# **Derivation of Young Children’s Interaction Strategies with Digital Educational Games from Gaze Sequences Analysis.**

**ABSTRACT**: Digital Education Games (DEGs) have been used to support children’s learning in various domains. A number of existing studies on DEGs has focused on whether they could improve children’s learning performance. However, only a few of them have attempted to address the critical question of how young children interact with DEGs. Bridging this gap was the main motivation underpinning this research study. With the use of eye-tracking technology, we explored our research goal by evaluating a bespoke DEG on numeracy and its cardboard version that we developed based on the UK Early Years Foundation Stage (EYFS) framework. A between-subject experiment study involving 94 five-year-olds was conducted. The research protocols and instruments were pilot tested and ethically approved. In analysing the eye-tracking data, we refined the *Gaze Sub-sequence Marking Scheme* to infer children’s interaction strategies. Results showed that the difference in the learning effect between the digital and cardboard game was insignificant, that the children’s interaction strategies varied significantly with their achievement level, and that children’s gender was not a significant factor in determining the impact of learning with the DEG. Implications for rendering eye-tracking technology more child-friendly and designing DEGs for young children are drawn.

**Keyword:** Human-computer Interface; Digital Educational Games; Early Years Education; Interaction Strategies; Gaze sequence analysis.

1. **Introduction**

Digital educational games (DEGs) have increasingly been used to support young children to learn (Neumann and Neumann, 2014), thanks to their motivational power and to the proliferation of digital devices, which have become more affordable and usable (Shuler et al., 2012). For instance, a study on the effect of using the DEGs - *Endless Alphabet* and *Letter School* - to enhance the literacy of children aged two to five years old reported positive results (Neumann, 2018). Another study aimed to improve the motivation of young children for writing with the DEG *AZBUKA* also reported encouraging outcomes (Duh et al., 2017).

While these and a handful of other studies (Aladé et al., 2016; Burnett, 2010; Peirce and Centre, 2013) focused on *whether* DEGs could help improve learning performance of young children, which is a meaningful and important research goal per se, very few attempts have been undertaken to understand *how* children interact and learn with DEGs. This understanding is critical as it can not only inform the design of DEGs to enhance their impact but also gain insights into teaching strategies for this age group. A plausible reason for the scantiness of research on this specific area is the methodological challenge of working with young children whose verbal ability is in general low (Parker et al., 2013), making it difficult to utilise research methods like think-aloud or interview. A viable alternative approach to understanding young children’s interaction with a DEG is *eye-tracking*.

The recent development of eye-tracking technology, especially its improved usability and applicability for mobile devices (Holmqvist et al., 2011; Jarodzka et al., 2017), has led to the notable increase of its usage in the field of Human-Computer Interaction (HCI) in general (e.g., usability testing (Wang et al., 2019), security (Katsini et al., 2020), Section 2.2), and has motivated us to apply this technology to study young children’s interaction with DEG in particular. Nonetheless, child-friendly equipment and practical guidelines for implementing such studies remain limited (Section 3.4). We aimed to address this issue with our hands-on experience of conducting our study, which was aimed to assess the learning effect of a bespoke DEG on numbers 1 to 20. Specifically, we investigated whether the DEG could improve five-year-olds’ numeracy and how two related factors - gender and prior knowledge - would influence the effect. Our particular focus on gender has been stimulated by the related work on studying the development of gender differences in mathematics education from an early age; the findings have been inconclusive (Section 2.3). We were motivated to explore whether the learning medium played a role in determining any difference.

More intriguing would be, through eye-tracking data, to identify interaction strategies young children applied when playing the DEG, which we developed with reference to the Early Years Foundation Stage (EYFS) framework in the UK educational system. Requirements for the DEG design were derived from multiple sources, theoretical as well as empirical (Section 3.3). One important observation was that learning materials in physical forms were predominantly used in classrooms, implying that it would be relevant to compare this conventional teaching approach with the alternative DEG-based one. This consideration contributed to our research design (Section 3). Early game prototypes and experimental setups were evaluated in a preliminary test and a pilot study with 5 and 31 children at their home and school, respectively. For the main study, 94 five-year-old children from two other schools were involved. A typical pretest-posttest experimental design was used to assess the learning effect. While questionnaires for evaluating gaming experience with young children exist (e.g. Read & MacFarlane, 2006), the issue of social desirability remains hard to address (Markopoulos et al., 2008). Given this reason and the concern of prolonging an experimental session with young children, who typically have a short attention span, we decided not to use such a questionnaire but to rely on observations and videos, whose details are not reported in this paper in order not to dilute its focus or lengthen it.

It is theoretically well-grounded that *fixation duration* as a prototypical eye-tracking measure is a good proxy for estimating attention paid to an area of interest (Reingold et al., 2001; Steichen et al., 2014). Long fixation duration can indicate depth of cognitive processing (Kuperman et al., 2008). Indeed, we analysed fixation durations to infer how intense children focused on direct (e.g. answer boxes) and indirect (e.g. static avatar) objects of the DEG (Mohd Nizam and Law, 2018)[[1]](#footnote-1). While there were some interesting findings such as the role of children’s prior knowledge in varying fixation durations, these static aggregated measures, which merely showed the amount of time that the children gazed at certain areas, could not reflect their cognitive strategies for interacting with the game. Consequently, we applied an alternative and more informative eye-tracking measure – *scanpaths*.

A scanpath can be defined as a “*repetitive sequence of saccades and fixations, idiosyncratic to a particular subject and to a particular target pattern*” (Choi et al., 1995). Scanpath analysis is essentially *gaze sequences analysis*, extracting information not only which objects but also the order they are viewed, and can thus be used to identify cognitive models (Henderson and Hollingworth, 2003). Recently, increasingly sophisticated methods for scanpath analysis such as String-Edit algorithm (Eraslan et al., 2016; Le Meur and Baccino, 2013) and visualisation have been developed, contributing to the creation of software applications such as *eyePatterns* (West et al., 2006).

Overall, in the light of the gap identified in the child-game interaction research, namely the lack of understanding of young children’s cognitive strategies for playing DEGs, and the viability of eye-tracking technology for bridging the gap, we aimed to address the following research questions (RQs) with empirical data:

1. To what extent does thelearning effect of a DEG on numeracy differ from its physical (cardboard) version?
2. How are children’s interaction strategies derived from gaze sequence analyses related to the learning effect induced by both game learning media?
3. How are children’s attributes - gender and prior knowledge - related to the learning effect induced by both game learning media and to the cognitive strategies?

The main contribution of our work is threefold:

* To demonstrate how gaze sequence analysis can be applied to infer cognitive strategies that children have when interacting with a DEG
* To provide further empirical evidence about the role of gender in the learning effect and cognitive strategies, drawing implications for DEG design.
* To provide practice insights in applying eye-tracking technology to young children, contributing to future research methodology on children-computer interaction.

1. **Related Work**

Three major strands of research - digital game-based learning, eye tracking, gender and technology - are highly relevant to our work. In this section, we present the reviews of the selected literature, which has inspired us to explore the topics of interest.

* 1. **Digital Game-based Learning (DGBL) Research**
     1. *DGBL Conceptual Models*

In designing DEGs for young children, it is essential to ground them in pedagogical approaches focusing on child development and game feature models. Specifically, Jean Piaget’s constructivist theory of learning and play are particularly relevant (Doherty and Hughes, 2009; Piaget, 1971). According to Piaget’s stage-based cognitive development model, children of five-years old - the target group of our research - are in the *preoperational* stage when they start to recognise symbols and use language to understand the world although their capability of logical thinking is still low (Johnston, 2005). In accord with Piaget’s theory, play is critical means for children to explore their environments. At the age of five, children can engage in symbolic and pretend play for acquiring experiences to construct knowledge whereas at the age of seven, they can enhance social skills through games. These two stages of development are marked by the two closely related but distinguishable concepts. Play is open to enable make-believe and world-building whereas games are constrained by rules and negotiated by strategies (Kampmann Walther, 2003); both play and games should have fun.

Nonetheless, Piaget’s developmental view on play, despite its popularity, has been challenged (e.g. Glenn et al., 2013; Sutton-Smith, 1966). Among others, one criticism is its heavy focus on the cognitive implication of play while neglecting, if not more important, the emotional one. This resonates with Sutton-Smith (2006)’s arguments pertaining to the first of his seven rhetorics of play, namely *the rhetoric of play as progress*. Accordingly, this rhetoric refers to the belief that children adapt and develop through play, contributing to their social, moral and cognitive competence. This notion, while embraced by many educators, should have emphasized fun and enjoyment as well (Sutton-Smith, 1966, 2006). Nevertheless, the prevailing work on game-based learning has complemented this gap, which, however, entails more empirical work to demonstrate the related theoretical assumptions.

A game is a system characterised and constituted by a set of features or elements (Fullerton, 2008; Salen and Zimmerman, 2004). The inclusion of specific features in a game defines its type, for instance, a story is essential for a role-play game but not for a mini-game (Heintz and Law, 2018). For a DEG with the dual aim of teaching and entertaining, it is critical to integrate learning content into game features that can support learning. (Wilson et al., 2008) identified a comprehensive list of such features, which, however, were not structured, making it hard to apply them (Bedwell et al., 2012). A recent attempt to produce such a structure resulted in the Game Elements-Attributes Model (GEAM; Heintz & Law, 2015), which was derived from reviewing the earlier work on motivation (Malone and Lepper, 1987) up to the contemporary game design literature (Adams, 2014). The ten game elements in GEAM are: Player, Input/Output, Rule, Perspective, Setting, Action, Challenge, Goal, Reward, and Story (cf. Prensky's (2001) six basic game features), with each associating with a set of attributes. The selection and combination of game elements and attributes are determined by several factors such as research goal, target group and learning domain (Heintz and Law, 2018).

* + 1. *Benefits of DGBL and influencing factors*

There exist a range of empirical studies on DGBL with different foci. While some research studies focused on the effectiveness of DGBL, others examine specific factors influencing the impact. Overall, DGBL proved to be a beneficial approach, as shown in the following literature review.

