University of Central Lancashire UCLan

Central Lancashire Online Knowledge (CLoK)

Title	Using Eye-Tracking to Demonstrate Children's Attention to Detail when Evaluating Low-Fidelity Prototypes.
Type	Article
URL	https://clok.uclan.ac.uk/49890/
DOI	https://doi.org/10.1093/iwc/iwad052
Date	2024
Citation	Sim, Gavin Robert and Read, Janet C (2024) Using Eye-Tracking to Demonstrate Children's Attention to Detail when Evaluating Low-Fidelity Prototypes. Interacting with Computers, 36 (2). pp. 100-112. ISSN 0953-5438
Creators	Sim, Gavin Robert and Read, Janet C

It is advisable to refer to the publisher’s version if you intend to cite from the work.
https://doi.org/10.1093/iwc/iwad052

For information about Research at UCLan please go to http://www.uclan.ac.uk/research/

All outputs in CLoK are protected by Intellectual Property Rights law, including Copyright law. Copyright, IPR and Moral Rights for the works on this site are retained by the individual authors and/or other copyright owners. Terms and conditions for use of this material are defined in the http://clok.uclan.ac.uk/policies/

Using Eye-Tracking to Demonstrate Children's Attention to Detail when Evaluating Low-Fidelity Prototypes.

This study used eye tracking glasses to better understand how children explore low-fidelity prototypes in the context of user experience studies and to explore the potential of eye tracking in this context. The main research question that was being explored was whether the aesthetic refinement, either wireframe or high-resolution images, would affect children's self-report and if so, or if not, 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 there were no differences in the time spent viewing the prototypes and most of the children focused on both the visuals and text on all the pages. However, there were a higher number of fixations recorded in the wireframe prototype compared to the photo-realistic version. This paper contributes to the design of prototypes through an understanding of how children interact with prototypes, demonstrating the importance of the text along with the visuals when evaluating game concepts with children. Further research is required to understand the differences and whether similar results are replicated with different games.

Keywords: Eye Tracking, Children, User Experience, Prototype, Games

1 INTRODUCTION

Child Computer Interaction is defined as "a discipline concerned with the design, evaluation and implementation of interactive computing systems for children's use and with the study of major phenomena surrounding them" (Read & Bekker, 2011). In terms of both evaluation and design, obtaining feedback from children has played a crucial role; and various methods have emerged for evaluating technology with children which deliver value to designers. Methods include Picture Cards (Bekker & Antle, 2011), This or That (Zaman, Abeele, & De Grooff, 2013) and The Fun Toolkit (Read & MacFarlane, 2006) all of which provide data on aspects such as preference, fun and usability. Technology being presented to children for evaluation can take different forms; examples have included paper prototypes (Rante & Basuki, 2019), interactive prototypes (Volioti, Tsiatsos, Mavropoulou, & Karagiannidis, 2016) and fully functional products (Sim & Horton, 2012).

When asking children to make a judgement on a technology, there is an assumption that the child is looking at that technology and making an assessment. There is very little research on understanding how children interact with technology when forming opinions; studies tend to focus on the results of the evaluation rather than on the process. Observing children evaluating technology can provide some insights (Kara & Cagiltay, 2020), asking them about the reasons for any reported choices can also help (Zaman & Abeele, 2010); a third option is to establish where a child's attention is during an evaluation activity (Gulz, Londos, & Haake, 2020).

Eye tracking is a method that can facilitate this latter choice. It involves the collection of eye movement data that relates to visual stimuli processed by the brain. Within the context of a technology evaluation, these stimuli may relate to a child looking at the visual elements of a product or to a child reading a description related to the product. Most current eye tracking systems use infrared light that is reflected off the cornea in the eye to be then captured by cameras (corneal reflection-based eye trackers) and subsequently processed by software. Current technology enables studies to be easily performed with children in natural environments, like schools and clubs, where they can be looking at almost anything (Potvin Kent, Pauzé, Roy, de Billy, & Czoli, 2019). Examples of such use include using eye tracking to explore how children interact with programming environments (Sharma, Papavlasopoulou, & Giannakos, 2019) and using it to examine hand - eye co-ordination whilst playing video games (Chen & Tsai, 2015). In considering how children carry out an evaluation of a product we can use eye tracking to confirm that children have been looking at the things that we have presented 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, rather than a fully functional, product (Bertou & Shahid, 2014; Hershman et al., 2018). Prototypes can be created in different forms and described in different ways. One categorization that is used is in terms of 'fidelity'. Low-fidelity is used to describe prototypes that are typically made from a material that is different from the final product, such as paper sketches or cardboard (Rudd, Stern, & Isesensee, 1996). These can be used to open the design space for new alternatives and to explore concepts like navigation or organisation of material (Buxton, 2010). For example, low-fidelity paper prototypes have been created to explore Olfactory interfaces, requiring the user to scratch and sniff stickers (Brooks & Lopes, 2023). High-fidelity prototypes usually offer a level of functional interactivity using materials that you would expect to find in the final product (Rudd et al., 1996). For example, a high-fidelity prototype of an Android Application could be developed and deployed to Android phone with scaled back functionality to be tested with users (Toolaroud et al., 2023), whereas a low fidelity prototype might be presented as a sequence of post it notes (Nebeling & Madier, 2019).

It is not trivial studying fidelity effects as it can be difficult to always understand to what extent findings from evaluations align to the concept being presented or to the characteristics of the prototype itself (Lim, Pangam, Periyasami, & Aneja, 2006). In a study with adults, it was suggested a lack of refined graphics may bias evaluators against products (Kohler, Haladjian, Simeonova, & Ismailovic, 2012). Studies can yield contradictory findings; in a study with children evaluating a single concept using different prototype fidelities results were found to be similar (Sim, Cassidy, & Read, 2013) and, in a different study with the same conditions, different (Sim & Cassidy, 2013) suggesting that there may be things happening during the evaluation that are as yet not well understood.

There is little work using eye tracking to gain insights into how users interact with low-fidelity prototypes; especially with children, although studies of children reading can give some insights. In an educational context, studying biology, eye tracking results revealed that children's comprehension was largely driven by the text rather than the images (Hannus & Hyönä, 1999) which may suggest that when children are evaluating prototypes the textual description of the game may play a crucial role in influencing results. Aspects of the prototype, such as colour images, may bias cognitive process (Damle & Smith, 2009) and potentially influence children's ratings of the prototype. Therefore, understanding how children interact with different elements within a prototype can potentially help inform the design of prototypes to facilitate improved validity in the results of an evaluation study.

This paper aims to explore the fidelity effect by using eye tracking technology to better understand how children interact with different prototypes. Virzi et al. (Virzi, Sokolov, & Karis, 1996) suggest that prototype fidelity can differ on four dimensions: degree of functionality, similarity of interaction, breadth of features and aesthetic refinement. This study compares children viewing a game concept with two different aesthetic refinements;

interaction, functionality and breadth of features were kept the same across both versions. Surveys and eye tracking technology are used to answer two research questions:

1. RQ1: When presenting a prototype to children, in different levels of aesthetic refinement, can this affect their overall rating of a game?
2. RQ2: What insights can eye tracking technology provide towards understanding how children interact with low-fidelity prototypes to help inform prototype design?

2 BACKGROUND AND RELATED WORK

2.1 Prototypes and Children in HCI

Prototypes have been used for a variety of different reasons for example, the evaluation of design ideas (Hanna, Neapolitan, & Risden, 2004), the exploration of ideas for interaction (Lee, Cagiltay, & Mutlu, 2022), and as props to assist with communication as part of a development process (Krishnakumar, Berdanier, Lauff, McComb, & Menold, 2022; Levi & Conrad, 1996; Reilly, Dearman, Welsman-Dinelle, & Inkpen, 2005). To evaluate design ideas, once a prototype is developed it is used to get feedback from end users to inform future iterations and identify potential usability problems (Mathis, O'hagan, Vaniea, & Khamis, 2022; Virzi et al., 1996; Wiklund, Thurrot, & Dumas, 1992).

