

Using Eye Tracking to Understand the Fidelity Effect when Evaluating Low-Fidelity Prototypes with Children

Gavin Sim
University of Central Lancashire
Preston, UK
grsim@uclan.ac.uk

Janet C Read
University of Central Lancashire
Preston, UK
jcread@uclan.ac.uk

This study used eye tracking glasses to understand how children explore low-fidelity paper prototypes in the context of user experience studies. Twenty seven children aged between 9 and 11 participated in the study examining either a colour or black and white prototype of a mobile game. The research question being explored was whether the aesthetic refinement, either wireframe or high-resolution colour images, would affect children's self-report and if so what could be learned from knowing where children looked when exploring the prototypes. The results showed that the aesthetic refinement had little influence over the children's overall ratings of the game. The eye tracking data demonstrated that most of the children focused on both the visuals and text on the pages of the prototype. However, there were a higher number of fixations recorded in the wireframe prototype suggesting this may have been more cognitively demanding. This paper contributes to understanding of fidelity effects when evaluating low-fidelity prototypes and shows how eye tracking technology can help inform HCI methodologies.

Children. Prototypes. Eye Tracking. Fidelity Effect. Games.

1. INTRODUCTION

Eye tracking technology collects eye movement data that relates to the visual stimuli that are processed by the brain. Within the context of HCI these stimuli may relate to the user interacting with visual elements of a prototype or text describing a game or app. To capture eye tracking data from users, eye tracking hardware and software are required. Most of these systems use infrared light that is reflected off the cornea in the eye and captured by cameras (corneal reflection-based eye trackers). Current technology enables studies to be easily performed with users in natural environments where they can be looking at almost anything (Potvin Kent, et al. 2019). Eye tracking has been used with children to capture visual behaviour in many contexts, including programming with children (Sharma, et al. 2019) and hand eye co-ordination in video games (Chen and Tsai 2015). Notably, in HCI, we can use eye tracking to confirm that children have been looking at things that we present to them, thus helping us build confidence in our assumptions about how children interact with artefacts and about the basis on which they report back their opinions.

When evaluating artefacts with children, it is common to present a child with a prototype (Bertou and Shahid 2014, Hershman, et al. 2018). Prototypes are used for several reasons for example, the evaluation of design ideas (Hanna, et al. 2004), the exploration of ideas, and as props to assist with communication as part of a development process (Levi and Conrad 1996, Reilly et al. 2005). To evaluate design ideas, a prototype can be used to get feedback to inform future iterations and identify potential usability problems (Wiklund et al. 1992, Virzi, et al. 1996). Software prototypes can take many forms and be described in different ways. One categorization that is used is in terms of fidelity. Low-fidelity describes prototypes that are typically made from a material that is different from the final product, such as paper sketches or cardboard (Rudd et al. 1996). These can be used to explore concepts like navigation or organisation of material (Buxton 2007). High-fidelity prototypes usually offer a level of functional interactivity using materials you would expect in the final product (Rudd et al. 1996). For example, a prototype of an Android app might be simulated on a touch-based screen or a phone with scaled back functionality. It is important to note that fidelity can be varied in several dimensions: degree

of functionality, similarity of interaction, breadth of features and aesthetic refinement (Virzi et al. 1996). The work presented in this paper is concerned with fidelity effect in the context of aesthetic refinement.

With adults, the fidelity of prototypes has been shown to influence the results obtained in evaluation studies (Rudd, et al. 1996, Hoggenmüller et al. 2021). When studying fidelity effects it can be difficult to determine to what extent findings from evaluations align to the concept being presented or to the characteristics of the prototype (Lim et al. 2006). In one study with adults, using low fidelity (low visuals) prototypes, authors claimed the lack of refined graphics could bias evaluators against the products (Kohler et al. 2012). There can also be contradictory findings; in a study with children evaluating a game concept using different prototype fidelities, there have been similar results (Sim et al. 2013) and significant differences (Sim and Cassidy 2013) suggesting things happening during the evaluation that are not well understood.

This paper uses eye tracking technology to better understand how children interact with prototypes of different fidelities by comparing children viewing the same game concept but with different visual (aesthetic) refinements of the prototype. This builds on prior work on understanding the fidelity effect when evaluating prototypes with children (Sim et al. 2016). Measured through surveys and eye tracking technology two research questions are identified:

- (i) RQ1: Does presenting a prototype to children in different levels of aesthetic refinement affect their rating of a game concept?
- (ii) RQ2: What insights can eye tracking technology provide towards understanding how children interact with a low-fidelity prototype?

2. BACKGROUND AND RELATED WORK

2.1 Eye Tracking

Eye tracking technology is composed of hardware and software. The hardware is camera technology that is either fixed to an object being viewed, like a car dashboard or screen, or worn by the viewer, as glasses or headsets. In recent years the hardware has become very versatile and enables studies to be performed away from a laboratory and within natural environments (Potvin Kent et al. 2019). This is especially well suited for use with children. The software processes the data which is gathered by the cameras. This is eye movement data that relates to the visual stimuli processed by the brain. In the context of our work the stimuli may relate to the text or images that the child views on a low-fidelity prototype. Researchers are often concerned

with certain areas of stimuli; this may be a line of text or a word (Kornev et al. 2018) or could relate to making eye contact with a peer (Kleberg et al. 2020).