Furió, Juan, Seguí and Vivó (2015) investigated with children of eight to ten years old the efficiency and satisfaction of learning the water cycle with a DEG on mobile phone in contrast to traditional classroom learning. While both approaches showed positive learning outcomes, the motivational effect was found to be higher for the DEG group. Papastergiou (2009) studied the impact of a computer memory concept game with high school students, and reported encouraging results on the learning motivation and effectiveness. Sung and Hwang (2013), in analysing the impact of playing DEGs on sciences in a collaborative learning environment on sixth-grade students, reported improvements in the students’ learning attitude, motivation, achievement and self-efficacy.

Tsai, Yu and Hsiao (2012) identified four sixth graders’ attributes - learning ability, playing skills, prior knowledge and game experience – that could influence the learning effectiveness of DEGs. Ku and colleagues (2014) reported on the significant increase in fourth graders’ confidence in mathematics as a result of learning the subject with a DEG as compared with a paper-based approach. Hung, Kuo, Sun and Yu (2014) studied the effect of different difficulty levels of a DEG for learning geographical concepts for third graders. They found that the DEG led to significant learning improvement and that moderate difficulty level contributed to better performance. On the other hand, Erhel and Jamet (2013) investigated the effect of different instruction types of a DEG with university students. Results showed the learning instruction enabled deeper learning as compared to the entertainment instruction, which, however, could also lead to deep learning if students were provided frequent feedback during the gameplay.

* 1. **Eye Tracking Research**
     1. *Eye-tracking applications*

In the field of Human-Computer Interaction (HCI) and Educational Science, eye tracking has increasingly been used (Jarodzka et al., 2017; Majaranta and Bulling, 2014). Previous research with eye tracking was primarily restricted to desktops and lab-based settings. However, as technologies become increasingly mobile, so does the equipment for eye tracking (Bulling et al., 2011). Capturing participants’ eye movements in natural conversations and behaviours is the primary benefit of mobile eye tracking (Bulling and Gellersen, 2010). Thanks to recent years of active research on eye-tracking methodologies, improvements in performance, usability and affordability of mobile eye-tracking devices have been witnessed (Holmqvist et al., 2011). At the same time, the quality of software packages for automatic analysis of eye movements has also been improved. These advanced features of eye-tracking technology have stimulated as well as supported research on cognition and learning (Bulling and Roggen, 2011; Rayner, 2009).

A number of eye tracking studies has been conducted to investigate cognitive processes. Liu and Chuang (2011) employed eye-tracking methods to investigate college students’ cognitive process when viewing multimedia elements on a webpage. They found that attention was paid more to text elements and concluded that eye-tracking technology could promote insights into cognitive process. Some researchers utilised eye-tracking technology to investigate differences in cognitive process of eleventh-grade students when viewing pictures with different abstractness (Mason et al., 2013a). She and Chen (2009) applied eye-tracking to examine the cognitive process of middle-school children using different interaction modalities for learning science materials. Results confirmed the attentional benefit of multimedia and the significant relation between fixation duration and depth of learning.

Ponce and Mayer (2014), based on eye-tracking data, identified two types of gaze pattern, suggesting two learning strategies – linear and generative - used by students in different study activities. Mason, Tornatora and Pluchino (2013) identified three visual behaviour patterns that were applied by fourth graders when reading science text. Tsai and colleagues (2012) utilised eye-tracking to examine scan patterns used by university students when solving multiple-choice science questions, and found that successful solvers focused on relevant objects whereas unsuccessful ones looked around. Overall, eye tracking technology proved to be useful for gaining insights into cognitive process in learning, especially gaze data can be analysed to infer cognitive states of learners.

Nonetheless, the number of eye-tracking studies on young children using multimedia learning remains limited, despite the growing number of young multimedia users (Mayer, 2017). Only 4% of eye-tracking studies were used with kindergarten and elementary children while most 77% were used with college-age students (Alemdag and Cagiltay, 2018). Our research could enrich the body of applied knowledge of applying eye-tracking technology to understand young children’s interaction with DEG as a form of multimedia learning.

* + 1. *Eye-tracking metrics*

Common eye-tracking indicators are fixation, saccade, heat map and scanpath (or gaze sequence). Fixation refers to the static position of an eye on a specific area being viewed whereas saccades are quick movements of the eye from one fixation to another in a sequence (Schall & Bergstrom, 2014). These two indicators are categorised as *synchronic* as they show events occurring at a specific point of time (Le Meur and Baccino, 2013). Fixation duration is proven as a good proxy for attention paid to objects of interest (Steichen et al., 2014) and are related to the depth of cognitive processing (Eraslan et al., 2016).

Heat maps visualise fixations gathered over a group of participants to indicate the distribution of focused attention at a time; they are static with no information about the order that stimuli are looked at (Rösler, 2012). In contrast, scanpaths are the combination of both static and dynamic aspects of eye movements that create a complete sequence (Poole and Ball, 2006). These indicators are categorised as *diachronic* as they take time into account, and are the least studied category of eye-tracking studies (Le Meur and Baccino, 2013).

While there exist some studies that have applied different eye-tracking metrics, the use of scanpaths remains low. For instance, Al-Wabil et al. (2010) focused on the use of fixation duration and count, and correlated them with subjective experiential measures. Birkett et al. (2011) reported their literature review on the use of eye-tracking methods with young children, and commented that, in addition to other eye-tracking measures, scanpath “may also be revealing” (p.2253) for evaluating the ease of locating a target. Papavlasopoulou et al. (2018) applied fixation duration, saccade amplitude and direction with 44 children of 8-17 years old to evaluate their coding attitudes; their research goal, target group and measures were different from ours. While none of these studies applied scanpath as their eye-tracking measure, they attested the potential use of scanpath as a research tool with young children.

Gaze sequences derived from scanpaths can be used to understand cognitive processes (West et al., 2006). A range of techniques and software tools have been developed to support scanpaths/gaze sequences analysis (see the review of Eraslan et al., 2016). For our research study, we applied position-based weighted model (Sutcliffe and Namoun, 2012) and *eyePatterns* (West et al., 2006) to analyse gaze sequences young children had when interacting with a bespoke DEG.

* 1. **Gender and Technology**

In the field of educational psychology, a plethora of research studies have been conducted to understand the development of gender differences in mathematics education from both the cognitive and emotional perspective (Arroyo et al., 2013). In earlier research, basic spatial ability (e.g. mental rotation) and verbal ability (e.g. retrieval of arithmetic facts) pertinent to mathematical tasks were found to have significant gender differences, with the former being stronger in boys (Hyde et al., 2008) and the latter in girls (Voyer and Voyer, 2014). However, additional empirical results (e.g., Arroyo, Royer, & Woolf, 2011; Feng, Spence, & Pratt, 2007) counter-argued that such differences could be reduced or even eliminated through exposing children of both genders to relevant activities (e.g. video games for enhancing spatial cognition). Similarly, children's feelings towards mathematics can be shaped by their learning experiences in schools, especially feedback from teachers and peers who may hold stereotypical views on gender-based mathematical abilities (Catsambis, 2004).

With the increasing use of digital technologies in STEM education (Aladé et al., 2016), whether the aforementioned observations on gender differences have changed is an intriguing topic to explore. Several studies on gender differences involving technology have been conducted (Jackson et al., 2008; Snell and Snell-siddle, 2013). An extensive research project was carried out by Arroyo et al.(2013) on the role of gender in learning mathematics via an adaptive digital tutoring system with altogether about 500 adolescents in different schools. They found significant gender differences in the affective aspect of learning. Specifically, female students sought and accepted help provided by the tutoring system more often than their male counterparts did. Female students also performed better in the presence of a same-sex digital learning companion, which, however, was rejected by male students.

While sharing the goal of Arroyo et al.(2013) in studying the development of gender differences, Sullivan and Bers (2016) in their pilot study focused on different age groups, children from kindergarten to second grade. They reported that at such a young age children already formed opinions on the suitability of certain technological tools for a specific gender, and that boys performed significantly better than girls in tasks on more difficult programming concepts. However, these findings could be challenged because of some methodological issues in Sullivan and Bers (2016). They interviewed the young children for their knowledge of and attitude towards robotics and programming in unstructured and play-based contexts. In fact, the appropriateness of using their interview method with the kindergarteners is questionable, because there was no objective pre-test to provide a baseline for comparing the post-test performance. For drawing more solid conclusions, more work beyond the pilot study reported in Sullivan and Bers (2016) is required.

However, in studying the potential gender effect of benefiting from DEGs on computer science for high school students, Papastergiou (2009) reported no significant differences in learning gain between the two genders. The same observation of non-significant gender effect was also reported in a study with pre-schoolers learning through touchscreen applications (Moser et al., 2015). Overall, the number of research studies on gender differences in the learning effect of educational technologies with young children remains low (Oliemat et al., 2018) and their findings are inconsistent. To address this gap, we explored the gender effect in learning with DEG in our study.