Within CCI, prototypes are used for a variety of different purposes including facilitating design (Hirschmanner, Lammer, & Vincze, 2015), learning programming (Sabuncuoglu & Sezgin, 2022), exploration of ideas (Hanna et al., 2004) and as props to assist with communication (Guha et al., 2004). When using prototypes it can be difficult to discern if the findings from evaluations are closely aligned to the concept of the artifact being evaluated or are more associated with the characteristics of the physical prototype itself (Lim et al., 2006). For example, in a study in an AR context, colour was shown to have influenced users ability to comprehend the interaction and user experience (Rehring & Ahlemann, 2020). It has also been suggested that participants may idealize a product that is in a low fidelity form, which can raise concerns about the validity of results (Christoforakos & Diefenbach, 2019). Counter wise, (Snyder, 2003) suggested that (in a high fidelity prototype) ‘’When something appears to be finished, minor flaws stand out and will catch the users’ attention”. In a study looking into the effect of visuals on usability studies using game prototypes, comparing low and high fidelity, it appeared that most of the participants paid very little attention to the visuals (Kohler et al., 2012). Within VR, prototypes that facilitated high immersion resulted in higher subjective ratings when comparing prototypes with different levels of interaction and immersion (Christoforakos & Diefenbach, 2019).

There have been several comparative research studies of prototypes at different fidelity levels (Hoggenmüller et al., 2021; Lim et al., 2006; Sefelin, Tscheligi, & Giller, 2003; Wiklund et al., 1992). In the 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 (Walker, Takayama, & Landay, 2002). Seflin *et al.* (Sefelin et al., 2003) investigated whether participants confronted with a paper-based low-fidelity prototype differed in their willingness to criticize a system, compared to a computer based prototype. The results showed that there was no difference in the number of criticisms, but the users preferred the computer prototype.

There are only a few research studies that have looked at the fidelity effect when evaluating prototypes with children. One study invited 16-17 year old pupils 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 towards a high fidelity prototype 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 interaction in low fidelity prototypes were highlighted 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 and the results showed that there was little difference in reported user experience between the three prototypes and very few usability problems were unique to a specific prototype. In a further study, the physical form factor was examined to determine whether a prototype presented on an iPad differed to that on paper (Sim, Horton, & McKnight, 2016); interestingly the children rated the aesthetics higher on the paper version compared to the iPad, despite the graphics being identical. This suggests that the ratings of the aesthetics may be influenced by the form in which a prototype is presented to children.

2.2 Attention and Eye Tracking

A key theoretical underpinning for eye tracking research is the eye-mind hypothesis (Just & Carpenter, 1980) which suggests that when a person gazes on a stimuli they will continue to look at it until it is comprehended; this suggests a link between eye gaze and the cognitive processing of information. This theory has been challenged over the years when depth is added to a visual stimuli (Underwood & Everatt, 1992) and as it is possible for mind wondering to occur (Gwizdka, 2019). However, it has been suggested that the eye-mind hypothesis holds true when information is presented in a visual modality (Ball & Richardson, 2022) such as a prototype of a software application. Therefore, it is conjectured that if a child is looking at a prototype, they are processing the stimuli.

Prototypes can be presented using a combination of pictures and text. Based on the cognitive theory of multimedia learning (Mayer, 2005, 2014), written words and static images are actively processed in separate channels of limited capacity which are dual processed and the information is integrated. In an educational context, studying biology, eye tracking results revealed that children's comprehension was largely driven by the text rather than the images (Hannus & Hyönä, 1999). Whilst in another study illustrations enhanced the comprehension of the text (Mason, Pluchino, Tornatora, & Ariasi, 2013). When evaluating prototypes with children the textual description of the game may play a crucial role in influencing the results. Understanding how children interact with these different elements within a prototype can help inform the design of prototypes to facilitate improved validity in the results of an evaluation study.

When viewing stimuli the eye does not move in a continuum but comes to rest for short periods of time, called fixations, before moving, described as a saccade, to the next fixation point. Information related to the stimuli, whether that is text or an image, is processed during fixation (Holmqvist & Andersson, 2017). Within the context of reading text, the majority of saccadic eye movement is forward but regression does occur when confusion occurs and the eye may move back to previous words (Rayner, Pollatsek, Ashby, & Clifton Jr, 2012). Return sweeps also occur when the eye moves from the end of one line to the beginning of the next line. Within reading studies, children have been shown to have longer fixations, shorter saccades, more re-fixations and many regressions when compared with adults (Huestegge, Radach, Corbic, & Huestegge, 2009; Wertli, Schötzau, & Palmowski-Wolfe, 2023).

When examining visual attention in static images one model of processing information is the saliency model (Koch & Ullman, 1985). Eye movement when viewing an image is driven by high image saliency whereby regions differ in factor such as luminance, contrast, colour, texture and contour density (Underwood, 2009). Research has also shown that age can influence visual attention on elements within an image, for example young children aged between 3-30 months are drawn to facial features within images (Frank, Vul, & Saxe, 2012). Whilst bright primary colours can also influence children's preferences for products (Gollety & Guichard, 2011). When presenting images to children in the form of a prototype, then the aesthetics may influence their fixations on regions if there is high saliency within the scene. However, it is important to note that this model may not hold true with dynamic content (Rider, Coutrot, Pellicano, Dakin, & Mareschal, 2018).

To facilitate eye tracking studies, technology composed of two artefacts, hardware and software, is needed. The hardware is either fixed to an object which is being viewed, for example this could be a car dashboard or a screen, or worn, in the form of glasses or VR headsets. In recent years 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. Eye trackers collect eye gaze data in the form of two-dimensional Cartesian coordinates (x, y) that represent a point on the screen display or a point in the world where the eyes have focused. Eye tracking software captures the time and duration of fixations and the movements (saccades) of the eyes from one point to another. Together, fixations and saccades can be interpreted to infer something about human visual processing or behaviour.

The evaluator has to interpret the data within the context of the work; depending on the activity the user is performing, a high number of fixations have been described as negatively correlating with search efficiency (Bilal & Gwizdka, 2016) or have suggested users having difficulty interpreting the information or layout (Al-Wabil, Al-Husian, Al-Murshad, & Al-Nafjan, 2010). 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, or indeed some other reason (Torney et al., 2018). Fixation-based metrics can also be useful to explore content, an important HCI design concern, where they can be used to compare and benchmark different areas of interest (AOI). AOI are predefined areas (e.g., the navigation bar) that are determined by the researcher. Using eye tracking software, the researcher can mark out an AOI and then the software can report visits to these AOI as counts, dwell times or frequencies. As with general fixations, interpretation of such data has to be done with caution: if certain AOI 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 or they are confused (Ball & Richardson, 2022; Currie et al., 2017; Rante & Basuki, 2019).

Eye tracking studies with children generally focus on reading, scanning for information and fixations (Sim & Bond, 2021). Much of the early work on eye tracking with children focused on understanding reading (Marcel & Patricia, 1980) with a move to understand differences between populations, for example children with numeric processing deficits (Moeller, Neuburger, Kaufmann, Landerl, & Nuerk, 2009), as a diagnostic tool (Vajs, Ković, Papić, Savić, & Janković, 2022) or with autism (Su, Chen, Li, Yan, & Wang, 2018). Although eye tracking has been used with children to evaluate prototypes of applications including augmented reality applications (Gomes, Oh, Chisik, & Chen, 2012) and robot interaction (Othman & Mohsin, 2017), there have been no studies to date examining children's interactions with low-fidelity prototypes. However, given the similarity in structure of combining images and text, eye tracking studies within media, including comic books and advertisements, may afford insights into how children visualize data 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 they required more cognitive effort to comprehend the stories (Martín-Arnal, León, van den Broek, & Olmos, 2019). In a study examining the effectiveness of geographical comics as a learning aid for children, eye tracking was used with 36 children aged 10-14 (von Reumont & Budke, 2020). Heat map data from the studies showed the highest concentration of fixations over the text elements of the comic book. This would be expected as more fixations are required to decipher text than images.

