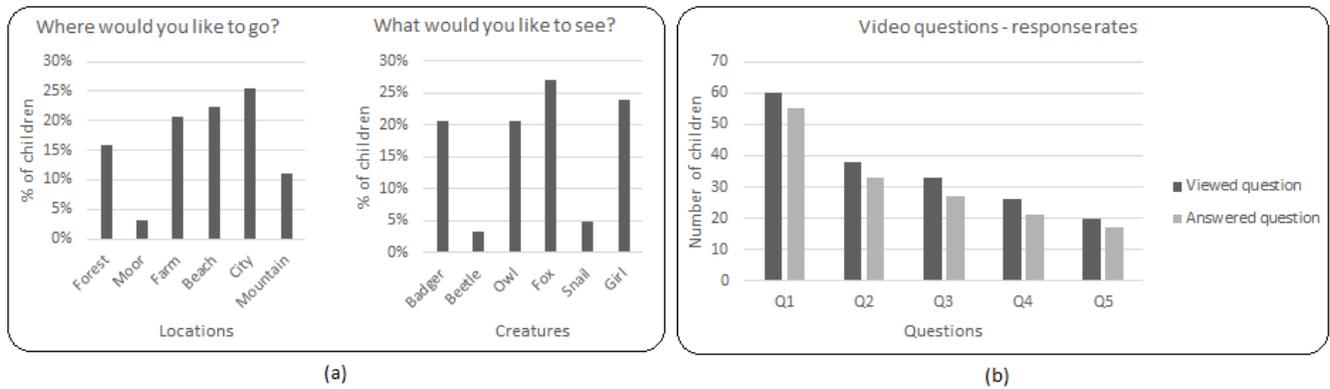


Figure 2. (a) the responses to questions at boxes 3 and 4; (b) the response rates to the five video questions.

he was searching for his friends and family and some talked about specific things he had done such as, "You went on a boat", "You ate a stick".

For question four ("Some things on my journey made me happy. What made me happy?"), the answers were widely varied capturing a range of happy moments from the play that impacted on the children. Responses included; "Seeing all the animals and seeing nature", "Swimming in the boat", "When you got back to the city", "When you saw the fox", "The glowing things".

Finally, the answers to question five ("Some things on my journey made me sad. What made me sad?") often mentioned something about the It being lost, trying to find friends and family or wanting to go home. However, several children mentioned specific, but poignant, experiences in the play that also upset the It. These included; "A girl running off", "When there was a machine cutting all the trees down", "It made you sad, um, when the tractor was smashing the trees down I think, yeah".

Interactions with the Video Box

Overall the children enjoyed interacting with box 5 and many were observed smiling broadly when the character first appeared on screen and greeted them. For example, on one occasion a little girl turned to her friend behind her and they both started giggling. It was clear that the children recognized the character on screen from the play, and everyone listened intently to his questions, even if they didn't answer. Several of the younger children (aged five and below) required some prompting, from researchers, to say their answers into the microphone. However, the 'listening gesture' performed by the It at the end of each question had the desired effect and the majority of children understood the cue to respond. On one occasion a little girl responded before the It pretended to listen, but repeated her answer again on cue, as though to make sure he heard her.

Three children were observed striking up a dialog with the character on screen and role-playing as if to make the box feel more interactive than it really was. For example, when the character said; "Hey...it's so good to see you again", one child responded; "Thanks, it's nice to see you too", before

listening to and answering the follow-up question. On another occasion, when the character said; "You were watching me on my journey", a young girl responded; "Yes, yes I was". Conversational body language and gestures were also common. They included the waving of hands, shrugging of shoulders and crossing arms with a finger on the cheek in a 'thinking' pose.

DISCUSSION

Based on our results, we discuss how effective SmallTalk was at evaluating the theater performance and the children's memories and experiences of it. Additionally, we use our results to reflect on the initial design goals, and what design implications were identified that could be applied more generally when designing future tangible devices to gather feedback from children.

Evaluation of the Theater Performance

Overall, the tangible method was able to elicit a range of reflections about the performance the children had just experienced. The answers to boxes 1 and 2 showed that the majority of children would like to return to the theater in the future having enjoyed their current experience. Interestingly, the most popular answer to the question at box 3, about where the children would like to go, was 'the city', beating what many would consider to be more idyllic settings such as the beach and the mountains. This reflects the positive emotions that the It had at the city - it being the place where he was most excited and happy as he was almost home. The moor was the least popular location and was also the place where the It was most desolate after he had just gotten lost. At box 4 there was a similar mapping of most popular creatures to the emotions of the It when he saw them on his journey. The fox (most popular) appeared in the city when he was almost home whereas the beetle (least popular) appeared when he was on the moor, alone and sad.