1. **Methodology**

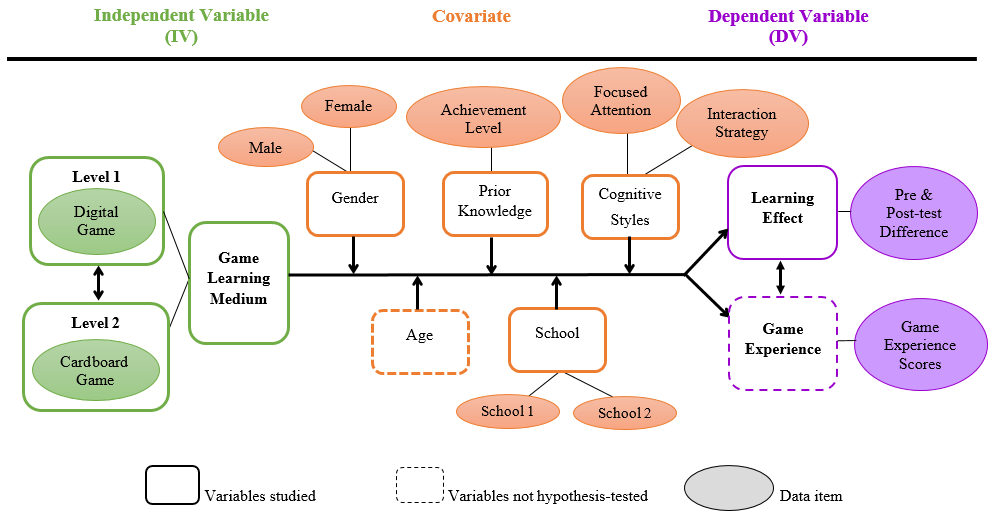
Our empirical work consisted of three evaluation studies: First, the Preliminary Test was conducted with five children at their respective home to test the early DEG prototype. Second, the Pilot Study was carried out with 31 children in their school to evaluate the initial research protocol. Third, the Main Study was conducted with 94 children in two schools (Furió et al., 2015) to verify research hypotheses (Section 4.2); two eye-tracking analysis methods, designated as Method 1 – focused attention inferring from fixation duration - and Method 2 – interaction strategy inferring from gaze sequences, were employed. Details of the Preliminary Test and Pilot Study and Main Study Method 1 are presented in *P* (footnote 1). In this paper we report on Method 2, which is more complex as well as more insightful for understanding the children’s interaction with the DEG. In addition, in this paper we report on the role of gender in the perceptual and behavioural responses to the gameplay; this issue is not explored in *P* either.

* 1. **Ethics Approval and Recruitment**

Participants involved in the empirical studies of this research project were 5-year-old children attending reception (or foundation) years at school. Approval and consent from their parents and head-teachers were obtained. Throughout the recruitment process, invitations, informed consent forms and study flyers were sent via post as well as emails to more than 100 schools. Upon the approval of the head-teacher of the participating school, the first author worked in the school voluntarily a few days before the experimental sessions. This volunteering work was to build rapport with the children so they would not feel uncomfortable or anxious that might arise from facing a stranger during the experiment. Research projects involving human participants are required to go through an ethical approval review. A Disclosure and Barring Service (DBS) check is also needed to ensure that researchers are eligible to work with young children, who are classified as a vulnerable group. Both DBS and ethical approval were obtained prior to working with the children.

* 1. **Research Model**

For the empirical studies, a between-subject experimental design was employed with the *Game Learning Medium* being the independent variable (IV), which consisted of two levels – digital game and cardboard game (Ku et al., 2014) with the same design and content. In the experiment, every participant had to complete a session on an individual basis, with the goal of studying whether the DEG or its cardboard version could lead to better *Learning Effect* (dependent variable- DV) (Hung et al., 2014), which was derived from the difference between pre-test and post-test score for both game learning media. *Game Experience*, another DV, was evaluated by facial expressions of individual children recorded by the eye tracker (NB: the related analysis and results are not reported in this paper). Furthermore, there were several covariates that characterised participants, including *Gender, Prior Knowledge, Cognitive Style, Age* and *School*. However, *Age* was irrelevant since all participants were from the same age group, 5-year-old. *Cognitive Style* (Mason et al., 2013a; She and Chen, 2009) was defined in terms of focused attention and interaction strategy, which were derived from fixation duration and scanpath as eye-tracking data. *Prior Knowledge* (Tsai et al., 2012) was the children’s existing achievement level in numeracy before the experiment and measured with the pre-test; it could not be measured independently because of the lack of formal assessment at such a young age in the UK educational system. Fig. 1 shows the research model depicting the relationships between the variables.



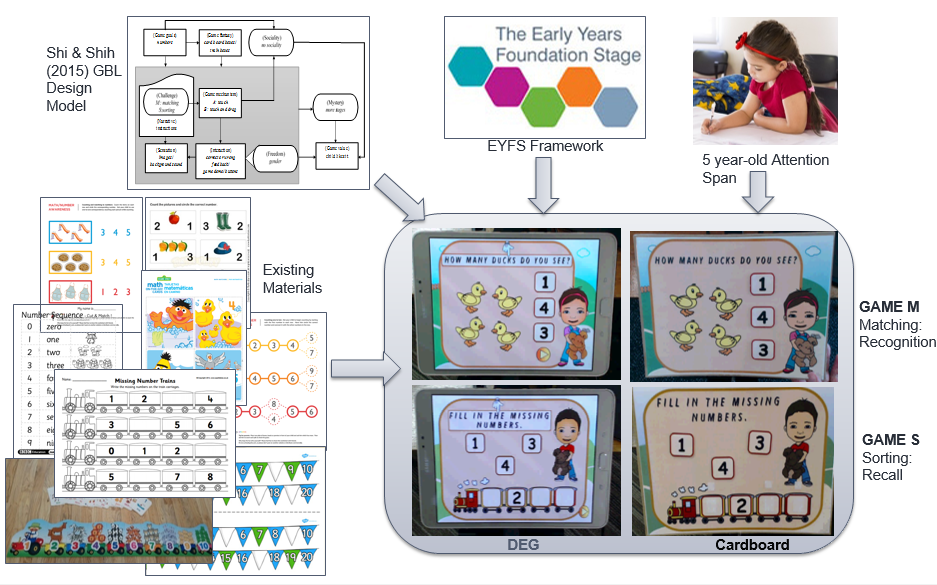
**Fig. 1**. The research model for this study.

* 1. **Game Design and Development**
     1. *Game content*

The content of the DEG was designed by referring to existing learning materials and the UK Early Years Foundation Stage (EYFS) framework on numeracy (Department for Education, 2014). Accordingly, a child between 40 to 60 months and over should be capable of counting numbers from 1 to 20 and sorting them. This helped define the scope of the DEG to support children’s ability to recognise and recall numeric symbols from 1 to 20.

Furthermore, informal observations in a kindergarten were made to gain insight into the DEG design. With the approval of the teachers of the kindergarten, the first author observed how numbers were typically introduced to young children in the classroom and what kind of toys as well as non-digital games children were given to play to help them build numeracy. It was observed that physical objects from everyday applications were used to introduce counting and grouping concepts.

The literature on the game-based learning (e.g. Shi & Shih, 2015), on young children’s attention span (e.g. Ruff & Capozzoli, 2003) and non-complex user interfaces for young children (e.g., Peirce & Centre, 2013; SesameWorkshop, 2012) was also referred to when considering the structure and duration of the game. Fig. 2 illustrates the resources used.



**Fig. 2.** Resources of the game content and design.

The game consisted of two parts: Matching (Game M) and Sorting (Game S) with recognition- and recall-based tasks, respectively. Both the digital and cardboard versions of the game progressed from easy to hard in four levels. This was to build children’s engagement with the game, which had to be short to avoid the loss of interest among young children.

In Game M, the task was to choose the correct answer from the three options presented on the page to the total number of objects (i.e., ducks, fish, bees, apples) given on the left (Fig. 3, left) whereas in Game S, the task was to sort the numbers correctly into the empty train compartments at the bottom of the page (Fig. 3, right). Interactions with the digital and cardboard games are different (Section 4.1). Throughout the game, a static avatar (Westerman et al., 2015) was present in the user interface. We name it as Static Avatar to distinguish it from a dynamic type, and reframe our analysis accordingly. While the avatar could give a sense of presence and thus some motivational effect (Bailey et al., 2009; Segaran et al., 2019), it might not have any direct effect on learning (Section 3.4.4). Two gender options are offered to choose at the beginning of the game because children of this age tend to prefer gender-oriented activities (Markopoulos et al., 2008).



**Fig. 3.** Page layout of level 1 (Page1) for Game M (left) and Game S (right).

* + 1. *Game technology*

The software used to develop the DEG was Adobe Animate CC (aka. Adobe Flash Professional). The ActionScript 3.0 language was used and coded via the action panel or script window in the authoring environment of the Animate CC software. ActionScript and JavaScript programming language share the same root standard making both languages very similar (Animate CC, 2017). The game was published with Adobe AIR, which was set for Android, the OS used for the game.

The game mechanics was simple for the young children: to tap an answer box in Game M and drag-and-drop an answer box in Game S. Apart from appealing images, another key game element – visual feedback for correct and wrong answer (star or sad smiley) with appropriate sound effect – was utilised. To facilitate the gameplay, an animated hand on the instruction bar and a demonstration page for Game S were shown. All instructions were presented in both text and audio format. Evaluation and improvement of the DEG was done through the Preliminary Test and Pilot Study, which were presented in paper *P*, and are not described here.

* + 1. *Cardboard Game Development*

A cardboard version of the DEG was employed to compare the learning effect between the two media (IV). Recycled cardboard boxes, Velcro strips and Blu Tack were used to create the game from scratch. Children could easily pull objects out and stick them back on the cardboard with the help of Blu Tack or Velcro strips similar to some conventional learning materials. The cardboard game was played on the recording area under a camera and above the eye tracker, which was supported by a bespoke frame (Fig. 4). During the experiment, a researcher’s presence was essential. The researcher (the first author) read out instructions and provided verbal as well as non-verbal feedback to every answer pulled-out and stuck back by a child. The researcher was also responsible for switching the cardboard pages throughout the game.

* 1. **Eye Tracker: Setup and Metrics**
     1. *Hardware*

The eye tracker used in this research project was Tobii X2-30 incorporating with a mobile device stand (MDS), a frame supporting the eye tracker, the game device, and an external camera together. The eye tracker is placed underneath the participant, allowing her or him to play in a normal way with the tablet/cardboard (Fig. 4). Technically, the MDS has eight standard configurations for Tobii X2-30 which can be selected based on device type used and participant height (Tobii AB, 2016). We used configuration C2 (i.e. eye-tracker angle -5,2o; eye-tracker distance -8cm; device angle 0o; height difference between device and eye-tracker 5,7cm) since young children are typically short.

|  |  |
| --- | --- |
|  |  |

**Fig. 4.** Setup of the cardboard game evaluation using the coloured star calibration board and the eye tracker is placed below the recording area (circle right). On the left is the setup using the original calibration board, which used small black identification numbers (oval in left).