Eye tracking studies examining children's viewing behaviour with advertisements has tended to focus on unhealthy products including food, tobacco, and alcohol. In a study looking at unhealthy food advertisements (Velazquez & Pasch, 2014) the children's food preference correlated with the length of time and number of times a child looked at an item. In a study examining children viewing tobacco advertisements the warning message was only viewed for 8% of the total time and in 43.6% of cases the children did not view the message at all (Fischer, Richards, Berman, & Krugman, 1989). Researchers in a study looking at alcohol advertisements (Thomsen & Fulton, 2007), and their associated warning messages, demonstrated that when the warning message was incorporated more prominently in the advertisement children spent more time attending to it. When text was small and placed at the bottom of the page only 7% of the total viewing time was spent looking at the warning message.

Given that some eye-tracking studies suggest a focus on text, and others a focus on images, it is interesting to the HCI community to learn how children might be attending to paper prototypes that they may see during an evaluation. In a prototype evaluation the child is not seeking to be entertained and is not really seeking information – he or she is looking at presentations in order to understand the mechanisms and look of a product. It is therefore important to understand how children attend to different areas within a storyboard if results of an evaluation will inform future iterations of a product.

This paper uses eye tracking technology to examine the fidelity effect when children are presented with prototypes with two different levels of aesthetic refinement whilst exploring the efficacy of eye tracking to gain insights into children's visual attention during an evaluation exercise.

3 METHOD

The study employed a between subject design methodology with prototype fidelity (wireframe vs. photo-realistic) as the independent variable and user experience ratings, and attention as captured by eye-tracking data, as the two dependent variables. A decision had been made to create two matched storyboards – one with all the images as wireframes and one with all the images being realistic copies from the actual game. When viewing an image, factor such as luminance, contrast and colour can influence fixations (Underwood, 2009) and the wireframe prototype lacks many of these attributes. The study took place in a school over a 6-week period to ensure that all the children had a turn to participate. The 1st week was an orientation exercise to help children understand the technology, the following five weeks saw they children coming to the study one at a time.

3.1 Participants

The participants were 28 school children from a UK primary school. In the school where the study took place, children in Year 5 and 6 are merged into a single school class, thus the children in the study were 9, 10 or 11 on the days of the study. Developmentally, the children were at a similar stage (Piaget, 1970), and the teacher said that there was considerable overlap in the ability of the children within both year groups. Good practice on child participation, and not being selective, dictated that all 28 children in the class be included. The same researcher (the first author) was present at all the sessions.

3.2 Apparatus

Two prototypes of a single game were needed: a wireframe prototype and a photo-realistic prototype. 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 meaning it was unlikely that children in the current study would have seen it before. As this was a ‘real’ game, the game mechanics, narrative, and visuals were established and so a paper prototype could be generated to be used as an experimental tool. The focus of the study was to understand how children interacted; it was not on gathering data to redesign the game.

Figure 1: Splode Game

To create the prototype the game was played by the first author who determined that ten screens would be needed to represent a game play event. These screens were captured as images and used in the photo-realistic prototype. To create the wireframe storyboard, the collected images were traced using Adobe Illustrator, see Figure 2; this allowed all the detail to be accurately captured for the non-photo-realistic version. This approach to create wireframes has been used by the authors in a previous study examining the fidelity effect (Sim et al., 2013).

To highlight movement and interaction, arrows and hands were added to the images in both versions. To complete the storyboard, text that explained the game concept and the mechanics (the same for both versions) was added underneath each image.

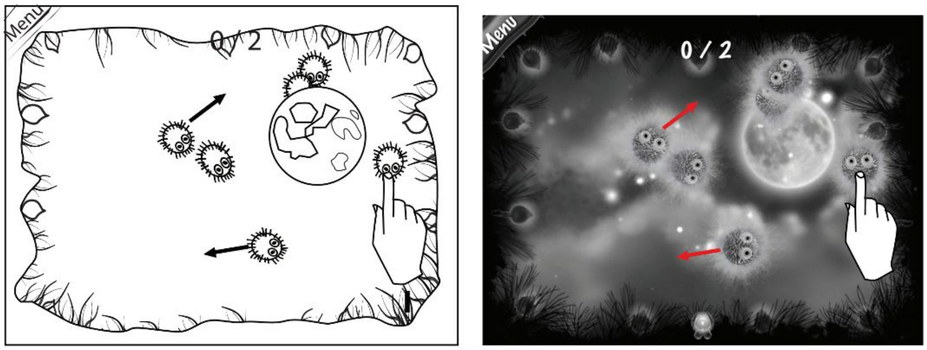

Figure 2: Wireframe (left) and Photo-realistic (right) prototypes of the Game

Both versions were presented to the children on ten sheets of A4 paper (printed on one side only), with one image and one block of text on each page, 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.

3.3 Capturing UX Ratings

To measure user experience, the decision was made to use an adaptation of the Fun Toolkit (Read, MacFarlane, & Casey, 2002) using the Smileyometer and Again Again table and to follow the same protocol as used within [45]. The Smileyometer, is a visual analogue scale with a coding based upon a 5-point scale, see Figure 3.

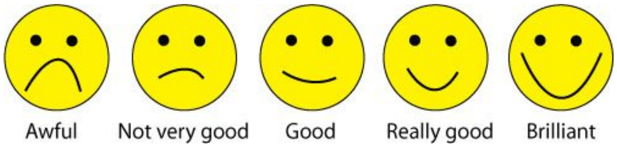

Figure 3: The Smileyometer

A data capture form was created that had two Smileyometers on it; one to measure initial expectations of the game after seeing the first screen, and one to measure expectations after the child had seen all the pages of the game explaining the concept in greater detail.

Three questions were also created to evaluate the prototype and game concept with the children, again using the Smileyometer to record their response, the questions were:

- The idea of the game is…
- How do you think the (real) game will look…
- Overall the game is…

An Again Again table was also used on the form to establish whether the children would download the game based on the storyboard presented. The table required the children to answer ‘yes’, ‘maybe’ or ‘no’ to these two 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?

Thus, the children were presented, on the day, with two documents; one was the ten-page storyboard in either a wireframe or photo-realistic form, and the second was the data capture form to complete the Smileyometers, and the Again Again table. Children brought their own pens from the classroom to complete the survey.

3.4 Capturing Attention

Tobii Glass 2 eye trackers were used to capture eye movement from the children; these have been successfully used with children in other studies looking at programming (Giannakos, Papavlasopoulou, & Sharma, 2020; Papavlasopoulou, Sharma, & Giannakos, 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. The laptop contained the Tobii Glasses Controller software and was used by the researcher to set up the children and monitor data capture during the study. The child was not required to view the monitor to perform the study and therefore aspects such as its screen size and resolution would not impact the results.

3.5 Procedure

The researcher initially talked through the study with the head teacher at a local school that had previously worked with the university. A class was identified that was suitable and an information sheet was sent to the parents along with a consent form to be returned to the school. Before the study started, the first author went into the school to show the chosen class the eye tracking technology, explain how it worked, and to let the children ask any questions. This was to ensure that the children understood the study, the technology, the data being captured and could make an informed decision about whether they wanted to participate. It also enabled the researcher to discuss with the teacher whether any child had any known visual impairment that may impact them being able to participate; none were reported.

Following this initial orientation session, the study with the children was conducted over a five-week period during which a single researcher went into the school, each Wednesday afternoon (as this would be less disruptive), to perform the study. The physical location used changed over this period and included the sports hall, headteacher’s office and a corridor. Within each location there was a desk and chair for the child and a separate chair for the researcher to sit at during the study. The procedure was the same in each location. Each child only attended on one day, the reason the study took five weeks was that there were 28 children and the study had to be completed individually; between 5 and 6 children participated in the study each week.