The video box questions asked about the actions of the It and what affected his emotions during the performance as well as how the child felt when watching the performance. The response rates were high across all questions showing that the children were able to reflect on their own feelings as well as empathizing with the feelings of the It. The responses also

covered a wide range of events from more broad concepts (e.g., losing his family) to quite specific moments, such as when a girl he meets runs away, or when he tries to eat a stick. This suggests how engrossed children were in the performance and specific moments that evoked emotion in themselves such as sadness for the It or laughter when he acted silly. As such, our results show how evaluation data could be collected using the tangible boxes – with data showing the extent to which the children connected with the performance at an emotional level, while also being able to reflect on specific events that happened and their consequence for the It.

Design Goals and Implications

SmallTalk achieved the aim of providing an *engaging and physical experience* for children, while providing feedback about something they had just experienced. The children were engrossed in their interaction with the boxes and eagerly went from one to the next. A variety of different physical inputs such as buttons, dials and spinners were used, as on other tangible feedback systems (not specifically designed for children [10]) such variety helped to maintain interest with the device and encouraged sustained engagement. However, in this case it was confusing for some, primarily due to their unfamiliarity with the affordances of physical buttons, dials and spinners. In several cases we observed children trying to apply touch-screen gestures, which highlights the fact that many are growing up with tablets and smartphones all around them. As such, they can be familiar with modern touch-screen gestures and less familiar with physical buttons and dials. We propose that on future systems for young children, greater continuity of physical interactions (e.g. only using buttons throughout) would help reduce the cognitive overheads of recognizing and understanding how to use multiple physical inputs.

Each deployment of SmallTalk was successfully run in the theater space immediately after the performance and therefore supported *in situ interaction* and captured opinions that were more ‘in the moment’. By deploying in this context, the atmosphere inside the cavernous theater space came into play during each evaluation. When the evaluation group was loud and busy, this created a buzzing atmosphere inside the theater and a more relaxed context for the children to interact with the boxes. However, when the group was quiet and timid (as the youngest school group were), this created a still, tense atmosphere inside the theater and a slightly oppressive context for the children interacting with the boxes. This raises an important design implication for future in situ deployments, in that the wider deployment context can potentially undermine the device design if a less than positive atmosphere occurs in that context. Therefore when deploying tangible devices in situ, thought is needed as to the wider deployment environment and determining the appropriate atmosphere to surround those interacting.

Another aim of SmallTalk was to support *minimal mediation*. This was achieved through the simplistic design

of the boxes and strong audio support throughout. However, we observed a threshold age of six, below which children tended to need guidance. This was related to familiarity with the affordances of physical inputs, but also to a greater tendency for younger children to look for adult confirmation that what they were doing was correct.

SmallTalk also supported *time-efficient evaluation* with each evaluation group, consisting of 9 children on average, taking around 10 minutes to complete all 5 boxes. The majority of time was spent at the video box that took much longer to complete than the other 4 boxes with multiple-choice questions. However, the video box arguably captured the richest and most interesting data. As such, there is a trade-off between the time taken to complete the interaction and the richness of data that can be gathered from the interaction. For a given application, if richer data is desirable but there are time constraints, we propose fewer open-ended questions over many multiple-choice ones.

SmallTalk was very successful at supporting *open-ended feedback* from children through talking to the character appearing in the video box. Compared to the tangible questionnaire VoxBox that used a telephone handset for open-ended questions [10], the video box used a familiar face and gestural cues (listening gesture), which had the desired effect of encouraging spoken responses from young children. A design consideration is the positioning of open-ended questions in relation to multiple-choice ones. From our results we saw that answers to previous multiple-choice questions can affect answers at the video box, all be it to a small degree. One could experiment with putting open-ended questions first to further mitigate any interplay. However, we suggest that this could introduce other issues. Our observations showed that the children became more relaxed and confident as they moved down the line of boxes and understood how to use them. If we had placed the video box at the start of this line the children would have been required to interact with the most outwardly expressive box (speaking their answer) while also being at their most anxious and least confident. As such we propose that several simple, multiple-choice questions are desirable as the first point of contact with a tangible feedback system to allow familiarity and confidence to grow before introducing more expressive, open-ended questions.

CONCLUSION

It can be difficult obtaining feedback from young children about something they have just experienced such as a production, an event, a lesson, a film or a product. While adapting adult-based usability methods to be more child-centric (such as asking them to draw and using emoticons on rating scales) has made some in-roads, they require much mediation and are limited as to the extent to which they can get children to reflect, remember and think more deeply in situ. An alternative approach is to use an embodied tangible system that asks children multiple questions that they answer through moving physical dials, sliders and pressing buttons

and through answering more open-ended questions by talking to a protagonist on a screen. Based on an in situ study of such a system, called SmallTalk, we demonstrated how such an embodied physical system could be used to good effect to evaluate an immersive theater performance, showing how the performance evoked emotional connections among the young audience and empathy with a lead character. To get the most out of children, however, also requires considering the atmosphere, the level of guidance and the setting within which to situate such tangible feedback systems. Children will provide feedback but need to feel comfortable, excited, and should themselves receive a level of reassurance when opening up their thoughts and feelings.