* + 1. *Calibration*

Calibration is a process where the attributes of a participant's eyes are calculated based on an internal 3D eye model within the eye-tracking software. By using a calibration plate, participants focus on specific points known as calibration points to enable the software to measure eye attributes. In this project, the number of calibration points is kept to be 5, which is necessary for ensuring the tracking accuracy. However, based on the observations of a preliminary study (Mohd Nizam and Law, 2018), the calibration plate (Fig.4) was modified to make it more child-friendly, especially for very young participants. We used coloured stars (i.e., black, green, yellow, blue, red) to replace plain number indicators, which were observed to be not attractive enough for capturing the children’s attention (ibid). As coloured stars are commonly used by teachers as well as in games to reward children for their good responses, children have positive association with them. Hence, using coloured stars can create a playful atmosphere that makes it easier for children to cooperate.

Technically, the gap between the eye-tracker surface and participants' eye ranged from 60 to 65 cm to allow a maximum viewing angle of 31 degrees that covers the recording area. The Tobii Studio software displays this range via the tracking status window. When the children moved out of the range, we used soft tones and simple instructions (e.g., “Uh-oh, can we sit still?”, “Oops, are we wiggling?”) to encourage them to stay within the calibration range.

The calibration process in this experiment started right after the participants had completed their paper-based pre-test and ended exactly after the interactive game-play was over. Before the calibration process began, each child was asked to identify each of the five star colours in a playful way to ensure they could follow the subsequent instructions such as “Mr Tumble has a yellow spotty bag, right? Can you help me look at the yellow star now?”, “George Pig has a green dinosaur, can you help me look at the green star?” (Note: “Mr Tumble” and “George Pig” are the character of a popular children TV programme and cartoon book series in England, respectively). This verification was essential to smoothen the calibration process. Depending on the cooperativeness of the participants and their physical characteristics (i.e. in general children with a smaller body size are more difficult to stay within the tracking range), repeated calibration had to be performed on nine children 2 or 3 times. Some of these children were observed to be slightly irritated (e.g. frowning) the third time they were asked to redo the calibration, but they could be gently reassured.

* + 1. *Segmenting recording scenes*

Segmenting and creating shorter scenes of the eye-tracking recordings enables statistical and visualisation data to be captured. For our study, each participant's recording was divided into short scenes based on the pages of the two parts of the game. In total there were 13 scene pages, including the welcome page, two selection pages, demonstration page, reward page and four task pages of each Game M and Game S. However, due to the large volume of data, only the scene pages of Game M (M1 to M4; matching activities based on recognition) and Game S (S1 to S4; sorting activities based on recall) were used for the analysis in this study.

* + 1. *Area of Interest (AOI)*

An AOI is used to link eye-tracking measures (e.g. fixation duration) to a specific area of a stimulus displayed, facilitating the interpretation of these measures (Hessels et al., 2016). According to Orquin, Ashby, & Clarke (2016), AOI can be defined in two ways: (i) based on expectations where AOI overlaps may occur due to the stimuli design, and high accuracy is not required, for instance, a usability test of a website design (Eraslan et al., 2016); (ii) based on quality criteria where the stimuli design allows maximising the distance between objects, and high accuracy is required, for instance, a research study on visual cognition (Rayner and Reingold, 2015). As overlapping AOIs may happen in our game design, a smaller AOI margin (≈0º margin) was used to balance the proportion of fixations (Orquin et al., 2016).

Six AOIs were selected for each page of Game M and six AOIs for each page of Game S (Fig. 5). In Game M, five user interface (UI) objects were identified as *direct* (i.e. directly related to learning): *instruction A, counting objects B, answer box C, answer box D* and *answer box E* whereas there was only one *indirect* UI object static avatar *F,* which was indirectly related to learning but could enhance game experience, which in turn might lead to higher learning effect. A slightly different structure was used for Game S: *instruction A, train B, answer box D, answer box E* and *answer box F* were identified as direct UI objects whereas *static avatar C* was the only indirect UI object.

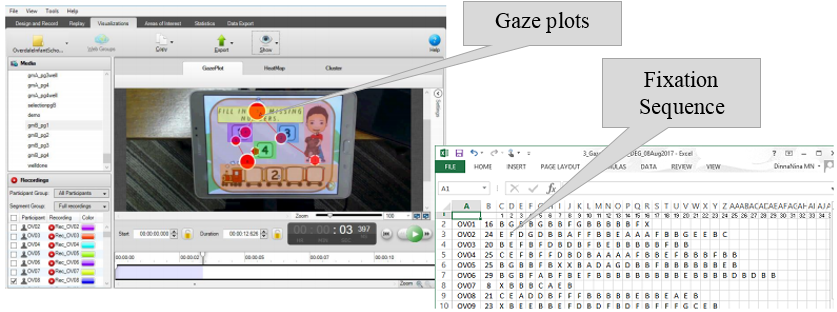


**Fig. 5.** Areas of Interest (AOIs) for Game M (matching, left) and for Game S (sorting, right)

Note that based on our initial check of the data in our preliminary study (Mohd Nizam and Law, 2018), we observed that the children hardly looked at the arrow icon (the navigational element) during the gameplay. It was only looked at when one level of the game had been done; gazing at this element would not contribute to learning. In contrast, quite some children gazed at the background image – static avatar - during the gameplay. To contain the complexity of data analysis, we did not select this navigational element icon as an AOI.

* + 1. *Gaze Sequences (EyePatterns)*

Gaze sequence analysis is a specific type of scanpath analysis that analyse eye-tracking data through visualisation. For our data analysis, gaze plots were observed by identifying all AOI visits made by each participant. Gaze plots are series of fixations and saccades that are presented by circles and lines in the eye-tracking recording (Rösler, 2012). All fixation visits of the gaze plots were extracted manually and converted into fixation sequence in Excel (Fig. 6). The fixation sequences were then fed into *eyePatterns*, a software program that identifies sequence patterns and similarity between fixation sequences (West et al., 2006). However, we only used eyePatterns to collapse fixation sequence into gaze sequence: a short and compact version of the long fixation sequence. This technique was to control human errors in performing the tedious and error-prone collapsing task. For example, eyePatterns can convert this 40-AOI fixation sequence (aka. *extended sequence*): “BABBBBCBBBFBCDBCBBBCBBAAAACBCABBCBBBBBDB” into a 24-AOI gaze sequence (aka. *collapsed sequence*) “BABCBFBCDBCBCBACBCABCBDB. In this collapsing process, consecutive fixations on the same AOI was marked by that AOI (e.g. “BBBB” to “B”; “AA” to “A”), forming a smaller sequence pattern. In particular, a gaze sequence focuses on the order of AOIs visited by the participant through eliminating the consecutive fixations, while fixation sequence focuses on the chronological aspects of the sequence shown by the repeated fixations (West et al., 2006), indicating the approximate duration taken by individual participant (Steichen et al., 2014). A detailed gaze sequence analysis is discussed in Section 4.2.



**Fig. 6.** Extracting gaze plots into fixation sequence in Excel

* 1. **Game Experience**

In our research study, game experience were derived from the children’s facial expressions during gameplay. Several studies have used facial expressions as an approach to study game experience (Tan et al., 2014; Taufik Akbar et al., 2019). Video-recording of a user’s face when interacting with the eye-tracker is a default feature of the Tobii eye-tracking software. Videos so produced can be used for the analysis of facial expressions. To derive the children’s game experience, we coded their facial expressions with the coding scheme developed by LoBue and colleagues (2017). Specifically, the coding scheme consists of three categories (Appendix A) which are Negative (e.g., Sad, Bored), Neutral, and Positive (Happy, Excited), with each being associated with a numerical value of 0, 1 or 2, respectively. Nevertheless, before each child's facial expressions were coded, we segmented the video recording into four sections that covered from the start to the end of the gameplay session (i.e., Introduction, Game *M*, Game *S* and End).

The first author and her colleague, who was not involved in conducting the experiment, coded the children's facial expressions using the coding scheme. While both raters were experienced in interacting with young children, they had not applied the coding scheme beforehand. They also had different disciplinary backgrounds, namely computer science and engineering, respectively. So, both raters practised with existing data and read through related facial expression articles (LoBue et al., 2017; Widen and A Russell, 2010) to get familiarised with the coding task. The initial, Cohen's Kappa between both raters evaluation was 0.49, which was not satisfactory. Therefore, the two raters reviewed the differences until they reached consensus. The improved Cohen's Kappa reached 0.71. Table 1 shows the descriptive statistics, which showed that the children generally had positive experience. Note, however, that we did not collect other data on game experience because there was a time constraint for an experimental session; we discussed this limitation in Section 6. Furthermore, we did not perform any hypothesis testing on the relation between the learning effect and game experience in order to sharpen the paper’s focus on gaze sequence scanpath analysis (Section 3.2).

**Table 1**.

Game Experience scores coded by the raters using the Coding Scheme

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Average Game Experience (N = 94) | | | | |
|  | **Intro** | **Game M** | **Game S** | **End** | **Total Score** |
| Mean | 1.2 | 1.6 | 1.4 | 1.4 | 5.5 |
| S.D. | 0.4 | 0.5 | 0.5 | 0.5 | 1.5 |
| Median | 1.0 | 2.0 | 1.0 | 1.0 | 5.0 |
| Range | 1.0 - 1.0 | 1.0 - 2.0 | 1.0 - 2.0 | 1.0 - 2.0 | 4.0 - 7.0 |

* 1. **Knowledge Test**

A paper-based knowledge test (i.e. pre-/post-test) was designed based on multiple EYFS materials. The complete draft of the test was given to two teachers of the participating schools to validate the difficulty level of the task for 5-year-olds. Both teachers agreed on the difficulty level and gave advice on constructing the instructions for the test. The test was divided into two sets of task - matching and sorting, which were designed to be similar to the DEG and cardboard game activities to evaluate the learning effect (Fig. 7). To assist children to focus on one question at a time, each question was revealed one at a time vertically downwards (as indicated by the dotted lines in Fig. 7) while covering the other questions with pieces of blank paper.