Eye gaze data are two dimensional Cartesian coordinates (x, y) that represent a point on the screen display or a point in the world. Eye tracking software captures the time and duration of these events which are called 'fixations'. These events need to be interpreted within the context of a study; depending on the activity the user is performing; a high number of fixations can negatively correlate with search efficiency (Bilal and Gwizdka 2016) or indicate difficulties in interpreting information (Al-Wabil et al. 2010). The software also captures movements of the eyes from one point to another, referred to as 'saccades'. Together, fixations and saccades can infer something about human visual processing or behaviour. Simple fixation-based metrics include fixation count, total fixation duration and average fixation duration. Fixation count relates to the total number of times a user focussed their attention, and the average fixation duration (with standard deviation) provides the typical amount of time that the user fixated or focused their attention. Fixation frequency can be used to infer how often the user fixated per second (Hz) or per minute, for example in HCI, a high fixation frequency might correspond with the user being confused or might suggest a lack of visual hierarchy in the user interface (Torney et al. 2018).

Eye tracking studies with children generally focus on fixation data (Sim and Bond 2021). Much of the early work on eye tracking with children focussed on understanding reading (Marcel and Patricia 1980) to understand differences between populations, for example children with numeric processing deficits (Moeller et al. 2009) or autism (Su et al. 2018). Within reading studies, children have been shown to have longer fixations, shorter saccades, more re-fixations and many regressions when compared with adults and this has been useful to understand beginner readers (Huestegge et al. 2009). Fixation-based metrics can be useful to explore content, an important HCI design concern, where they can be used to compare and benchmark areas of interest (AOIs). AOIs are predefined areas (e.g., the navigation bar) that are identified by the researcher. Using eye tracking software, the researcher can mark out an AOI and then the software can report visits to these AOIs as counts, dwell times or frequencies. As with general fixations, interpretation of such data has to be done with caution: if certain AOIs have a high visit count, it can be that perhaps that AOI is valuable, or it could be an area of interest in a decision-making experiment (Currie et al. 2017).

Calibration of the eye tracker is an important part of any study; during the study, system-inherent drift may occur resulting in the recording of eye tracking movement becoming inaccurate. With small AOIs

this drift is problematic and requires re-calibration during a study (Holmqvist and Andersson 2017). Data can also be lost through the participants' physical movement during the study, for example moving towards the screen (Tragant Mestres and Pellicer-Sánchez 2019). One key challenge for any experimental study using eye tracking with children is to explore and retain the accuracy of the data.

Eye gaze data can be presented in a tabular form or can be visualised as heat maps (akin to contour maps) which are basically attention maps showing which areas had the least and most fixations (or fixation durations, saccades, or visits). Researchers can also visually inspect videos of the users' attention, generated by the software, where the video is superimposed with eye gaze fixations, allowing for qualitative analysis of human behaviour.

Although eye tracking has been used with children to evaluate prototypes of applications including augmented reality (Gomes et al. 2012) and robot interaction (Othman and Mohsin 2017), there have been no studies to date examining children's interactions with low-fidelity prototypes. Given the similarity in structure of combining images and text, eye tracking studies within media, including comic books and advertisements, may afford insights that may apply to low-fidelity prototypes. In a study comparing comprehension of an image based comic book (no text) children had more, and longer, fixations than adults, suggesting more cognitive effort was needed to comprehend the stories (Martín-Arnal et al. 2019). Eye tracking was used with 36 children aged 10-14 to examine the effectiveness of geographical comics as a learning aid for children and showed the highest concentration of fixations were on the text elements of the comic book, (von Reumont and Budke 2020). These studies have highlighted that children's attention can be captured through eye tracking. In the (Martín-Arnal et al. 2019) study, the comic book was presented on a computer screen and data captured using a Tobii-x120, it is interesting to consider if children would interact with a physical comic book differently and new eye-tracking techniques, like glasses will allow this sort of inquiry.

Eye tracking studies examining children's viewing behaviour with adverts have included food, tobacco, and alcohol. In a study looking at unhealthy food advertisements (Velazquez and Pasch 2014) the children's food preference correlated with the length of time and number of times a child looked at an item. With children viewing tobacco adverts, the warning message was only viewed for 8% of the total time and in 44% of cases the children did not view the message at all (Fischer et al. 1989). It has been suggested that text towards the bottom of an advert is not considered important and thus was ignored by the children; or it could be that the text was too small. Researchers looking at alcohol

adverts (Thomsen and Fulton 2007), demonstrated that when the warning message was incorporated more prominently, children spent more time attending to it. When text was small and at the bottom of the page only 7% of the total viewing time was spent looking at it. If these findings translate to traditional storyboards where text is often at the bottom of a screen, children may miss this and not understand the information the designer is trying to convey. It is important to understand how children attend to different areas within a storyboard if results of an evaluation are used to improve future products.

2.2 Prototype Fidelity

Within the HCI literature, researchers have discussed the merits of different prototype fidelities (Rudd et al. 1996, Snyder 2003), but there are contradictions in the literature and many claims have not been empirically validated. For example (Snyder 2003) suggested that 'When something appears to be finished, minor flaws stand out and will catch the users' attention'. This was contradicted in a study looking into the effect of visuals on usability studies using game prototypes that empirically compared two different form factors, low and high fidelity, and showed that most of the participants paid very little attention to the visuals (Kohler et al. 2012). When the focus of a study is on capturing user's emotional responses, developers tend to use higher-fidelity prototypes characterized by considerable aesthetic refinement (Sauer and Sonderegger 2009). When capturing usability problems and emotional responses, understanding how visuals and text interact in such studies is important.