ACKNOWLEDGEMENTS

This research was funded by the EPSRC and the Intel Collaborative Research Institute: Cities. We would like to thank Fevered Sleep production company, Alison Clark from the Open University, and all the schools and children who participated in the study.

REFERENCES

1. Alissa N. Antle. 2007. The CTI framework: informing the design of tangible systems for children. In *Proc. TEI 2007*, 195-202.
2. Alissa N. Antle. 2012. Knowledge gaps in hands-on tangible interaction research. In *Proc. 2012*, 233-240.
3. Alissa N. Antle, et al. 2009. Hands on what?: comparing children's mouse-based and tangible-based interaction. In *Proc. IDC 2009*, 80-88.
4. Ester Baauw and Panos Markopoulos. 2004. A comparison of think-aloud and post-task interview for usability testing with children. In *Proc. IDC 2004*, 115-116.
5. Saskia Bakker, Alissa N. Antle, and Elise van den Hoven. 2012. Embodied metaphors in tangible interaction design. *Personal Ubiquitous Comput.* 16, 4, 433-449.
6. Wolmet Barendregt, Mathilde M. Bekker, and Ester Baauw. 2008. Development and evaluation of the problem identification picture cards method. *Cogn. Technol. Work* 10, 2, 95-105.
7. Afke Donker and Pieter Reitsma. 2004. Usability testing with young children. In *Proc. IDC 2004*, 43-48.
8. Paul Dourish. 2001. *Where the action is: The foundations of embodied interaction*. MIT Press, Cambridge, UK.
9. Helen Edwards and Rachel Benedyk. 2007. A comparison of usability evaluation methods for child participants in a school setting. In *Proc. IDC 2007*, 9-16.
10. Connie Golsteijn, et al. 2015. VoxBox: A Tangible Machine that Gathers Opinions from the Public at Events. In *Proc. TEI '15*, 201-208.
11. Libby Hanna, Denise Neopolitan, and Kirsten Ridsen. 2004. Evaluating computer game concepts with children. In *Proc. IDC 2004*, 49-56.
12. Libby Hanna, Kirsten Ridsen, and Kirsten Alexander. 1997. Guidelines for usability testing with children. *interactions* 4, 5, 9-14.
13. Bart Hengeveld, et al. 2008. Let me actuate you. In *Proc. TEI 2008*, 159-166.
14. Eva Hornecker and Jacob Buur. 2006. Getting a grip on tangible interaction: A framework on physical space and social interaction. In *Proc. CHI 2006*, 437-446.
15. Ilse E. H. van Kesteren, et al. 2003. Assessing usability evaluation methods on their effectiveness to elicit verbal comments from children subjects. In *Proc. IDC 2003*, 41-49.
16. Pranee Liamputtong and Douglas Ezzy. 2005. *Qualitative research methods*. 2nd ed. Oxford University Press, Oxford, UK.
17. Paul Marshall. 2007. Do tangible interfaces enhance learning? In *Proc. TEI 2007*, 163-170.
18. Lorna McKnight and Janet C. Read. 2011. PLU-E: a proposed framework for planning and conducting evaluation studies with children. In *Proc. BCS HCI 2011*, 126-131.
19. Emma Nicol and Eva Hornecker. 2012. Using children's drawings to elicit feedback on interactive museum prototypes. In *Proc. IDC 2012*, 276-279.
20. Janet C. Read and Stuart MacFarlane. 2006. Using the fun toolkit and other survey methods to gather opinions in child computer interaction. In *Proc. IDC 2006*, 81-88.
21. Janet C. Read, Stuart MacFarlane, and Chris Casey. 2002. Endurability, engagement and expectations: Measuring children's fun. *Interaction Design and Children* 2, 1-23.
22. Theodosios Sapounidis et al. 2013. Tangible versus graphical user interfaces for robot programming: exploring cross-age children's preferences. *Personal Ubiquitous Comput.* 17, 8, 1775-1786.
23. Lesley Xie, Alissa N. Antle, and Nima Motamedi. 2008. Are tangibles more fun?: comparing children's enjoyment and engagement using physical, graphical and tangible user interfaces. In *Proc. TEI 2008*, 191-198.
24. Bieke Zaman, et al. 2009. Tangibles for children, the challenges. In *Proc. CHI 2009 Ext. Abstr.*, 4729-4732.
25. Bieke Zaman, et al. 2012. Editorial: the evolving field of tangible interaction for children: the challenge of empirical validation. *Personal Ubiquitous Comput.* 16, 4, 367-378.