****

**Fig. 7.** Knowledge test with two parts – matching (left) and sorting (right)

1. **Empirical Study**

The Preliminary Test and Pilot Study were conducted to test the research protocol and instruments used to identify improvement suggestions and address them before a larger scale study (Leon et al., 2012). Specifically, we evaluated the eye tracker setup and the early game prototype with the participants to assess their acceptance towards the experimental setup and elicit their requirements for the game (Read et al., 2016; Walsh et al., 2013). Consequently, some key improvements were made, including: modifying the calibration plate; optimising the room lighting condition for the eye tracker; revising the knowledge test structure to avoid confusion; redesigning game elements (Mohd Nizam and Law, 2018).

In this paper, we report primarily the Main Study, especially the use of the gaze sequence analysis method to infer interaction strategy of children used when interacting with the bespoke game implemented in both media – digital and cardboard.

## **Data Collection: Participant and Procedure**

The Main Study was conducted in two primary schools in the UK for seven weeks. The experimental sessions were carried out on an individual basis, and the schedule of the sessions was bound by the respective school’s timetable and activities. 94 signed consent forms were returned by parents of 50 girls and 44 boys, all aged five and were in the foundation (playgroup) stage. Each session took place in an uninterrupted room in the respective school where 59 children played the DEG, and 35 played the cardboard game.

First, a child was asked to complete the pre-test. Next, the child moved to the eye-tracking setup area, playing the DEG or cardboard game on the MDS. The eye-tracking recording started with the calibration process, which would be repeated, if required, to ensure the data quality. The built-in camera of a laptop captured data into the Tobii Studio software to record the child’s gameplay. The researcher (the first author) was present in the room to observe the child’s game interaction behaviour and provided help when necessary. After the child completed the game, the child was asked to complete the post-test.

The average duration of the experimental session for the DEG and that for the cardboard version was 22 and 25 minutes, respectively. The duration included the calibration task (DEG = 5 minutes, Cardboard = 5.5 minutes) and actual gameplay (DEG = 3.7 minutes, Cardboard = 5.4 minutes). The 5-minute gameplay, albeit short, could be effective in enhancing young children’s performance, as shown by some other studies (Wexler et al., 2016). The remaining time was accounted for by other activities – greeting and seating the participant, giving the instructions, filling out the pre and post-test, and debriefing. Overall, the range of the duration of gameplay was within the attention span of 5-year-old (Ruff et al., 1998).

As described in Section 3.2 and depicted in Fig. 1, our research model focused on analysing the learning effect (DV) of both game learning media (IV) and other factors (covariates) were also considered. While the relation between Game Experience and Learning Effect was studied, it is not reported here in order not to prolong the paper.

## **Data Analysis: Gaze Sequence Analysis**

A methodological challenge for our research work was to analyse how gaze sequences (interaction strategy) were related to the Learning Effect, Gender and Achievement Level on both game learning media. All 94 participants’ data from both the DEG and Cardboard game were used in the analyses. In eye-tracking studies, gaze sequences are identified to be associated with the human cognitive processes (West et al., 2006). A number of techniques and software applications have emerged for applying gaze sequences empirically (Eraslan et al., 2016; Steichen et al., 2014; West et al., 2006).

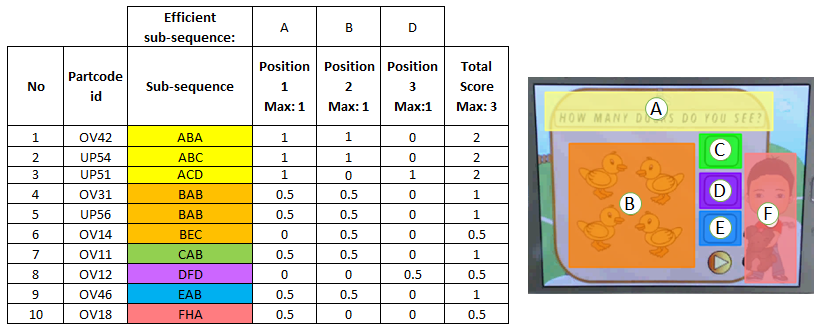
***Gaze Sequence****:* The process of extracting a fixation sequence from gaze plots and then converting the fixation sequence into a gaze sequence was discussed in Section 3.4.4. To study the interaction strategy at the beginning of the game task from these sequences, the first 4 AOIs fixated on each level of both Game *M* and Game *S* were extracted from the gaze sequences to form sub-sequences (Steichen et al., 2014). Fig. 8 shows the process flow of converting gaze plots into sub-sequences. The gaze sub-sequences were used to study the relations of gaze sub-sequence with Learning Effect, Gender and Achievement Level by computing scores for individual gaze sub-sequences. The gaze sub-sequences were also used to study the interaction strategy applied by the child when entering each level of the game task by sorting, colour-coding, clustering and calculating the gaze sub-sequence scores.



**Fig. 8.** Converting Sequence process.

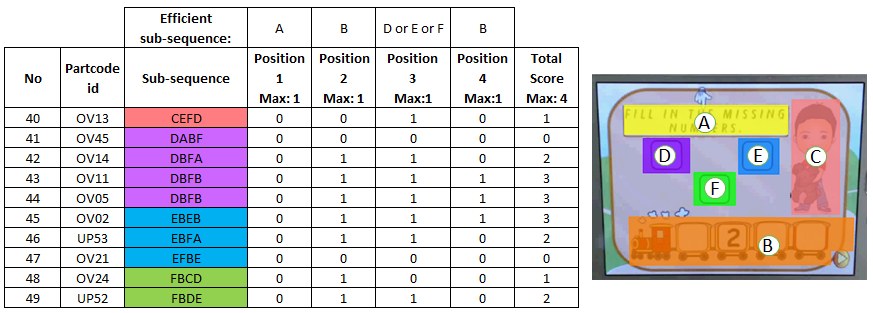
***Scoring Gaze Sub-sequence****:* To use the gaze sub-sequence for analysis and study the children interaction strategy the sub-sequence had to be scored (i.e. gaze sub-sequence score, GSSS). The scoring scheme used in this study was based on the position-based weighted model developed in Sutcliffe & Namoun (2012) and adapted by Eraslan et al's (2016). The model compared the participant’s gaze sub-sequence with the efficient sub-sequence (i.e., the shortest gaze sub-sequence for answering the game task) and applied a scoring scheme based on the AOI position in the sub-sequence.

In Game *M*, the efficient gaze sub-sequence involved only the first three of the four AOIs fixated for a gaze sub-sequence, this was because a single task could efficiently and effectively be completed by visiting three AOIs. Fig. 9 shows how the scores were given to each participant in page 1 of Game *M*. For example, the participant OV42’s gaze sub-sequence was “ABA” and the efficient sub-sequence was “ABD”. Therefore, only 1st and 2nd position of this gaze sub-sequence were given points while 3rd position was not given any point because “A” did not match with the 3rd position of the efficient sub-sequence. Another example is the participant OV31’s gaze sub-sequence for the same task. With the gaze subsequence “BAB”, 1st and 2nd position were given 0.5 point each, although “B” and “A” were not in the efficient sub-sequence position they were still among the first three positions of the efficient sub-sequence. However, the 3rd position of OV31’s sub-sequence was given 0, because no repetition was accepted in calculating the scores.



**Fig. 9.** Scoring the gaze sub-sequence for Page 1 in Game *M* (left). Game *M* design (right).

As for Game S, given its different design the game scores were calculated slightly different but still adapted the position-based weighted model approach. In Game *S*, the efficient gaze sub-sequence involved fixating the first four AOIs to complete a single task (i.e. selecting and dragging one of the number boxes into the train). Fig. 10 shows how the scores were given to participants in page 3 of Game *S*.



**Fig. 10.** Scoring the gaze sub-sequence in Game *S* (left). Game *S* design (right).

For example, OV14’s gaze sub-sequence was “DBFA” and the efficient sub-sequence for Game *S* was “AB(X) B”. The X in the efficient sub-sequence can either be D, E or F in the 3rd position, depending on which number box the participant chose to move it into the train. Unlike Game M where there is one definite scanpath for the correct answer, this first move of Game S has the chance of 1/3 to get the right answer and altogether 9 combinations. To simplify the matter, the efficient sub-sequence does not take the correctness of the move (i.e. there is no restriction which train box should be answered first) into account, focusing on the relevance of the related AOI. Therefore, for OV14 only 2nd and 3rd position were given points while 1st and 4th position were not given any points. Another example is the participant OV02’s gaze sub-sequence, which was “EBEB”, and the efficient sub-sequence was still “AB(X) B”. 1 point was given to 2nd, 3rd and 4th position each, while no point was given for 1st position, giving a total of three. In Game *S* repetitions were accepted but no scores were given for ‘not in the efficient sub-sequence position’ within the first four AOIs fixated, in contrast to the scoring scheme for Game *M*.

With the basic statistical analysis and more specific gaze sequences analysis methods, we answered the three main research questions (**RQs**) of our study (Section 1). Accordingly, a set of associated hypotheses are hypothesized. As there are some closely related and similar terms, their descriptions are reiterated in Table 2 to facilitate the comprehension of the hypotheses and results (Section 5).

**Table 2**

Definition of terms used in this study

|  |  |
| --- | --- |
| **Terms** | **Description** |
| Learning Effect | is referred to the difference between the pre-test and post-test score. |
| Performance Level | are categories of Learning Effect: ‘high’ and ‘low’ performance. |
| Achievement Level | is derived from the pre-test score and categorised as ‘high’ and ‘low’ achievers. |
| Prior Knowledge | is referred to the knowledge the child had before participating in the experiment; Achievement Level is one of the Prior Knowledge components. |
| Interaction Strategy | is referred to the approach to playing the game as derived from gaze sequences; one aspect of Cognitive Style |

**RQ1:** Learning Effect (i.e. difference score = posttest score - pretest score)

* **H1.1:** There is a significant Learning Effect for the DEG.
* **H1.2:** There is a significant Learning Effect for the Cardboard game.
* **H1.3:** There is a significant difference in Learning Effect between the DEG and Cardboard game.