On the day, children were sent by the teacher, one at a time, to where the study was taking place. The child was greeted by the researcher, who again explained the purpose of the study and sought assent. At this point the child was allocated to one of the two conditions either wireframe or photo-realistic paper prototype. The child was then presented with the data capture form, a pen and the paper prototype 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. Once the Tobii glasses were on, the child was shown the computer screen which showed them their eyes and the video output from the glasses. The researcher then explained the calibration process and calibrated the glasses for the individual child. 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. During the study, by focusing on the laptop, sampling data was checked, and if lost, the next participant would repeat the 'lost' condition (e.g., Photo-realistic prototype) to ensure equal numbers for both conditions.

Once initial calibration was completed, recording from the glasses 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 it being distracting. The children were asked to look at the first page of the storyboard, taking as long as they wanted to view the image and read the information relating to the description of the game which was as described on the app store. Once they had finished examining the 1st page, they were asked to answer the first question of the Smileyometer regarding their expectations.

The children then went through the remaining nine pages of the storyboard, and once complete, they answered the remaining questions on the survey sheet rating 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. Before leaving the study, children were shown the video capture of their eye movement from the glasses and assent was sought again for using the video and survey data. The whole procedure lasted about 20-25 minutes per child.

3.6 Analysis

3.6.1 Survey Data

The questions using the Smileyometer, see Figure 3, 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 in order to show differences (Read, 2012).

3.6.2 Eye Tracking Data

Data collected was in the form of numeric logging data of eye movement and video recorded via the Tobii glasses, the information was stored on the SD card before being transferred to the Tobii Pro Lab software for analysis. Analysis of the eye tracking data was carried out in three phases; firstly, there was a quantity / quality check to determine whether the data from an individual child was suitable for inclusion. This was also monitored during the study and 7 participants data was removed.

Each child's data was examined to establish whether sufficient gaze samples had been captured. The sampling rate of the eye tracker is the number of data samples per second collected for each eye and the Tobii Glasses has a rate of 50Hz. Around 5-10% of samples can be lost via blinking and other data can be lost if children look underneath the glass or to the side, thus samples can be lost for a variety of different reasons (Hollander & Huette, 2022). A decision was therefore made to only use logged data with a sample of at least 80%, which is higher than previous research whereby sampling rates of less than 66% have been discounted (Niehorster et al., 2020). This was essential as data was lost during the study for various reasons, including lighting and children looking underneath the glasses.

The second phase of analysis was the mapping to AOIs. Two areas of interest (AOI) were established for each page of the storyboard; one for the text and the other for the image. This enabled a comparison of the fixation data between the two areas. The researcher initially divided the video in separate clips, one for each page of the storyboard which resulted in 10 videos for each child. To do this mapping, given that children had been looking at images and text in a booklet, the researcher uploaded twenty still images, corresponding to each page in the booklet (10 each for the wireframe and photo-realistic prototypes). . The software enabled automated mapping of fixation data to the still images, but as this can still be unreliable the decision was made to manually plot fixations onto each image. Whilst watching the video with the fixations visible, the first fixation on the page was captured by clicking on the corresponding point in the still image and then, all fixation points, until the child moved to a next page, were recorded. This was necessary as sometimes the child would look away from the prototype at the researcher, especially between turns of the pages. Only fixation points that mapped to one of the two AOIs were considered for analysis, despite some being logged on the page as they fell outside the AOI. It was evident that sometimes the child's gaze moved outside of the AOI and these are visible on the heat maps (see figure 4) but the data was not used within any fixation counts or analysis. Fixations outside the AOI also occurred when they had finished reading the text and were turning over the pages of the paper prototype. This analysis enabled heatmaps to be generated for each image and fixation date relating to the AOIs to be exported for statistical analysis in SPSS.

In the final stage of analysis the pages on the storyboard had different lengths of text and therefore it would not be feasible to do a direct comparison between individual pages as you would expect the number of fixations to rise if the number of words increased. To mitigate this problem a ratio was produced based on the number of words per page. In text entry studies characters are counted (without the white space) and words are judged to be 5 characters (MacKenzie & Soukoreff, 2003). We chose to apply this standardised measure to each screen so we could consider, and compare, fixations per word. For example, on screen 1, there were 534 characters which translates to 107 words, participant 1, viewing the photo-realistic storyboard had 65 fixations on this area of text giving a ratio of 0.61 fixations per word. See table 5 for words per page. In addition, for each AOI on a page, a percentage was calculated for each child based on their total fixations for all pages.

4 RESULTS

Twenty-seven children participated with consent; one did not have consent but used the technology, so they did not feel left out, but no data was captured. As expected, the Splode game was new to all the children, none of them reported having seen it, or played it, before. The children were all able to manage wearing the Tobii glasses with a couple needing a bit of assistance. One child wore glasses but did not need them for reading and so took them off during the study. There were some difficulties in calibrating the glasses for a few children, for example one child's left eye could not be detected. In these instances, so as not to cause anxiety for the child, the child was reassured by the researcher and continued anyway but inevitably the data was too poor to use. None of the children asked for assistance or clarification whilst completing the survey tools, they could all complete the task independently.

4.1 Survey data

The results for the first question of the Smileyometer, after the children had viewed the 1st page of the storyboard prototype, are presented in table 1:

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

Group		Wireframe		Photo-realistic	
		Mean	SD	Mean	SD
Q1	What do you think the game will be like?	3.45	.69	3.83	.94

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

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

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

Group		Wireframe		Photo-realistic	
		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			Photo-realistic		
	Yes	Maybe	No	Yes	Maybe	No
free	N=6	N=3	N=2	N=7	N=5	N=0
0.99p	N=4	N=1	N=6	N=3	N=6	N=3

It would appear as though children would be less willing to download the game having seen it as a wireframe prototype. In answer to RQ1: These results appear to show that the different levels of aesthetic refinement did not affect children's overall rating of the game but did potentially affect whether or not they might download it.

4.2 Eye Tracking Results

Seven recordings were discarded as the sample rate was not sufficient (<80%). There were several reasons for this; for example, even though the glasses were calibrated, one child was reading underneath the glasses and therefore no data was captured. It is a common problem when using eye tracking where data loss of around 25% has been reported in eye gaze studies with children (Abe, Hamada, Nagai, Shiomi, & Omori, 2017). The consequence of this was that eye tracking data was gathered from 20 children; 10 in each condition.

RQ2 was looking for insights from the use of eye tracking technology towards understanding how children interacted with the prototypes. Fixation data from the 1st screen showed that the mean number of fixations was higher for both AOIs for the wireframe in comparison to the photo-realistic version, see table 4. A t-test revealed no significant difference for the Text AOI *t(18) = 3.639, p = .073 and t(18) = 2.948, p = .103* for the Image AOI.

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

Group	Wireframe		Photo-realistic	
	Mean	SD	Mean	SD
Text	57.30	33.61	49.60	13.11
Image	16.70	19.02	15.20	8.63

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, where the children focus on the text and navigation options, but there does appear to be 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 Photo-Realistic

Table 5 shows the means for the total number of fixations for the area of interests for the 10 pages of the each prototype. The 10 participants who received the wireframe prototype (*M* = 165.10, *SD* = 64.67) compared to the 10 participants viewing the photo-realistic 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.*

Table 5: Total Fixations for the two AOI in the Wireframe and Photo-realistic Prototype

Group	Wireframe		Photo-realistic	
	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, see table 5, the mean number of fixations was higher for the wireframe across all 9 of the 10 pages for the Image AOI and the ratio for the Text AOI was higher on 7 of the 10 pages. Page nine demonstrated the largest difference between the two groups with the Image AOI of the wireframe prototype (*M* = 35.50, *SD* = 17.28.67) compared to the photo-realistic prototype (*M* = 5.4, *SD* = 3.92). Differences in fixation could have been attributed to a difference in the amount of time the children interacted with the prototype, with the time captured in seconds, but a t-test revealed no significant difference in the time spent viewing with the prototype between the two groups. For the 10 participants who received the wireframe prototype the total time was *M* = 338.76, *SD* = 24.56 compared to the 10 participants viewing the photo-realistic prototype *M* = 285.57, *SD* = 67.05, t(18) = 1.639, p = .496.