There have been several comparative research studies of prototypes at different fidelity levels (Wiklund et al. 1992, Sefelin et al. 2003, Lim et al. 2006, Hoggenmüller et al. 2021), however results are inconclusive. In a study examining context-based interfaces (Hoggenmüller et al. 2021) compared three prototype representations and the quantitative results show that while the real-world VR representation resulted in a higher sense of presence, there were no significant differences in the user experience or trust. There have been studies showing that, with low-fidelity prototypes, results can be gathered that are equivalent to those gained from evaluating fully operational products; other studies report the additional benefits of higher fidelity prototypes e.g., (Sauer et al. 2008). Seflin *et al.* (Sefelin et al. 2003) investigated whether participants differed in their willingness to criticize a system with a paper-based low-fidelity prototype, as opposed to a computer based one; there was no difference in the number of criticisms, but the users preferred the computer prototype.

There are a limited number of research studies that have looked at the fidelity effect when evaluating prototypes with children. One study invited 16-17

year olds to compare low and high fidelity prototypes for tabletop surfaces (Derboven et al. 2010), the findings cautioned against generalizing high-level user interactions from low-fidelity prototypes as the interaction differed; for example it was feasible to layer information on top of each other in a 3D space and this was not feasible in the 2D space. Difficulties in simulating the interaction have also been reported in (Kohler et al. 2012) with one participant struggling to understand the concept of an accelerometer. In a study by (Sim et al. 2013) three low-fidelity prototypes were evaluated with results showing very little difference in reported user experience between the three and very few usability problems unique to a specific prototype. Physical form factor was also examined to determine whether a prototype presented on a iPad differed to that on paper (Sim et al. 2016), interestingly children rated aesthetics higher on the paper version compared to the iPad, despite the graphics being identical. This suggests that ratings of aesthetics may be influenced by the form in which the prototype is presented.

This paper aims to use eye tracking software to examine the fidelity effect when children are presented with prototypes with different levels of visual refinement; it additionally aims to gain insights into any differences in their visual attention.

3 METHOD

The study adopted a between subject design methodology to evaluate the user experience of a prototype game in two different forms: a wireframe prototype and a higher fidelity colour prototype. The prototype was in the form of an annotated storyboard, and this was presented to the children in a paper booklet consisting of 10 pages.

3.1 Participants

The participants were 27 children from a UK primary school, aged 9-11 years old and from a single school class. The study took place in the school over a 6-week period so all the children in the class could participate. The 1st week was an orientation exercise to help children understand the technology.

3.2 Apparatus

Two prototypes of a single game were needed: one low fidelity and one higher fidelity. The Splode game (see Figure 1) was selected as this had been used in previous work examining the fidelity effect (Sim et al. 2013) and was popular several years ago so had probably not been seen by the participating children.

To create the low fidelity storyboard, the game was reverse engineered by capturing screen grabs of the game on the iPad and then tracing these using Adobe Illustrator, see Figure 2. This approach has been previously used in studies examining the fidelity effect (Sim et al. 2013). Reverse engineering games is not how conventional prototyping occurs,

but it does isolate fidelity from maturity of design which is important to reduce confounds.

The high-fidelity version showed the same screens that were used to create the low-fidelity version, see Figure 1. In both cases ten screens, representing the same interaction points in the game, were shown in the form of a storyboard that conveyed the game mechanic to the children. To highlight movement and interaction, arrows and hands were added to the storyboard in both versions. Text was added underneath the images to explain the game concept and the mechanics.

Both versions were presented to the children as ten sheets of A4, stapled together in the left-hand corner and intended to be looked at as one would a book, turning over the pages as the child walked through the game idea.

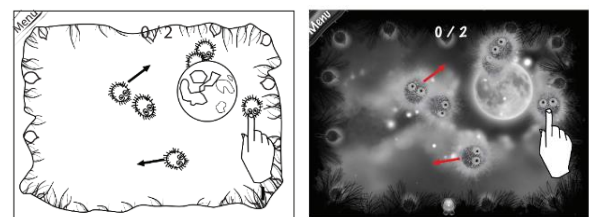


Figure 1: Wireframe (left) and Colour Version of the Game (right)

3.3 UX Measures

To measure user experience an adaptation of the Fun Toolkit (Read, et al. 2002) using the Smileyometer and Again Again table was used. This followed the same protocol as used within (Sim et al. 2013). The Smileyometer, is a visual analogue scale coded using a 5-point scale, see Figure 3.

The Smileyometer is presented before children interact with technology to measure their expectations, and then afterwards when it is assumed that the child is reporting their experience. In the context of the prototypes, expectations were captured following a look at the 1st screen and the text presented. A Smileyometer was presented on paper and the child chose a face and was also given space to say why they had chosen that option.