**RQ2**: Gaze Sub-Sequence Score (GSSS)

* **H2.1**: For the DEG, there is a statistically significant correlation between Learning Effect and GSSS for (i) Game M; (ii) Game S.
* **H2.2**: For the Cardboard game, there is a statistically significant correlation between Learning Effect and GSSS for (i) Game M; (ii) Game S.
* **H2.3**: For the DEG, there is a statistically significant difference in GSSS between the two groups with high and low Performance Level for (i) Game M; (ii) Game S.
* **H2.4**: For the Cardboard game, there is a statistically significant difference in GSSS between the two groups with high and low Performance Level for (i) Game M; (ii) Game S.

**RQ3**: Children’s Attributes - Gender and Achievement Level

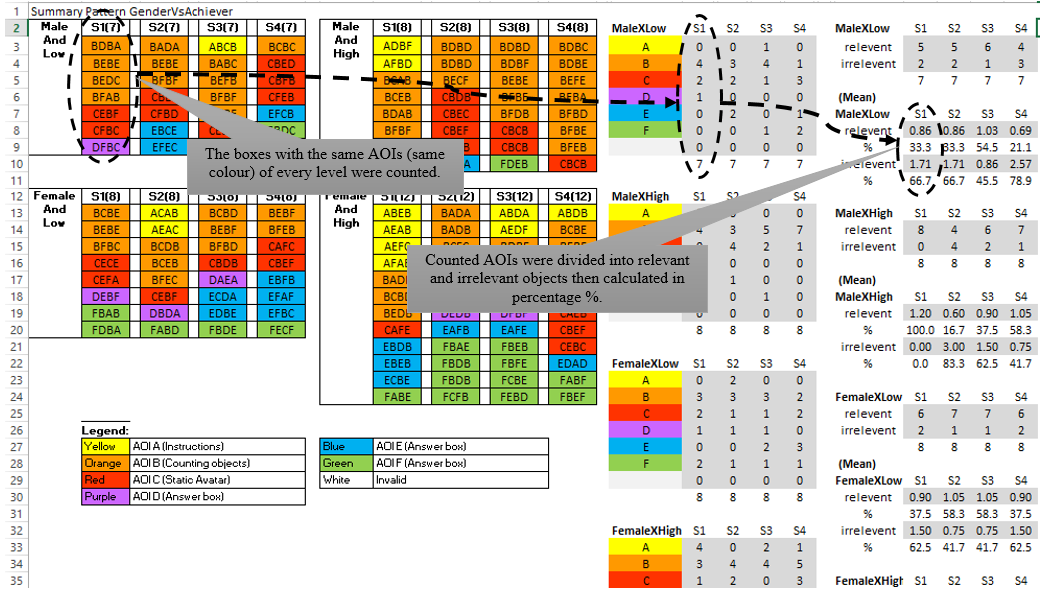
* **H3.1**: For the DEG, there is a statistically significant difference in GSSS between Genders for (i) Game M; (ii) Game S.
* **H3.2**: For the Cardboard game, there is a statistically significant difference in GSSS between Genders for (i) Game M; (ii) Game S.
* **H3.3**: For the DEG, there is a statistically significant difference in GSSS between the two groups with high and low Achievement Level for (i) Game M; (ii) Game S.
* **H3.4**: For the Cardboard game, there is a statistically significant difference in GSSS between the two groups with high and low Achievement Level for (i) Game M; (ii) Game S.

**Clustering the gaze sub-sequences and calculating the gaze sub-sequence percentage:** In studying the interaction strategy applied by a child when entering each level of the game, the gaze sub-sequence had to be sorted alphabetically and colour-coded, based on which AOI the child first fixated on a spreadsheet (see the legend in Fig. 11). Boxes coded in red represent *Indirect AOI* objects while other colours such as yellow, orange, purple, blue and green are *Direct AOI* objects (Section 3.4.3). The numbers of Direct and Indirect AOIs are 5 and 1, resulting in the corresponding fixation probability of 5/6 and 1/6. In other words, it is 5 times more likely for a child to first fixate a direct AOI than an indirect one. For example, Fig. 11, the Male-Low cluster (NB: High vs. Low cluster is differentiated by Performance Level) playing the cardboard game, 5 out of the 7 children fixated first on Direct AOIs and 2 Indirect ones; the ratios were 5/7 and 2/7. However, to control the bias, it is necessary to divide the two ratios by the respective fixation probability as follows:

*Direct AOIs: 5/7 divided by 5/6=0.86= 0.86/ (0.86+1.71) =33.3%*

*Indirect AOIs: 2/7 divided by 1/6=1.71=1.71/ (0.86+1.71) =66.7%*

Similarly, the same approach was applied to the Male-High, Female-Low and Female-High clusters and the results are discussed in the gaze sub-sequence percentage (%) (Section 5.6). The weighted averages are converted into percentages, allowing easier comparisons across the clusters.

****

**Fig. 11.** Calculating the gaze sub-sequence percentage from Game *S* of the Cardboard Game.

1. **Results**

In this section we first present general preliminary data analysis approaches to check for outliers and normality, and such approaches were applied to all data. Then we report on the findings per research question and its associated set of hypotheses (Section 4.2).

* 1. **Preliminary Analysis Approaches**

A preliminary analysis was performed to make sure no missing data, outliers were dealt with to minimise bias to the dataset and to identify which statistical methods, parametric or non-parametric, to apply based on the data distribution of the variables. The paper-based pre-test and post-test evaluation, which was actually the same test administered before and after the gameplay, had a maximum score of 5 for Game *M* and 7 for Game *S*.

*Outlier*: By using the boxplots (Appendix B), the variables (pre-test, post-test, prepost-test differences and in-game variables) were examined and observed for outliers (Tabachnick and Fidell, 2007; Warner, 2008). 7 cases with outliers (values more than three IQRs (Hoaglin and Iglewicz, 1987) from the end of the boxplot) were transformed to the next highest/lowest non-outlier value, maintaining the total number of cases for both the DEG: 59 and Cardboard game: 35.

*Normality*: A Shapiro-Wilk test was used on the variables to study the data distribution and thus identify which statically method to apply.

* 1. **Learning Effect**

Non-parametric Wilcoxon signed-ranks tests were used to study Learning Effect of both the DEG and Cardboard game. Statistical results for the DEG (Mdn*pre*=7; Mdn*post*=8) indicated that no statistically significant improvement in learning was achieved (Z=1.10, *p*=0.27), thus **rejecting H1.1**. In contrast, for the Cardboard game (Mdn*pre*=9; Mdn*post*=10), there was a statistically significant improvement in learning (Z=1.90, *p*<.05), thus **accepting H1.2**. These results could be attributed to the typical learning environment of the classroom where interacting with physical objects is the prevailing educational method. Young children are taught through active learning which involves sensing and manipulating physical materials (Essa, 2012). Another possible explanation for the results is that children participating in this study might not be exposed to DEGs or have restricted access to them.

A Mann Whitney test was used to evaluate the difference in Learning Effect between the DEG and Cardboard game. No statistically significant difference in Learning Effect was found (U =948.5, Z=0.67, p=.5) between the two game learning media (Mdn*media* =0). **H1.3 was rejected**.

* 1. **Gaze Sub-Sequence Score (GSSS) on Learning Effect**

The Learning Effect threshold for classifying the high and low performance was the average *difference score* for each game in both media: DEG: Game *M* (Mean = 0.17, SD = 0.93); Game *S* (Mean = 0.24, SD = 1.006); Cardboard: Game *M* (DEG Mean = 0.09, SD = 0.74); Game *S* (Mean = 0.37, SD = 1.003). Those who were equal or above the threshold were classified as high performers, otherwise as low performers.

A Spearman’s rank test was performed to assess the relation between Learning Effect and GSSS for both Game *M* and Game *S* on the DEG and Cardboard game. For the DEG, neither Game *M* (N = 59, r=0.03, p=.82) nor Game *S* (N = 59, r=0.20, p=.14) had any significant correlation between the two variables. A similar result was observed for the Cardboard game; no statistically significant correlation was obtained between Learning Effect and GSSS for either Game *M* (N = 35, r=0.10, p=.57) or Game *S* (N = 35, r=0.20, p=.24). **H2.1and H2.2were rejected**.

Mann Whitney tests were used to evaluate the difference in GSSS between the high and low performers on Game *M* and Game *S* of both game learning media. For the DEG, no statistically significant difference in GSSS was found for Game *M* (U =336, Z=0.35, p=.73) and Game *S* (U =282, Z=1.43, p=.15) between high and low performance (Table 3). A similar result was seen for the Cardboard game, where no statistically significant difference in GSSS was found in Game *M* (U =64, Z=1.42, p=.16) and Game *S* (U =117, Z=0.74, p=.46) between high and low (performance (Table 3).

The statistically non-significant difference between the high and low performers was probably due to the game design, which might not be a good differentiator in studying Learning Effect of the game. The game design was developed based on young children’s attention span and was tested in the Pilot Study. There might also be the ceiling effect of the pre-/post-test, although it was pilot tested and improved. **H2.3 and H2.4 were rejected**.

**Table 3**

Median GSSS of high and low performers for the DEG and Cardboard game

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | DEG | | Cardboard Game | |
| High Performers | Low Performers | High Performers | Low Performers |
| Game *M* | 2.0 | 1.0 | 2.0 | 1.0 |
| Game *S* | 2.0 | 1.0 | 2.0 | 1.0 |

* 1. **Gaze Sub-Sequence Score on Gender and Achievement Level**

Mann Whitney tests were used to evaluate the difference in GSSS between Genders on Game *M* and Game *S* of both game learning media. For the DEG, no statistically significant difference in GSSS was found for Game *M* (U =432, Z=0.05, p=.96) or Game *S* (U =332, Z=1.58, p=.12) between boys and girls (Table 4). A similar result was seen for the Cardboard game, where no statistically significant difference in GSSS was found for Game *M* (U =112, Z=1.28, p=.2) or Game *S* (U =134, Z=0.54, p=.59) between both genders (Table 4). These statistically non-significant results showed that girls and boys at this age would not have any inherent differences in applying interaction strategies when interacting with games on numeracy. **H3.1 and H3.2 were rejected**.