To determine whether there was a difference in the way the children viewed the content over time, a ratio was calculated based on fixation per words for each of the 10 screens. Table 6 shows the ratio of fixation per word for the Text AOI and the mean fixations for the Image AOI for the two prototypes.

Table 6: Fixations per page based on ratio of fixation per word for the Text AOI and average fixations for the Image AOI

			Wireframe		Photo-realistic	
Page	Characters	Words per Character	Text – Mean (SD)	Image – Mean (SD)	Text – Mean (SD)	Image – Mean (SD)
1	534	107	.54 (.31)	16.70 (19.03)	.46 (.12)	15.20 (8.63)
2	84	17	.58 (.33)	19.40 (13.98)	.61 (.45)	11.00 (9.58)
3	74	15	.45 (.30)	12.20 (9.76)	.59 (.48)	12.60 (6.87)
4	73	15	.67 (.46)	12.90 (10.45)	.53 (.37)	7.60 (4.09)
5	164	33	.64 (.29)	15.20(15.47)	.51 (.27)	9.80 (6.12)
6	131	26	.52 (.38)	13.50 (7.64)	.40 (.21)	12.10 (6.04)
7	64	13	.55 (.46)	13.80 (12.39)	.23 (.28)	6.90 (4.07)
8	115	23	.73 (.38)	17.70 (12.15)	.54 (.35)	7.40 (5.42)
9	264	53	.21 (.13)	35.50 (17.28)	.42 (.22)	5.40 (3.92)
10	81	16	.71 (.29)	14.60 (9.45)	.58 (.40)	10.10(5.13)

The data shows that the number of fixations fluctuated between pages when examining the data from the two AOI. It appears that the children are engaging with the prototype across all 10 pages. Within the image AOI it does appear to decline after the first few pages and then increase again towards the end. Within the text AOI the highest ratio of fixation per word was on page 8 of the wireframe and page 2 on the photo-realistic prototype. The data would suggest that the number of characters on the page may not appear to influence the number of fixations when looking at the text AOI. It is important to note that not all children looked at the content of all the pages when examining the two prototypes. The average time to view the wireframe storyboard was Mean=339 second SD=78 second whilst the photo-realistic prototype was Mean=286 second SD=67 seconds. Although the participants on average spent more time examining the black and white prototype a Wilcoxon test revealed there was no significant difference between the two groups.

In table 7, this shows the percentage of total fixations for an individual child and the average of all the children's fixation per page and there is large individual variability. Two of the children from wireframe group had 33% and over of their total fixations in the text on the first page and only 1 participant in the photo-realistic having a similar number. Given the amount of text on page 1 it may have been anticipated that there would be a higher number of fixations on this page. This would only account for potential differences in the Text AOI.

Table 7: Percentage of fixations per page for each child in the two AOI (I Image AOI and T Text AOI)

	P1		P2		P3		P4		P5		P6		P7		P8		P9		P10	
	I	T	I	T	I	T	I	T	I	T	I	T	I	T	I	T	I	T	I	T
B1	3	36	3	3	1	4	0	1	2	6	2	7	1	5	1	8	13	1	2	4
B2	0	23	5	1	5	3	3	7	4	1	5	7	3	1	6	0	15	3	5	3
B3	5	6	8	4	7	0	6	3	4	9	7	3	7	0	4	5	4	8	5	5
B4	5	11	4	5	4	3	4	5	3	9	4	4	4	2	3	8	12	4	4	2
B5	1	33	4	4	4	2	2	3	3	5	4	6	1	3	4	5	7	4	2	5
B6	1	14	4	4	3	3	5	4	4	8	5	7	2	3	6	4	12	2	4	3
B7	4	21	4	0	1	0	1	0	2	12	2	0	6	4	6	5	17	2	5	5
B8	9	9	5	2	5	1	5	3	8	5	4	3	6	2	6	5	10	2	5	2
B9	4	13	13	3	1	2	4	4	6	5	3	3	3	0	8	4	11	6	3	3
B10	13	17	3	2	4	1	5	2	7	5	7	4	6	0	5	3	3	5	6	2
C1	4	21	2	7	4	3	1	4	3	8	3	4	2	4	1	6	3	11	3	7
C2	3	24	2	2	13	5	3	6	2	8	7	6	2	0	2	0	3	5	4	5
C3	2	38	1	4	5	1	2	0	0	4	2	3	4	2	3	10	1	12	3	4
C4	6	17	2	3	6	3	4	3	3	7	4	5	3	2	5	4	4	9	6	4
C5	8	17	9	3	5	2	2	4	3	5	4	5	3	0	5	7	3	8	4	3
C6	7	20	11	10	4	10	4	6	4	13	9	1	2	0	0	0	0	0	0	0
C7	6	8	3	5	2	4	2	4	5	7	4	6	3	2	4	8	1	14	5	6
C8	13	17	3	2	4	1	5	2	7	5	7	4	6	0	5	3	3	5	6	2
C9	6	20	9	3	4	4	6	1	9	8	2	1	3	1	3	3	0	8	5	3
C10	6	27	1	0	5	0	1	0	2	0	8	8	0	1	0	10	3	20	4	4
Ave	5	19	5	3	4	3	3	3	4	6	4	4	3	2	4	5	6	6	4	4

5 DISCUSSION

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

The first question *RQ1: When presenting a prototype to children in different levels of aesthetic refinement can this affect their overall rating of a game?* aimed to establish whether children rated the game differently based upon the fidelity of the prototype. The survey data revealed 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 the physical form factor of the prototype has on children self-reporting of a game concept that was presented on a iPad or paper, 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 (Sauer & Sonderegger, 2009) examining aesthetics in prototypes with adults, users appeared to compensate for deficiencies in aesthetic design by overrating the aesthetic qualities of reduced fidelity prototypes. More work is clearly required to understand what factors might influence and confound the results when evaluating prototypes with children.

When the data from the two conditions was aggregated, there was a significant difference in the rating of the game based on viewing the 1st page and on viewing the entire storyboard. In this instance the children rated the game higher at the end compared to their initial expectations after viewing the 1st page. This is useful for researchers as it shows not only that the children rated the game higher, in other words seeing how the game panned out was beneficial, but, because we had the eye tracking data, we understand their interactions with the prototype.

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 photo-realistic prototype. This is important as intention to play / purchase is an indicator that the children might enjoy this 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….?' Which liberates the child from judging something the 'researcher' has brought for evaluation. For this reason, we may conjecture that giving children a more visually appealing prototype might positively influence their intentions to play or download the game.

In response to RQ2: *What insights can eye tracking technology provide towards understanding how children interact with a low-fidelity prototype to help inform prototype design?* the eye tracking data that was obtained provided valuable insights into how the children interacted with the two versions of the prototype. 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 AOI in the wireframe prototype compared to the photo-realistic version, as seen in table 5. This may suggest that the wireframe prototype is more cognitively demanding due to salient differences in the visual refinement than the photo-realistic prototype, but it could also be that there may have been differences in the reading ability of the children between the two groups. Proficient readers are known to make fewer fixations than beginners (Rayner et al., 2012).