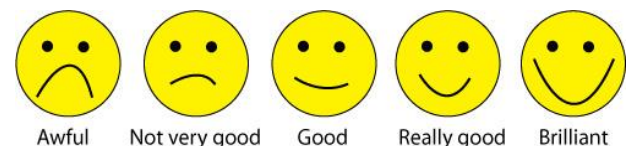


Figure 3: The Smileyometer

Once the child had worked through all the pages and gathered a firmer understanding of the game, experience was measured by asking children to rate the game idea, the graphics, and their overall experience; the wording of these questions can be seen in Tables 1 and 2 – in each case the child responded using a Smileyometer. The children

were asked to imagine that the prototype was transformed into a playable game on a tablet prior to answering the questions. Thus, when asking about the graphics they were encouraged to imagine these in a fully functioning tablet game.

Finally, the Again Again table was used to establish whether the children would download the game based on the storyboard presented. They answered 'yes', 'maybe' or 'no' to the following questions:

- If the game were free would you download it from the app store?
- If the game were 99p would you download it from the app store?

On the day, children saw two documents; one was the ten-page storyboard in a wireframe or colour version and the other was the data capture form with the Smileyometers, and the Again Again table.

3.4 Eye Tracking hardware and software

Tobii Glass 2 eye trackers were used to capture eye movement; these have previously been used with children in studies of programming (Giannakos et al. 2020, Papavaslopoulou et al. 2020). The sampling rate for the eye-tracking glasses was 60 Hz and the average accuracy was 0.5°. The glasses were paired with a laptop to capture the data and calibrate the eye tracker for each child.

3.5 Procedure

The study was conducted in the school. Before the study started, the first author went into the school to show the class the eye tracking technology, explain how it worked and to let the children ask any questions. This ensured that children understood the study, the technology, the data being captured and could make an informed decision about participation. Following this initial orientation session, the study was conducted over a five-week period during which time a single researcher went into the school, each Wednesday to perform the study. The physical location of where the study changed over this period, including the sports hall, the headteachers office and a corridor. Each child only attended on one day, the reason the study took five weeks was that there were 27 children and the study had to be done with one child at a time.

The children were sent from a single class by the teacher individually to where the study was taking place. Each child was greeted by the researcher, who again explained the purpose of the study and sought assent. The child was then presented with the data capture form and the storyboard turned over so they could not see the 1st screen. The Tobii eye tracking glasses were on the desk and the child was asked to put them on themselves and fix in place with the fastener. Occasionally a child would need some assistance from the researcher, but most

did not. Once the glasses were on, the child was shown the computer screen which showed them their eyes and the video output, and the researcher explained the calibration process. The glasses were then calibrated. There were some difficulties in calibrating the glasses for a few children, for example one child's left eye could not be detected. It was important not to cause any anxiety to the children and in these instances the child was reassured by the researcher and asked to continue the study anyway. To ensure that the study was as natural as possible the decision was made not to re-calibrate after every page of the storyboard. This would have increased accuracy and mitigated data loss through movement of the glasses during the study but would have increased the length of the study and potentially stressed the children and caused a loss of focus on the task.

Once calibration was completed, recording started, and the child was asked to turn over the storyboard. From this point on, the researcher was positioned to the side of the child and the computer screen was turned away from the child to avoid distraction. The children were asked to look at the first page of the storyboard, taking as long as they wanted to read the information relating to the description of the game as described on the app store. Once they had finished examining the 1st page, they answered the first question of the Smileyometer regarding their expectations and commented on their selection.

The children then went through the remaining nine pages of the storyboard, and then answered the remaining questions on the survey sheet. Once that was completed, they were shown the video capture of their eye movement and assent was sought again for using the video and survey data. The whole procedure lasted about 20 - 25 minutes per child. One child did not have parental consent to participate but still had a turn using the technology, so they did not feel left out; no data was captured.

3.6 Analysis

3.6.1 Survey Data

Of the 27 children who participated with consent, all completed all the questions in the survey. The questions Smileyometer questions were coded in an ordinal way 1 - 5, where 5 represented Brilliant and 1 Awful. For the Again Again table, Yes was coded as 2, Maybe as 1 and No as 0. In line with other studies using this scale, arithmetic averages have been taken of these scores (Read 2012).

3.6.2 Eye Tracking

The eye tracking data was analysed using Tobii Pro Labs. Two areas of interest (AOI) were established for each page, one for the text and the other for the image. This enabled a comparison of the fixation data between the two areas as we had hypothesised that some children might not read the text.

Each video was examined to establish whether sufficient gaze samples had been captured. This was essential as data was lost during the study for various reasons. For example, one child was reading underneath the glasses and therefore no gaze data was captured. In total 7 recordings were discarded as the sample rate was insufficient. For the remaining 20 participants the fixation data was manually plotted to pages of the storyboard as sometimes the child would look at the researcher, especially between turns of the pages, and so that fixation data had to be discarded. Only fixations on areas of interest were considered for analysis.

4 RESULTS

4.1 Survey data

Table 1. shows the results for the first question of the Smileyometer, after the children had viewed the 1st page of the storyboard prototype.

A Mann-Whitney U test revealed no significant difference between the two groups wireframe ($M_d = 4.00$) and colour ($M_d = 4.00$) for $U = 49.5$, $z = -1.94$, $p = 0.316$.

Table 1: Mean scores and standard deviation for the first question relating to expectations.

	Wireframe		Colour	
	mean	SD	Mean	SD
Q1 What do you think the game will be like?	3.45	.69	3.83	.94

The results of the questions that were asked after interaction with the prototype are shown in table 2.