**Table 4**

Median GSSS of girls and boys for the DEG and Cardboard game

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | DEG | | Cardboard Game | |
| Girls | Boys | Girls | Boys |
| Game *M* | 4.0 (3.5 - 5.0) | 4.0 (3.5 - 4.7) | 4.0 (2.6 - 4.5) | 4.5 (3.5 – 5.0) |
| Game *S* | 5.0 (2.8 - 6.3) | 4.0 (2.0 - 4.5) | 4.0 (3.0 – 6.8) | 4.0 (3.0 - 5.0) |

As described in Table 2, the Achievement Level was defined by the pre-test score. The threshold for classifying the high and low achievers was the mean pre-test for each game in both media: DEG: Game *M* (Mean = 3.51, SD = 1.32); Game *S* (Mean = 3.68, SD = 2.71); Cardboard: Game *M* (Mean = 3.94, SD = 1.26); Game *S* (Mean = 4.03, SD = 2.75). Those who were equal or above the threshold were classified as high achievers, otherwise as low achievers.

Mann Whitney tests were used to evaluate the difference in GSSS between high and low achievers on Game *M* and Game *S* of both game learning media. For the DEG, there was a statistically significant difference in GSSS for both Game *M* (U =288, Z=2.27, p=.02) and Game *S* (U =261, Z=2.67, p=.01) between high and low achievers (Table 5). However, for the Cardboard game, no statistically significant difference in GSSS was found for Game *M* (U =137, Z=0.44, p=.66) or Game *S* (U =106, Z=1.48, p=.14) between high and low achievers (Table 5). The statistically significant difference result in the DEG showed that the children’s interaction strategy could be influenced by their prior knowledge (Achievement Level) when playing the game digitally. However, no statistically significant difference was found for the Cardboard game. A possible reason for this result may be due to the nature of the physical game where interaction with the game was more sensory-based (concrete interaction). In the DEG, children were less distracted since they did not interact with any concrete objects, allowing the children to think abstractly and develop interaction strategies. While for the Cardboard game, the children’s attention tended to be on the concrete objects they were interacting with, compromising their cognitive resources for interaction strategies. **H3.3** **was accepted** **and** **H3.4** **was rejected**.

**Table 5**

Median GSSS of high and low achievers for the DEG and Cardboard game

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | DEG | | Cardboard Game | |
| High Achievers | Low Achievers | High Achievers | Low Achievers |
| Game *M* | 4.5 (3.8 – 5.0) | 3.8 (3.5 – 4.5) | 4.3 (3.5 – 5.0) | 4.0 (2.5 – 4.5) |
| Game *S* | 5.0 (4.0 – 5.5) | 3.0 (1.0 – 5.0) | 4.5 (3.3 – 7.0) | 4.0 (3.0 – 5.0) |

* 1. **Game Sub-sequence Matrices Results**

In Game *M* of the DEG (Fig. 12), the game sub-sequence matrices showed that for three out of 15 girls with high achievement their first-fixated AOI of the first page of Game *M* (*M1)* was the “instruction” *A* (yellow), nine was the “counting objects” *B* (orange) and the other three was “answer boxes” *D* (purple) and *E* (blue). The matrices also showed that across Game *M*, the attention of several low achievement children, both male and female, was diverted by the static avatar *F* (red) at the beginning of the game pages. Note that white boxes in the matrices (Fig. 12) indicate invalid data and they are not included in the gaze sub-sequence percentage calculation.



**Fig. 12.** Gaze sub-sequence matrices between genders and achievement levels for the four pages of Game *M* (*M1, M2, M3 and M4*) of the DEG.

As for Game *S* of the DEG (Fig. 13), for boys with low achievement,red boxes can be seen across the levels from *S1* to *S4*, indicating the number of children whose attention was diverted by the static avatar *C* (red) in the beginning of each page in Game *S.* The matrices also showed that boys’ attention was prone to be diverted by the static avatar *C* compared to girls’. Although a majority of children fixated the “answer box” *B* (orange) in the beginning, a number of children gazed at the centre of the screen and fixated the “answer box” *F* (green). This may be because some children tended to stare in the middle of the tablet screen while waiting for the next page to appear.



**Fig. 13.** Game sub-sequence matrices between genders and achievement levels for the four pages of Game *S (S1, S2, S3 and S4)* of the DEG.

As for Game *M* of the Cardboard game (Fig. 14), the gaze sub-sequence matrices showed that girls with high achievement were likely to fixate the “answer box” *C* (green) across the levels from *M1* to *M4*, indicating that a number of children from this group scanned the number boxes from the top and then navigated to the other areas. This group also tended to look at the “instruction” *A* (yellow) first. Overall, from the matrices of Game *M*, the attention of all four groups of children was diverted by static avatar *F* (red) but boys with low achievement tended to be so distracted more frequently.



**Fig. 14.** Gaze sub-sequence matrices between genders and achievement levels for Game *M* of the Cardboard game.

For Game *S* of the Cardboard game (Fig. 15), similar to Game *M* of the Cardboard game*,* the gaze sub-sequence matrices showed that girls with high achievement tended to look at “instruction” *A* (yellow) first in the beginning of each page in Game *S*. The girls, irrespective of the achievement level, consistently looked at the “answer box” *F* (green) from *S1* to *S4*, indicating many from this group scanned from the middle to the other areas of the game. Overall, the attention of all four groups of Game *S* was diverted by the static avatar *C* (red), but the attention of the low achievement children of both genders tended to be so diverted more frequently. In summary, the attention of the children using the Cardboard game was more diverted by the static avatar as compared to those using the DEG game. This may be because children tended to think the static avatar on the Cardboard game had a function to play while the DEG was only a decorative character.



**Fig. 15.** Gaze sub-sequence matrices between genders and achievement levels for Game *S* of the Cardboard game.

* 1. **Sub-sequence Percentages on UI objects Results**

For Game *M* of the DEG (Table 6), both boys and girls with high achievement on average gazed 100% on Direct UI objects in the beginning of the game pages as compared to 87% and 80% for the children with low achievement. This showed that higher achieving children’s attention was not diverted by the static avatar when entering each page.

**Table 6**

Sub-sequence percentages on Direct UI objects between genders and achievement level on both Games *M* and *S* of the DEG.

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Groups (DEG) | Game *M* (%) | | | | Mean  *M* (%) | Game *S* (%) | | | | Mean  *S* (%) |
|  | *M1* | *M2* | *M3* | *M4* |  | *S1* | *S2* | *S3* | *S4* |  |
| Boy x Low | 100 | 74 | 100 | 74 | 87 | 44 | 74 | 74 | 44 | 59 |
| Boy x High | 100 | 100 | 100 | 100 | 100 | 55 | 100 | 55 | 100 | 78 |
| Girl x Low | 74 | 72 | 100 | 72 | 80 | 74 | 100 | 100 | 100 | 94 |
| Girl x High | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 74 | 100 | 94 |

However, in Game *S* of the DEG a different pattern can be observed between the genders. Girls, both high and low achievers on average gazed 94% on a Direct UI object in the beginning of the game pages compared to boys with 59% and 78% for the high and low achievement group, respectively. The result implied that in Game *S* of the DEG, boys’ attention was easily diverted by the static avatar and the attention of boys with low achievement was constantly distracted across the four pages of Game *S*. Overall, for the DEG, girls with high achievement had the highest average percentage of gazing at Direct UI objects when entering each page with 100% in Game *M* and 94% for Game *S.*

As for Game *M* of the Cardboard game (Table 7), on average girls, low and high, had a higher percentage in gazing at Direct UI objects, 61% and 47%, compared to boys with 38% and 28%, respectively. None of the groups maintained a higher percentage gazing at Direct UI object across the four pages of Game *M*. The attention of all groups of children playing Game *M* of the Cardboard game was most distracted with the static avatar on page *M3*. However, an obvious drastic drop in percentage on direct objects from page *M2* to *M3* can be observed for girls with low achievement. This may be because of the difficulty in progressing from the easy to hard level. The hard level task began on page *M3* for numbers 11-15 (*M4* was the hardest level for numbers 16-20), which might prompt the children to wander around the Cardboard game, and girls with low achievement tended to wander more than the other groups.

**Table 7**

Sub-sequence percentage on Direct UI objects between genders and achievement level on both Games M and S of the Cardboard game.

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Groups (Cardboard) | Game *M* (%) | | | | Mean  *M* (%) | Game *S* (%) | | | | Mean  *S* (%) |
|  | *M1* | *M2* | *M3* | *M4* |  | *S1* | *S2* | *S3* | *S4* |  |
| Boy x Low | 13 | 33 | 13 | 33 | 23 | 33 | 33 | 55 | 21 | 36 |
| Boy x High | 21 | 55 | 17 | 58 | 38 | 100 | 17 | 38 | 58 | 53 |
| Girl x Low | 38 | 100 | 11 | 38 | 47 | 38 | 58 | 58 | 38 | 48 |
| Girl x High | 69 | 69 | 38 | 67 | 61 | 69 | 50 | 100 | 38 | 64 |

As for Game *S* of the Cardboard game (Table 7), children with high achievement of both genders had a higher percentage in gazing at Direct UI objects, 64% (girls) and 53% (boys), compared to low achievement children with 48% (girls) and 36% (boys). None of the four groups maintained a high percentage. A drastic fall of percentage on the direct object from page *S1* to *S2* can be observed for boys with high achievement. A possible explanation for this was that this group might build curiosity upon getting familiar in page *S1*. Once gaining confidence from the previous page, this group tended to look elsewhere to explore beyond what they had seen. Overall, similar to the DEG, girls with high achievement had the highest average percentage of gazing at the direct UI objects with 61% for Game *M* and 64% for Game *S.*

Based on the results, no consistent gaze sequences can be identified among the boys or girls. However, for both DEG and Cardboard game, girls with high achievement showed a higher percentage of gazing at direct UI objects as compared to their counterparts for both Game *M* and *S*. The relatively low percentage in Game *M* and Game *S* of the Cardboard game may be because of the nature of the game, where physical interactions with the game (attaching and detaching objects) were more distributed on the cardboard.