Table 7 showed the percentage of fixations in each AOI, highlighting a large proportion of fixations in the text on the 1st page with an average of 19% of total fixations. It was conjectured that children may not read the text, and this was not always the case. An explanation for this may have been if the children could understand the game mechanics from just the image, then the text would be redundant. There would be no reason for the children to read the text if they can comprehend the storyboard through just the visual stimuli. It is important to note that the images presented to the children were prominently in grey scale as the game was set at night, there was only two pages within the storyboard that consisted of bright primary colours but they did differ in contrast and luminance. Replicating the study with a different game may help understand whether the wireframe prototype is more cognitively demanding. For example, with small samples within each group, the total number of fixations may have been skewed by a few individuals within the wireframe group, see table 6 above. In total nine different children failed to look at least one of the areas of interest on at least one page. Notably C6 disengaged after the 6th page of the photo-realistic version, they were the only child who did not look through all the pages. Reviewing the eye tracking video for this participant, they appeared to flick through the remaining pages without looking at them. There would appear to be some behavioral differences on how the children interact with the content which is not consistent between pages. It is important to note that even though a small sample was used eye tracking studies involving children have tended to have small sample sizes (Sim & Bond, 2021) and conclusions can be drawn from the experimental studies.

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, Al-Ghanim, & Al-Wakeel, 2014).

6 LIMITATIONS

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 data needed to be examined after each participant to ensure an even distribution across the two conditions. Data was lost for a variety of reasons including the children moving the glasses during the study, calibration issues, light reflecting of 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, Höbel, & Nürnberger, 2014) and calibration issues (Krejtz, Krejtz, Duchowski, Szarkowska, & Walczak, 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, it would have increased the duration of the study and been more tiring, and on the ecological validity of the study, it would have made the interaction with the prototype less natural, 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 AOI, then it may have been essential to recalibrate more often given the impact of drift as reported in studies with adults (Holmqvist & Andersson, 2017).

Reading comprehension of the text within the prototype was not assessed in this study, and neither was the children's reading ability. These factors may have influenced the results and accounted for the fact there were more fixations within the wireframe group. The children viewing the wireframe prototype may have struggled to process the information to develop a mental model of the game when compared to the photo-realistic version. When pictures and text are presented dual coding is required to produce accurate mental models and the integration processes are crucial (Schnotz et al., 2014). The final limitation is the relatively low sample size of 20 children, with ten in each condition. It is unlikely that the two groups differed significantly as the teacher assigned children to each of the groups, examined the material beforehand and all the children appeared to be able to read the text, but differences cannot be ruled out.

7 CONCLUSIONS AND FUTURE WORK

This paper aimed to examine 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 wireframe and photo-realistic prototype when presented as a storyboard within a booklet but the Again Again table showed a stronger preference for the photo-realistic prototype. This highlights some of the challenges in interpreting findings when using low-fidelity paper prototypes to evaluate game concepts with children.

The eye tracking technology demonstrated that children engaged with all 10 pages of the prototypes, giving confidence that the children are reporting their perceptions of the entire game and not disengaging from the evaluation. It was conjectured that children may not read the text, but this appeared to be an important element of the storyboard and care should be taken on ensuring appropriate text is included as this may inform or influence the rating of the game. Thus, when developing games concepts the text may appear to play a crucial role in their comprehension of the game mechanics and interaction. CCI researchers can take confidence that when presenting prototypes of up to 10 pages the children are prepared to read the text description and view the images. However, the number of fixations was higher on the wireframe prototype suggesting this may potentially have higher cognitive demands. Further work is required to understand the reason behind these differences, and it would be interest to compare adult behaviour to children's when viewing the prototypes. In addition, given the potential importance text may play in the decision when children are presented with a prototype, evaluating children's comprehension may offer insights into understand and how this influences their ratings of a game. 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.

References

Abe, K., Hamada, Y., Nagai, T., Shiomi, M., & Omori, T. (2017). *Estimation of child personality for child-robot interaction*. Paper presented at the 26th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN).

Al-Wabil, A., Al-Husian, L., Al-Murshad, R., & Al-Nafjan, A. (2010). *Applying the retrospective think-aloud protocol in usability evaluations with children: Seeing through children's eyes*. Paper presented at the 2010 International Conference on User Science and Engineering (i-USEr).

Al-Zeer, S., Al-Ghanim, A., & Al-Wakeel, L. y. (2014). *Visual attention in interaction with Arabic augmentative and alternative communication apps*. Paper presented at the 2014 3rd International Conference on User Science and Engineering (i-USEr).

Ball, L. J., & Richardson, B. H. (2022). Eye Movement in User Experience and Human–Computer Interaction Research. In *Eye Tracking: Background, Methods, and Applications* (pp. 165-183): Springer.

Bekker, T., & Antle, A. N. (2011). *Developmentally situated design (DSD): making theoretical knowledge accessible to designers of children's technology*. Paper presented at the Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, Vancouver, BC, Canada.

Bertou, E., & Shahid, S. (2014). *Lowfidelity prototyping tablet applications for children*. Paper presented at the Proceedings of the 2014 conference on Interaction design and children, Aarhus, Denmark.

Bilal, D., & Gwizdka, J. (2016). *Children's eye-fixations on google search results*. Paper presented at the Proceedings of the 79th ASIS&T Annual Meeting: Creating Knowledge, Enhancing Lives through Information and Technology, Copenhagen, Denmark.

Brooks, J., & Lopes, P. (2023). *Smell & Paste: Low-Fidelity Prototyping for Olfactory Experiences*. Paper presented at the Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems, Hamburg, Germany.

Buxton, B. (2010). *Sketching user experiences: getting the design right and the right design*: Morgan kaufmann.

Chen, Y., & Tsai, M.-J. (2015). Eye-hand coordination strategies during active video game playing: An eye-tracking study. *Computers in Human Behavior, 51*, 8-14.

Christoforakos, L., & Diefenbach, S. (2019). *Idealization effects in UX evaluation at early concept stages: challenges of low-fidelity prototyping.* Paper presented at the Advances in Usability, User Experience and Assistive Technology: Proceedings of the AHFE 2018 International Conferences on Usability & User Experience and Human Factors and Assistive Technology, Held on July 21–25, 2018, in Loews Sapphire Falls Resort at Universal Studios, Orlando, Florida, USA 9.

Currie, J., Bond, R. R., McCullagh, P., Black, P., Finlay, D. D., & 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.

Damle, A., & Smith, P. J. (2009). Biasing cognitive processes during design: the effects of color. *Design Studies, 30*(5), 521-540.

Derboven, J., De Roeck, D., Verstraete, M., Geerts, D., Schneider-Barnes, J., & Luten, K. (2010). *Comparing User Interaction with Low and High Fidelity Prototypes of Tabletop Surfaces*. Paper presented at the NordiCHI, Reykjavik, Iceland.

Fischer, P. M., Richards, J. W., Berman, E. J., & Krugman, D. M. (1989). Recall and eye tracking study of adolescents viewing tobacco advertisements. *Jama, 261*(1), 84-89.

Frank, M. C., Vul, E., & Saxe, R. (2012). Measuring the development of social attention using free-viewing. *Infancy, 17*(4), 355-375.

Giannakos, M. N., Papavlasopoulou, S., & Sharma, K. (2020). Monitoring Children's Learning Through Wearable Eye-Tracking: The Case of a Making-Based Coding Activity. *IEEE Pervasive Computing, 19*(1), 10-21.

Gollety, M., & Guichard, N. (2011). The dilemma of flavor and color in the choice of packaging by children. *Young Consumers, 12*(1), 82-90.

Gomes, A., Oh, H., Chisik, Y., & Chen, M. (2012). *Ilha Musical: a CAVE for nurturing cultural appreciation*. Paper presented at the Proceedings of the 11th International Conference on Interaction Design and Children, Bremen, Germany.

Gossen, T., Höbel, J., & Nürnberger, A. (2014). *Usability and perception of young users and adults on targeted web search engines*. Paper presented at the Proceedings of the 5th Information Interaction in Context Symposium, Regensburg, Germany.

Guha, M. L., Druin, A., Chipman, G., Fails, J. A., Simms, S., & Farber, A. (2004). *Mixing ideas: a new technique for working with young children as design partners.* Paper presented at the Proceedings of the 2004 conference on Interaction design and children: building a community, Maryland, USA.