Table 2: Mean scores and standard deviation for the three questions following interaction

	Wireframe		Colour	
	mean	SD	Mean	SD
Q2 The idea of the game is	3.55	.82	3.67	.65
Q3 How do you think the (real) game will look	3.36	.92	3.67	.98
Q4 Overall the game is	3.90	1.22	4.25	.86

A Mann-Whitney U test revealed no significant difference for Q2 to Q4. Based on an aggregation of the data from the two groups, a Wilcoxon test revealed a significant difference between the Smileyometer results after they had viewed the entire storyboard for Q4 compared to Q1 which was completed after they had seen the 1st screen, $z = -1.995$, $p = 0.046$, with a moderate effect size $r = .41$.

The results from the Again Again table as to whether the child would download the game if it were free or 99p are presented in table 3.

Table 3: Frequency of responses to whether the child would download the game if it was free or 99p

	Wireframe			Colour		
	Yes	Maybe	No	Yes	Maybe	No
free	6	3	2	7	5	0
0.99p	4	1	6	3	6	3

It appears that children were less willing to download the game if they had seen the wireframe prototype.

4.2 Eye Tracking Results

RQ1 aimed to establish whether there was any difference in the way children looked at the pages of the prototype dependent on the prototype version.

Table 4: Mean number of fixations for the two AOIs on the 1st screen of the prototypes

	Wireframe		Colour	
	mean	SD	Mean	SD
Text	57.30	33.61	49.60	13.11
Image	16.70	19.02	15.20	8.63

Although on the 1st screen the mean number of fixations was higher for the wireframe compared to the coloured version, see table 4, a t-test revealed no significant difference for the two areas of interest.

Figure 4 shows a heat map generated using absolute fixation count for the two prototype versions of the 1st page. The viewing behaviour of the children appears similar for the Image AOI, whilst there is a larger clustering of fixations in the lower-middle of the text AOI in the wireframe prototype.

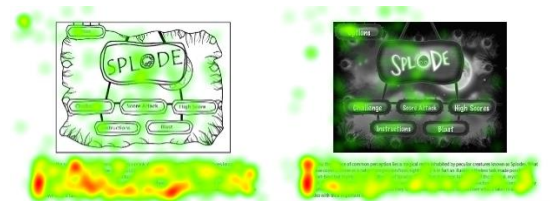


Figure 4: Heat map of the 1st page of the prototype left image wireframe and right the colour

Table 5 below shows the means for the total number of fixations for the areas of interest for the 10 pages of the prototype. The 10 participants who received the wireframe prototype ($M = 165.10$, $SD = 64.67$) compared to the 10 participants viewing the colour prototype ($M = 150.60$, $SD = 36.32$) demonstrated significantly more fixations when focusing on the text, $t(18) = 2.058$, $p = .027$. There was also a significant difference on the number of fixations based on viewing the image, $t(18) = .618$, $p = .008$. Given the higher number of fixations overall, it might

be deduced that the wireframe prototype was perhaps more cognitively demanding.

Table 5: Means of Total Fixations for the two AOIs in the Wireframe and Colour Prototype

Group	Wireframe		Colour	
	Mean	SD	Mean	SD
Text	165.10	64.67	150.60	36.32
Image	171.50	107.57	98.10	33.94

When looking at individual pages, the mean number of fixations per AOI was higher for the wireframe across all 10 pages. Page nine demonstrated the largest difference between the two groups with the wireframe ($M = 35.5$, $SD = 17.3$) compared to the colour ($M = 5.4$, $SD = 3.92$) prototype. Differences in the number of fixations could have been attributed to a difference in the time the children interacted with the prototypes, but a t-test revealed no significant difference in the overall time spent viewing the prototypes between the two groups. For the 10 participants who received the wireframe prototype ($M = 338.76$, $SD = 24.56$) compared to the 10 participants viewing the colour prototype ($M = 285.57$, $SD = 67.05$), $t(18) = 1.639$, $p = .496$.

It appears, from this data, that the children were engaging with the prototype across all 10 pages. Within the image AOI fixations appear to decline after the first few pages and then increase again towards the end, for the text AOI, the highest fixations were on page 8 of the wireframe and page 2 on the colour prototype.

5 DISCUSSION

Our study aimed to understand the impact that fidelity, in the form of visual refinement, would have when evaluating user experience and gain insights into how children interact with a low-fidelity prototype. To explore this, an iPad game was reverse engineered to construct two low fidelity prototypes that varied in aesthetic refinement and a between subjects eye tracking study was carried out with schoolchildren aged 9 - 11.

The first question *RQ1: Does presenting a prototype to children in different levels of aesthetic refinement affect their rating of a game concept?* aimed to establish whether children rated the game differently based upon the fidelity of the prototype. There was no significant difference on any of the constructs examined. The presentation of the game in wireframe yielded similar results to the prototype of higher visual refinement. This suggests that it may be possible to have confidence in the results of an evaluation based on a wireframe prototype and this is in line with other studies (Sim et al. 2013). However, care needs to be taken as other studies have potentially contradictory results with respect to aesthetic refinement. In a study comparing the impact physical form factor of a prototype has on

children, children rated the graphics lower (despite these being identical) on the iPad compared to the paper version (Sim et al. 2016). In another study Sauer and Sonderegger (2009) examined aesthetics in prototypes with adults who appeared to compensate for deficiencies in aesthetic design by overrating the aesthetic qualities of reduced fidelity prototypes. More work is required to understand what factors might influence and confound the results when evaluating prototypes with children.

Children rated the game higher at the end compared to their initial expectations. This is useful for researchers as, given the eye tracking data we can have some confidence that the ratings were aligned to experience rather than just guessed at.

Interestingly the Again Again table showed a higher percentage of children not wanting to download the game based on having seen the wireframe version in comparison to the colour prototype. This is important as intention to play / purchase is an indicator that the children enjoyed the game. The Again Again table has been suggested as a more objective tool for children in doing evaluations as it de-personalises the study by asking the children 'would you....?' This liberates the child from judging something the 'researcher' has brought for evaluation. For this reason, we hypothesise that giving children a more visually appealing prototype might positively influence their intention to play.

Obtaining quality data from children using the eye tracking glasses was challenging. From the 27 children who participated only 20 useable data sets were collected and we were very lucky that the twenty were evenly distributed across the two conditions! Data was lost for a variety of reasons including the children moving the glasses during the study, light reflecting off bright surfaces and children looking underneath the glasses. These challenges are not unique to this study or to glasses-based technology. Data has previously been reported lost when using screen-based hardware due to fidgety children (Gossen, et al. 2014) and calibration issues (Krejtz et al. 2012). The quality of the data could have been improved if the eye tracker was calibrated after each page but the impact of this on the children, supported our choice not to do this. If the study was examining specific interface attributes or fixation at a word level, rather than the two large AOIs, recalibration would be needed more often (Holmqvist and Andersson 2017).

The second question *RQ2: What insights can eye tracking technology provide towards understanding how children interact with a low-fidelity prototype?* Aimed to inform the use of eye tracking in this context. Based on the fixation data there was no difference in how children viewed the 1st page of the prototype; this being where they had to read about it and make a judgement. However, when the fixation data from all the 10 pages was aggregated together

there was a significant difference in the number of fixations with children fixating more on both of the AOIs in the wireframe prototype compared to the colour version. This may suggest that the wireframe prototype is more cognitively demanding than the colour version but it could also be that there may have been differences in the reading ability of the children between the two groups as proficient readers are known to make fewer fixations than beginners (Rayner et al. 2012). Another explanation could be based on the similarities of different pages; with the colour version, the overall pictures appeared more similar than they did in the wireframe which could have subliminally suggested to the child that there was little to attend to there. The children may also have been able to process the visuals more efficiently due to the added detail; this may require further research within the context of prototypes.

Despite these potential limitations the eye tracking showed that most children viewed all the pages of the prototype, and in most instances, looked at both the text and the visuals. Notably eye tracking can validate the study, providing evidence that the children examined all the pages of the prototype to make an informed judgement of the game being evaluated. With greater refinement of the AOI it would be possible to know where the child's visual attention was, and we could explore whether any visual or textual hierarchy existed. In addition, by examining scan paths and regression from one AOI to another it may be possible to identify interface components that are confusing for the children which could help improve the overall design of the game. For example, fixation data helped identify distracting content for the user and improve the usability (Al-Zeer et al. 2014).

6 CONCLUSIONS AND FUTURE WORK

This paper examined the fidelity effect and the insights that eye tracking technology can offer when evaluating prototypes with children. There were no differences in the ratings of children when viewing the low-fidelity and colour prototype when presented as a storyboard within a booklet, but the Again Again table showed stronger preferences for the coloured prototype. Overall the study gives confidence in the use of low-fidelity paper prototypes to evaluate game concepts with children.

The eye tracking technology demonstrated that most of the children engaged with all 10 pages of the prototypes, giving some evidence that the children reported their perceptions of the entire game and did not disengage from the evaluation. CCI researchers can take confidence that when presenting prototypes using multiple pages, as demonstrated here, children are prepared to read the text description and view the images. However, the number of fixations was higher on the wireframe

prototype which may suggest higher cognitive demands; further work is required to understand these differences and establish if they are repeated in similar studies. It would be interest to compare data from adults 'to children's in a similar study to see if there was less variability. Overall eye tracking glasses, despite some of their limitations, offer good insights into understanding children's behaviour and could help inform design by understanding what children are looking at within low-fidelity prototypes.

7 ACKNOWLEDGEMENTS

We would like to thank the children from class 5 and teachers from Heskin Pemberton's Church of England Primary School for their participation in the study.