Gulz, A., Londos, L., & Haake, M. (2020). Preschoolers' understanding of a teachable agent-based game in early mathematics as reflected in their gaze behaviors–an experimental study. *International Journal of Artificial Intelligence in Education, 30*, 38-73.

Gwizdka, J. (2019). *Exploring eye-tracking data for detection of mind-wandering on web tasks.* Paper presented at the Information Systems and Neuroscience: NeuroIS Retreat 2018, Vienna, Austria.

Hanna, L., Neapolitan, D., & Risden, K. (2004). *Evaluating computer game concepts with children.* Paper presented at the Proceedings of the 2004 conference on Interaction design and children: building a community, Maryland, USA.

Hannus, M., & Hyönä, J. (1999). Utilization of illustrations during learning of science textbook passages among low-and high-ability children. *Contemporary Educational Psychology, 24*(2), 95-123.

Hershman, A., Nazare, J., Qi, J., Saveski, M., Roy, D., & Resnick, M. (2018). *Light it up: Using paper circuitry to enhance low-fidelity paper prototypes for children.* Paper presented at the Proceedings of the 17th ACM conference on interaction design and children, Trondheim, Norway.

Hirschmanner, M., Lammer, L., & Vincze, M. (2015). *Mattie: A simple educational platform for children to realize their first robot prototype.* Paper presented at the Proceedings of the 14th international conference on interaction design and children, Aarhus, Denmark.

Hoggenmüller, M., Tomitsch, M., Hespanhol, L., Tran, T. T. M., Worrall, S., & Nebot, E. (2021). *Context-Based Interface Prototyping: Understanding the Effect of Prototype Representation on User Feedback.* Paper presented at the Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems, Yokohama, Japan.

Hollander, J., & Huette, S. (2022). Extracting blinks from continuous eye-tracking data in a mind wandering paradigm. *Consciousness and Cognition, 100*, 103303.

Holmqvist, K., & Andersson, R. (2017). Eye tracking: A comprehensive guide to methods. *paradigms and measures*.

Huestegge, L., Radach, R., Corbic, D., & Huestegge, S. M. (2009). Oculomotor and linguistic determinants of reading development: A longitudinal study. *Vision research, 49*(24), 2948-2959.

Just, M. A., & Carpenter, P. A. (1980). A theory of reading: from eye fixations to comprehension. *Psychological Review, 87*(4), 329.

Kara, N., & Cagiltay, K. (2020). Smart toys for preschool children: A design and development research. *Electronic Commerce Research and Applications, 39*, 100909.

Koch, C., & Ullman, S. (1985). Shifts in selective visual attention: towards the underlying neural circuitry. *Human neurobiology, 4*(4), 219-227.

Kohler, B., Haladjian, J., Simeonova, B., & Ismailovic, D. (2012). *Feedback in Low vs High Fidelity Visuals for Game Prototypes*. Paper presented at the Games and Software Engineering, Zurich.

Krejtz, K., Krejtz, I., Duchowski, A., Szarkowska, A., & Walczak, A. (2012). *Multimodal learning with audio description: an eye tracking study of children's gaze during a visual recognition task*. Paper presented at the Proceedings of the ACM Symposium on Applied Perception, Los Angeles, California.

Krishnakumar, S., Berdanier, C., Lauff, C., McComb, C., & Menold, J. (2022). The story novice designers tell: How rhetorical structures and prototyping shape communication with external audiences. *Design Studies, 82*, 101133.

Lee, C. P., Cagiltay, B., & Mutlu, B. (2022). *The unboxing experience: Exploration and design of initial interactions between children and social robots.* Paper presented at the Proceedings of the 2022 CHI conference on human factors in computing systems.

Levi, M. D., & Conrad, F. r. G. (1996). A heuristic evaluation of a world wide web prototype. *Interactions, 3*(4), 50-61.

Lim, Y. K., Pangam, A., Periyasami, S., & Aneja, S. (2006). *Comparative Analysis of High- and Low-fidelity Prototypes for More Valid Usability Evaluations of Mobile Devices*. Paper presented at the NordiCHI, Oslo, Norway.

MacKenzie, I. S., & Soukoreff, R. W. (2003). *Phrase sets for evaluating text entry techniques.* Paper presented at the CHI'03 extended abstracts on Human factors in computing systems, Fort Lauderdale, Florida.

Marcel, A., Just. , & Patricia, A., Carpenter. (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., & 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.

Mason, L., Pluchino, P., Tornatora, M. C., & Ariasi, N. (2013). An eye-tracking study of learning from science text with concrete and abstract illustrations. *The Journal of Experimental Education, 81*(3), 356-384.

Mathis, F., O'hagan, J., Vaniea, K., & Khamis, M. (2022). *Stay home! conducting remote usability evaluations of novel real-world authentication systems using virtual reality.* Paper presented at the Proceedings of the 2022 International Conference on Advanced Visual Interfaces, Rome, Italy.

Mayer, R. E. (2005). Cognitive theory of multimedia learning. *The Cambridge handbook of multimedia learning, 41*, 31-48.

Mayer, R. E. (2014). Introduction to multimedia learning.

Moeller, K., Neuburger, S., Kaufmann, L., Landerl, K., & Nuerk, H. C. (2009). Basic number processing deficits in developmental dyscalculia: Evidence from eye tracking. *Cognitive Development, 24*(4), 371-386.

Nebeling, M., & Madier, K. (2019). *360proto: Making interactive virtual reality & augmented reality prototypes from paper.* Paper presented at the Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, Glasgow, Scotland.

Niehorster, D. C., Santini, T., Hessels, R. S., Hooge, I. T., Kasneci, E., & Nyström, M. (2020). The impact of slippage on the data quality of head-worn eye trackers. *Behavior research methods, 52*, 1140-1160.

Othman, A., & Mohsin, M. (2017, 19-21 Dec. 2017). *How could robots improve social skills in children with Autism?* Paper presented at the 6th International Conference on Information and Communication Technology and Accessibility (ICTA), Muscat, Oman.

Papavlasopoulou, S., Sharma, K., & Giannakos, M. N. (2020). Coding activities for children: Coupling eye-tracking with qualitative data to investigate gender differences. *Computers in Human Behavior, 105*, 105939.

Piaget, J. (1970). Piaget's theory.

Potvin Kent, M., Pauzé, E., Roy, E. A., de Billy, N., & Czoli, C. (2019). Children and adolescents' exposure to food and beverage marketing in social media apps. *Pediatric obesity, 14*(6), e12508.

Rante, H., & Basuki, A. (2019). *The Role of Paper Prototyping in Designing Visual Novel Game as Learning Media for Children.* Paper presented at the 5th International Conference on Education and Technology (ICET), Kota Batu, Jawa Timur, Indonesia.

Rayner, K., Pollatsek, A., Ashby, J., & Clifton Jr, C. (2012). Psychology of reading.

Read, J. C. (2012). *Evaluating artefacts with children: age and technology effects in the reporting of expected and experienced fun.* Paper presented at the 14th ACM International Conference on Multimodal Interaction, Santa Monica.

Read, J. C., & Bekker, M. M. (2011). *The naure of child computer interaction*. Paper presented at the 25th BCS Conference on Human Computer Interaction, Northumbria.

Read, J. C., & MacFarlane, S. (2006). *Using the Fun Toolkit and Other Survey Methods to Gather Opinions in Child Computer Interaction*. Paper presented at the Interaction Design and Children, Tampere.

Read, J. C., MacFarlane, S., & Casey, C. (2002). *Endurability, Engagement and Expectations: Measuring Children's Fun*. Paper presented at the Interaction Design and Children, Eindhoven.

Rehring, K., & Ahlemann, F. (2020). *Evaluating the User Experience of an Augmented Reality Prototype for Enterprise Architecture.* Paper presented at the Wirtschaftsinformatik (Zentrale Tracks).

Reilly, D., Dearman, D., Welsman-Dinelle, M., & Inkpen, K. (2005). Evalusating Early Prototypes in Context: Trade-offs, Challenges, and Successe. *Pervasive Computing, 4*(4), 42-50.

Rider, A. T., Coutrot, A., Pellicano, E., Dakin, S. C., & Mareschal, I. (2018). Semantic content outweighs low-level saliency in determining children’s and adults’ fixation of movies. *Journal of Experimental Child Psychology, 166*, 293-309.

Rudd, J., Stern, K., & Isesensee, S. (1996). Low vs high-fidelity prototyping debate. *Interactions, 3*(1), 75-85.

Sabuncuoglu, A., & Sezgin, T. M. (2022). *Prototyping Products using Web-based AI Tools: Designing a Tangible Programming Environment with Children.* Paper presented at the 6th FabLearn Europe/MakeEd Conference 2022, Aalborg, Denmark.

Sauer, J., & 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.

Schnotz, W., Ludewig, U., Ullrich, M., Horz, H., McElvany, N., & Baumert, J. (2014). Strategy shifts during learning from texts and pictures. *Journal of Educational Psychology, 106*(4), 974.

Sefelin, R., Tscheligi, M., & Giller, V. (2003). *Paper Prototyping - What is it good for? A comparison of Paper and Computer based Low fidelity prototyping*. Paper presented at the Conference on Human Factors in Computing Systems, Fort Lauderdale.

Sharma, K., Papavlasopoulou, S., & Giannakos, M. (2019). Coding games and robots to enhance computational thinking: How collaboration and engagement moderate children’s attitudes? *International Journal of Child-Computer Interaction, 21*, 65-76.

Sim, G., & Bond, R. (2021). Eye tracking in child computer interaction: Challenges and opportunities. *International Journal of Child-Computer Interaction*(30), 100345.

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. (2012). *Investigating children's opinions of games: Fun Toolkit vs This or That*. Paper presented at the Interaction Design and Children, Bremen, Germany.

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, Manchester, UK.

Snyder, C. (2003). *Paper Prototyping: The fast and Easy Way to Define and Refine User Interfaces*. San Francisco: Morgan Kaufmann.

Su, Q., Chen, F., Li, H., Yan, N., & Wang, L. (2018, 3-6 Dec. 2018). *Multimodal Emotion Perception in Children with Autism Spectrum Disorder by Eye Tracking Study.* Paper presented at the 2018 IEEE-EMBS Conference on Biomedical Engineering and Sciences (IECBES), Sarawak, MALAYSIA.

Thomsen, S. R., & Fulton, K. (2007). Adolescents’ Attention to Responsibility Messages in Magazine Alcohol Advertisements: An Eye-Tracking Approach. *Journal of Adolescent Health, 41*(1), 27-34.

Toolaroud, P. B., Nabovati, E., Mobayen, M., Akbari, H., Feizkhah, A., Farrahi, R., & Jeddi, F. R. (2023). Design and usability evaluation of a mobile-based-self-management application for caregivers of children with severe burns. *International wound journal, 20*(7), 2571-2581.

Torney, H., Harvey, A., Finlay, D., Magee, J., Funston, R., & Bond, R. R. (2018). *Eye-tracking analysis to compute the visual hierarchy of user interfaces on automated external defibrillators.* Paper presented at the British HCI Conference 2018, Belfast, UK.

Underwood, G. (2009). Cognitive processes in eye guidance: algorithms for attention in image processing. *Cognitive Computation, 1*, 64-76.

Underwood, G., & Everatt, J. (1992). The role of eye movements in reading: some limitations of the eye-mind assumption. In *Advances in psychology* (Vol. 88, pp. 111-169): Elsevier.

Vajs, I., Ković, V., Papić, T., Savić, A. M., & Janković, M. M. (2022). *Dyslexia detection in children using eye tracking data based on VGG16 network.* Paper presented at the 2022 30th European Signal Processing Conference (EUSIPCO), Belgrade, Serbia.

Velazquez, C. E., & 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., & Karis, D. (1996). *Usability problem identification using both low- and high-fidelity prototypes*. Paper presented at the Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, Vancouver, British Columbia, Canada.

Volioti, C., Tsiatsos, T., Mavropoulou, S., & Karagiannidis, C. (2016). VLEs, social stories and children with autism: A prototype implementation and evaluation. *Education and Information Technologies, 21*, 1679-1697.

von Reumont, F., & Budke, A. (2020). Strategies for Successful Learning with Geographical Comics: An Eye-Tracking Study with Young Learners. *Education Sciences, 10*(10), 293-330.

Walker, M., Takayama, L., & Landay, J. A. (2002). *High-fidelity or low-fidelity, paper or computer? Choosing attributes when testing web prototypes.* Paper presented at the Proceedings of the human factors and ergonomics society annual meeting, Maryland, USA.

Wertli, J., Schötzau, A., & Palmowski-Wolfe, A. (2023). The Influence of Age on Eye Movements during Reading in Early Elementary School Children. *Klinische Monatsblätter für Augenheilkunde, 240*(04), 591-598.

Wiklund, M., Thurrot, C., & Dumas, J. (1992). *Does the Fidelity of Software Prototypes Affect the Perception of Usability*. Paper presented at the Proceedings of the Human Factors and Ergonomics Society Annual Meeting, Atlanta, USA.

Zaman, B., & Abeele, V. V. (2010). *Laddering with Young Children in User Experence Evaluations: Theoretical Groundings and a Practical Case*. Paper presented at the IDC, Barcelona.

Zaman, B., Abeele, V. V., & De Grooff, D. (2013). Measuring product liking in preschool children: An evaluation of the Smileyometer and This or That methods. *International Journal of Child-Computer Interaction, 1*(2), 61-70.

Figure 1: Splode Game

166x95mm (144 x 144 DPI)

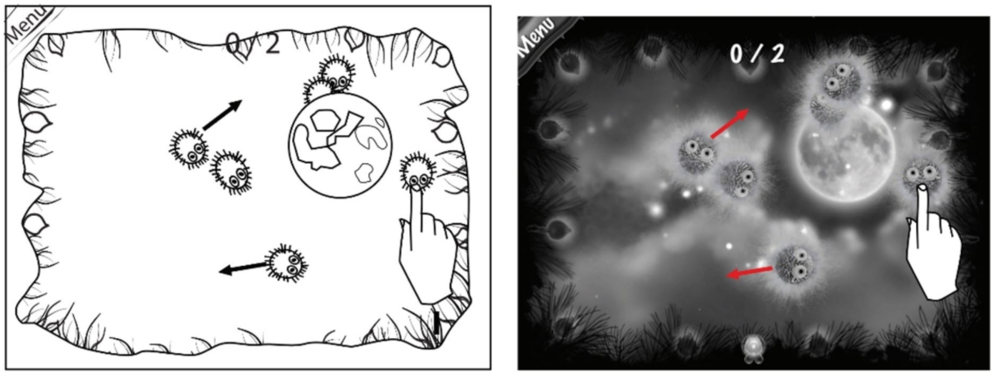

Figure 2: Wireframe (left) and Photo-realistic (right) prototypes of the Game

333x126mm (144 x 144 DPI)

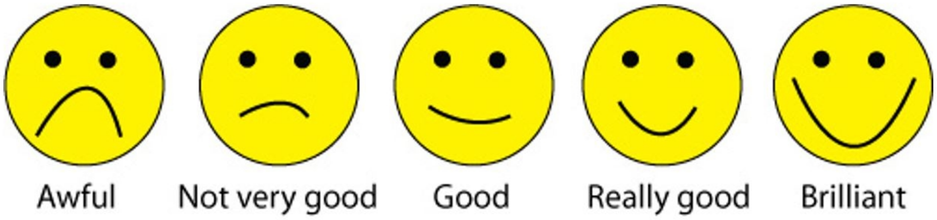

Figure 3: The Smileyometer

234x65mm (144 x 144 DPI)

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

146x223mm (144 x 144 DPI)