REFERENCES

- Al-Wabil, A., Al-Husian, L. Al-Murshad, R. and Al-Nafjan, A. (2010). Applying the retrospective think-aloud protocol in usability evaluations with children: Seeing through children's eyes. *In Proceedings of the International Conference on User Science and Engineering. 13-15 Dec., 2010.*
- Al-Zeer, S. Al-Ghanim, A. and Al-Wakeel, L.A. (2010). Visual attention in interaction with Arabic augmentative and alternative communication apps. *In Proceedings of the International Conference on User Science and Engineering. 2-5 Sept. 2014.*
- Bertou, E. and Shahid, S. (2014). Lowfidelity prototyping tablet applications for children. *Proceedings of the Interaction design and children 2014.*
- Bilal, D. and Gwizdka, J. (2016) Children's eye-fixations on google search results. *In Proceedings of the 79th ASIS Annual Meeting: Creating Knowledge, Enhancing Lives through Information Technology. Copenhagen, Denmark. American Society for Information Science.*
- Buxton, B. *Sketching User Experiences - getting the design right and right design.* (2007). Morgan Kaufmann, San Francisco.
- Chen, Y. and Tsai, M.Y. (2015) Eye-hand coordination strategies during active video game playing: An eye-tracking study. *Computers in Human Behavior*, 51.
- Currie, J. Bond, R., McCullagh, P. Black, P., Finlay D., and Peace, A. (2017). Eye tracking the visual attention of nurses interpreting simulated vital signs scenarios: mining metrics to discriminate between performance level. *IEEE Transactions on Human-Machine Systems*, 48, 2, 113-124.
- Derboven, J. Roeck, D.D., Verstraete, M., Geerts, D., Schneider-Barnes, J. and Luten, K. (2010).

- Comparing User Interaction with Low and High Fidelity Prototypes of Tabletop Surfaces. In *Proceedings of the NordiCHI. Reykjavik, Iceland*. ACM.
- Fischer, P.M., Richards, J.W., Berman E.J., and Krugman, D.M. (1989) Recall and eye tracking study of adolescents viewing tobacco advertisements. *Jama*, 261, 1, 84-89
- .Giannakos, M.N., Papavlasopoulou, S. and Sharma, K. (2020). Monitoring Children's Learning Through Wearable Eye-Tracking: The Case of a Making-Based Coding Activity. *IEEE Pervasive Computing*, 19, 1.
- Gomes, A., Oh, H., Chisik, Y. and Chen, M. (2012) Ilha Musical: a CAVE for nurturing cultural appreciation. In *Proceedings of the Proceedings of the 11th International Conference on Interaction Design and Children. Bremen, Germany*.
- Gossen, T., Höbel, J., and Nürnberger, A. (2014). Usability and perception of young users and adults on targeted web search engines. In *Proceedings of the 5th Information Interaction in Context Symposium. Regensburg, Germany*.
- Hanna, L., Neapolitan, D. and Risdén, K. (2004) Evaluating computer game concepts with children. Proceedings of the 2004 conference on Interaction design and children: building a community.
- Hershman, A., Nazare, J., Qi, J., Saveski, M., Roy, D. and Resnick, M. (2018). Light it up: Using paper circuitry to enhance low-fidelity paper prototypes for children. Proceedings of the 17th ACM conference on interaction design and children.
- Hirschmanner, M., Lammer, L. and Vincze, M. (2015). Mattie: A simple educational platform for children to realize their first robot prototype. Proceedings of the 14th international conference on interaction design and children
- Hoggenmüller, M., Tomitsch, M., Hespanhol, L., Tran, T.T.M., Worrall, S. and Nebot, E. (2021). *Context-Based Interface Prototyping: Understanding the Effect of Prototype Representation on User Feedback*. *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*.
- Holmqvist, K. and Andersson, R. (2017)/ Eye tracking: A comprehensive guide to methods. *paradigms and measures*.
- Huestegge, L., Radach, R., Corbic, D. and Huestegge, S.M. (2009). Oculomotor and linguistic determinants of reading development: A longitudinal study. *Vision research*, 49, 24, 2948-2959
- Kleberg, J.L., Högström, J., Sundström, K., Frick, A. and Serlachius, E. (2020). Delayed gaze shifts away from others' eyes in children and adolescents with social anxiety disorder. *Journal of Affective Disorders*.
- Kohler, B., Haladjian, J., Simeonova, B. and Ismailovic, D. (2012) Feedback in Low vs High Fidelity Visuals for Game Prototypes. In *Proceedings of the Games and Software Engineering*. Zurich, 2012. IEEE.
- A. N. Kornev, I. Balciuniene and S. R. Oganov. (2018). The strategic reading brain development: An eye-tracking study of the text reading in typically-developing and dyslexic children. *International Journal of Psychophysiology*, 131.
- Krejtz, K., Krejtz, I., Duchowski, A., Szarkowska, A. and Walczak, A. (2012) Multimodal learning with audio description: an eye tracking study of children's gaze during a visual recognition task. In *Proceedings of the ACM Symposium on Applied Perception*. Los Angeles, California.
- Levi, M.D. and Conrad, F.G. (1996). A heuristic evaluation of a world wide web prototype. *Interactions*, 3, 4, 50-61.
- Lim, Y-K., Pangam, A., Periyasami, S. and Aneja, S. (2006). Comparative Analysis of High- and Low-fidelity Prototypes for More Valid Usability Evaluations of Mobile Devices. In *Proceedings of the NordiCHI*. Oslo, Norway. ACM.
- Just, M.A. and Carpenter, P.A. (1980). A theory of reading: from eye fixations to comprehension. *Psychol Rev*, 87, 4, 329-354
- Martín-Arnal, L.A., León, J.A. van den Broek, P. and Olmos, R. (2019). Understanding comics. A comparison between children and adults through a coherence/incoherence paradigm in an eye-tracking study. *Psicología Educativa. Revista de los Psicólogos de la Educación*, 25, 2, 127-137.
- Moeller, K., Neuburger, S., Kaufmann, L., Landerl, K. and Nuerk, H. C. (2009). Basic number processing deficits in developmental dyscalculia: Evidence from eye tracking. *Cognitive Development*, 24, 4.
- Othman, A. and Mohsin, M. (2017). How could robots improve social skills in children with Autism? *6th International Conference on Information and Communication Technology and Accessibility (ICTA)*.
- Papavlasopoulou, S., Sharma, K. and Giannakos, M.N. (2020) Coding activities for children: Coupling eye-tracking with qualitative data to investigate gender differences. *Computers in Human Behavior*, 105.
- Kent, M.P. Pauzé, E., Roy, E-A., de Billy N. and Czoli, C. (2019). Children and adolescents'

- exposure to food and beverage marketing in social media apps. *Pediatric obesity*, 14, 6.
- Rayner, K. Pollatsek, A., Ashby, J and Clifton Jr, C. (2012). *Psychology of reading*. Psychology Press.
- Read, J.C. MacFarlane, S and Casey, C. (2002). Endurability, Engagement and Expectations: Measuring Children's Fun. In *Proceedings of the Interaction Design and Children*. Eindhoven.
- Read, J.C. (2012). Evaluating artefacts with children: age and technology effects in the reporting of expected and experienced fun. In *Proceedings of the 14th ACM International Conference on Multimodal Interaction*. Santa Monica. ACM.
- Reilly, D., Dearman, D., Welsman-Dinelle, M. and Inkpen, K. (2005). Evaluating Early Prototypes in Context: Trade-offs, Challenges, and Successes. *Pervasive Computing*, 4, 4, 42-50.
- Rudd, J. Stern, K. and Isesensee, S. (1996). Low vs high-fidelity prototyping debate. *Interactions*, 3, 1, 75-85
- Sauer, J., Holger, F. and Ruttinger, B. (2008). Designing interactive consumer products: utility of low-fidelity prototypes and effectiveness of enhanced control labeling. *Applied Ergonomics*, 39, 1, 71-85.
- Sauer, J. and Sonderegger, A. (2009). The influence of prototype fidelity and aesthetics of design in usability tests: Effects on user behaviour, subjective evaluation and emotion. *Applied Ergonomics*, 40, 670-677.
- Sefelin, R., Tscheligi, M. and Giller, V. (2003) Paper Prototyping - What is it good for? A comparison of Paper and Computer based Low fidelity prototyping. In *Proceedings of the Conference on Human Factors in Computing Systems*. Fort Lauderdale.
- Sharma, K., Papavaslopoulou, S. and Giannakos, M.N. (2019). Coding games and robots to enhance computational thinking: How collaboration and engagement moderate children's attitudes? *International Journal of Child-Computer Interaction*, 21.
- Sim, G., & Cassidy, B. (2013). Investigating the fidelity effect when evaluating game prototypes with children. Paper presented at the Proceedings of the 27th International BCS Human Computer Interaction Conference, London, UK.
- Sim, G., Cassidy, B., & Read, J. C. (2013). Understanding the Fidelity Effect when Evaluating Games with children. Paper presented at the Interaction Design and Children, New York.
- Sim, G., Horton, M., & McKnight, L. (2016). iPad vs Paper Prototypes: Does Form Factor Affect Children's Ratings of a Game Concept? Paper presented at the Proceedings of the The 15th International Conference on Interaction Design and Children.
- Sim, G., & Bond, R. (2021). Eye tracking in child computer interaction: Challenges and opportunities. *International Journal of Child-Computer Interaction*. 30.
- Snyder, C. (2003) *Paper Prototyping: The fast and Easy Way to Define and Refine User Interfaces*. Morgan Kaufmann, San Francisco.
- Su, Q., Chen, F., Li, H., Yan, N. and Wang, L. (2018). Multimodal Emotion Perception in Children with Autism Spectrum Disorder by Eye Tracking Study. *Conference on Biomedical Engineering and Sciences (IECBES)*.
- Thomsen, S.R. and Fulton, K. (2007). Adolescents' Attention to Responsibility Messages in Magazine Alcohol Advertisements: An Eye-Tracking Approach. *Journal of Adolescent Health*, 41, 1.
- Torney, H., Harvey, A., Finlay, D., Magee, J., Funston, R. and Bond, R. (2018) Eye-tracking analysis to compute the visual hierarchy of user interfaces on automated external defibrillators. *British HCI Conference 2018*.
- Mestres, E.T. and Pellicer-Sánchez, A. (2019) Young EFL learners' processing of multimodal input: Examining learners' eye movements. *System*, 80.
- Velazquez, C.E. and Pasch, K.E. (2014). Attention to food and beverage advertisements as measured by eye-tracking technology and the food preferences and choices of youth. *Journal of the Academy of Nutrition and Dietetics*, 114, 4, 578-582.
- Virzi, R.A. Sokolov, J.L. and Karis, D. (1996) Usability problem identification using both low- and high-fidelity prototypes. In *Proceedings of the Proceedings of the SIGCHI Conference on Human Factors in Computing Systems Vancouver, British Columbia, Canada*.
- von Reumont, R. and Budke, A. (2020) Strategies for Successful Learning with Geographical Comics: An Eye-Tracking Study with Young Learners. *Education Sciences*, 10, 10, 293.
- Wiklund, M., Thurrot, C. and Dumas, J. (1992) Does the Fidelity of Software Prototypes Affect the Perception of Usability. In *Proceedings of the Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, Atlanta, USA.