1. **Discussion**

In conducting this study, we have dealt with different challenges. Among others, we highlight those pertaining to applying eye-tracking technology with young children. Next, we present implications for designing DEGs for young children. Then, we reflect on the limitations of our work.

* 1. **Challenges for Eye-tracking Research**

Despite the recent advances in eye-tracking technology, researchers still face different challenges when applying it, especially with young children. We discuss such challenges from the practical and conceptual perspectives in the following.

* + 1. **Practical Challenges**

Eye-tracking technology is known to support studies related to focused attention and cognitive processing (Alemdag and Cagiltay, 2018; Eraslan et al., 2016; Lai et al., 2013) among adult participants. However, applying eye-tracking methodology with young children can be very challenging and time-consuming. Preparations such as lighting in the experiment room had to be known before setting up the experiment, ensuring that the reflection does not interfere with the eye-tracking recording, and the distance the child sits within the eye tracker’s ideal range also needs to be identified before the experiment. The most challenging part with young children was the calibration process (Section 3.4.2). Depending on the cooperation of the child and eye feature, multiple calibration rounds need to be done on a single child. The extra time required for this process needs to be taken into account for an eye-tracking study.

As scanpath analysis involves the process of positioning AOIs, extracting plots, clustering and comparing (matrices), it is very time–consuming, tedious, and error-prone, especially one needs to manually extract scanpath from the AOIs visited (Section 3.4.5). However, unlike heatmaps illustrating how much UI objects being fixated, scanpath data can reveal in which order a participant views those objects, making the analysis more informative. Given the goal of gaining insights into how children interact with both game learning media to derive their learning strategies, the efforts are worthwhile.

* + 1. **Conceptual Challenges**

One contribution of our work is a novel way of classifying gaze scanpaths and quantifying them into a measure known as Gaze Sub-Sequence Scores (GSSS) (Section 5.3). This measure is operationalised as an indicator of efficiency - the shortest possible path to attain a goal. However, alternative paths can lead to the same goal as well. A caveat is that GSSS is not a success indicator. In fact, our empirical findings showed no significant relationships between GSSS and the learning effect in either of the two game versions (Section 5.4). In other words, those children who gained more from interacting with the game did not necessarily take more efficient paths. This raises the question what implications we can draw from derivatives of gaze data such as GSSS. One clear premise is that we should not expect all learners, irrespective of their background and preference, follow a ‘normative’ path to solve a problem. Take it further, it could even become an ethical concern if learners of different abilities were trained to follow the same path. This would undermine the pedagogical principle of accommodating individual differences. Nonetheless, it is intriguing to find out what key factor accounts for different GSSS and the potential use of this measure. Our empirical findings hint at some insight. Specifically, children with a higher level of prior knowledge of numeracy, as measured by the pre-test, had a significantly higher GSSS than their counterparts with a lower level of prior knowledge (Section 5.4). This suggests that GSSS can be a measure of prior understanding. One may speculate that it can also be a measure of confidence (or confusion). However, this entails further research to establish this hypothesized relation between GSSS and level of confidence.

* 1. **Implications for Game Design**

*Game Duration*. The game design was developed based on the literature of young children’s attention span (e.g., Ruff et al., 1998; Ruff and Capozzoli, 2003). However, from the results and observations of our above described work, the short attention span of young children derived from previous studies, which were conducted in non-technological contexts, appears to be unsuitable. Specifically, the focused attention span for free play with a mix of construction and symbolic toys was reported to be 104 seconds (~1.7 mins) for 50-months-old (Ruff et al., 1998) and 181 (~3 mins) seconds for 54-months-old (Ruff and Lawson, 1990). More recently, Hallze and Droit-Volet (2017) conducted a study to investigate the relation between time perception and attention capacity of 5 to 7-years-old. They applied a two-decade-old instrument, namely the "Test of Everyday Attention for Children" (TEA-Ch) (Manly et al., 1999). While they confirmed significant differences in time perception among the three age groups, they did not provide any age-specific attention spans. Overall, there is a strong need for more research work on identifying young children's attention span, which is different from that identified decades ago, given the increasing exposure to digital technologies. This information is critical for informing the design of learning activities, including game-based ones, for young children.

Since interactive technologies, such as educational games, can engage users, both young and old, for a longer period of attention span. Furthermore, the attention span of young children when using DEGs may be different due to the engagement factor of a game. Therefore, the duration of a DEG should not be too short as suggested by literature on children’s attention span. However, it should not be too long either because of the potential fatigue effect.

*Game Difficulty*. A DEG for 5-year-old children game content and difficulty level (Section 3.3.1) should not be limited by the EYFS framework (number 1 to 20). The difficulty level can go beyond the number 20 to avoid the ceiling effect of these young children which have different learning paces. The number of levels in the game should also be added to enable the learning effect to be differentiated and to support abler children.

*Instruction*. Typically, when a person starts to plays a new game, she reads the instruction before doing anything else. However, the instruction text did not play a role in the game due to the children’s low reading ability (Section 5.5). Instead, they listened to the audio instruction. Game designers can drop the instruction text that occupies quite some real estate of the game, replacing it with an object that is more engaging for young children.

*Non-gender specificity*. The non-significant statistical results on gender difference can mitigate the stereotypical view (Arroyo et al., 2013; Sullivan and Bers, 2016) that one gender is better than the other in terms of learning via DEG (Section 2.3). The results showed that young children did not have any inherent differences in applying interaction strategy to learn arithmetic via games. It can be inferred that 5-year-olds, irrespective of gender, may have the comparable capacity for learning through game. Boys and girls of this age can be trained in the same way without any bias.

*Design of static avatar*. The function of the static avatar in our numeracy games was decorative, making the game look aesthetically pleasing and thus more motivating to play with. The avatar could also have the role of eliciting positive gameplay experience, given that the aesthetic quality is identified as a critical factor for user experience (Turner, 2017). In general, a happy learner can be a more effective problem-solver (Isen et al., 1987). Some of the children tended to look at the avatar in every page of the game (Section 5.3), despite its remaining static. This observation suggests that these children could have some emotional responses to the avatar. While we cannot determine the exact nature of those responses and the reasons why the children paid attention to the avatar, it could be interpreted with the lens of emotional attachment, albeit in its crude form (Schöbel et al., 2019). Nonetheless, designing an avatar for a DEG needs to strive a delicate balance between eliciting positive emotions and diluting attention to learning materials. The avatar of our study is 2D static with gender-based version available; it is also non-interactive, large (cf. the screen estate), visual mode only, and non-configurable. Without related data, we cannot evaluate their respective impact on user experience. Hence, we propose studying factors such as dynamicity, interactivity, size, colour, location, realism, and multimodality of an avatar in order to inform its design that fits the goal of a DEG.

* 1. **Limitations**

*Only a particular target group*. For a practicable comparison on young children between both game learning media and game tasks (Game *M* and Game *S*), a specific target group had to be selected on a selected topic area of the EYFS curriculum. There were seven areas in the UK early years’ education system but for this study arithmetic was selected due to insufficient empirical work on DEG on numeracy. The findings of this study may only be generalizable for DEGs on numeracy for 5-year-old children in the UK.

*Lack of Prior Knowledge data*. At such a young age no formal assessment was available and applied in the playgroup classroom. Milestone observation is the most applicable method used by teachers at the playgroup level, which is inevitably rather subjective. Nonetheless, for the concern of privacy, the related data were not shared by the teachers. Hence, the Achievement Level derived from the pre-test result was used as the measurement for children’s Prior Knowledge in the study.

*Device for data collection*. The use of the screen mounted eye-tracking device was to minimise the intrusiveness of the device to be used by young children such as wearable eye-tracker, which can cause discomfort or distress. Nevertheless, parents or guardians of young children may not give their consent for their child to participate in experiments involving intrusive devices.

*Statistically non-significant Learning Effect*. The ceiling effect of the test material makes it improbable to demonstrate any learning gain. This could probably be due to the fact that some children involved in the study may have different tutorial support and/or better home education, which allows them to know numeracy better than average. The extraneous influences of the social factors are beyond control of the study. Ideally, if a much bigger sample of children from foundation years was recruited, the impact of the confounding variables would be mitigated.

*Study Duration*. The study spread over several weeks due to the fact that there was only one eye-tracker, which was relatively expensive, for the experimental sessions, which were thus carried out on an individual basis. The schedule for the experiment sessions was constrained by the school term calendar and classroom activities. Inevitably our data collection work needed to revolve around these school-related constraints. Nonetheless, we mitigated the impact of spreading the study over weeks by the pre-test. In other words, any change in pre-knowledge in the study’s duration could be assessed and controlled statistically. For future work, strategies such as budgeting for more than one eye-tracker and the involvement of additional researchers can help optimise the study duration.

*Gaze Sub-Sequence Score.*Our proposed GSSS approach, while novel and feasible, could have further been substantiated with a parallel analysis of touch sequences. The integration of the two types of data sequences – gaze and touch – can be useful for assessing decision-making strategies (Jaswal et al., 2020; Miles et al., 2017; Vickers, 2011). Specifically, gaze sequences can inform us how a child takes in the objects of a game user interface, thereby enabling us to infer patterns in information processing. Motor choices can provide us behavioural data how the child actually interacts with the game, lending further evidence to the inference drawn from the gaze data. Nonetheless, we have not collected touch data in our empirical study. The lack of such gaze-motor is a limitation of our study, which we aspire to explore in our future studies.

*Game Experience*. The aim of our study was to study the learning effect and explore the interaction strategies using the GSSS approach (Section 3.2). The technique we used to evaluate the game experience was inspired by other studies where facial expressions were emotional data (LoBue et al., 2017; Tan et al., 2014; Taufik Akbar et al., 2019). While this proved viable to derive some basic and useful information, complementary information could have been collected with other approaches, such as semi-structured interviews with the children. However, this was not possible in our study because the experimental session with individual child was bounded to the school tight timetable and activities (Section 4.1).

1. **Conclusion**

Our research project contributed to the applied body of knowledge in the field of Human-Computer Interaction (HCI) and Digital Game-Based Learning (DGBL). For the field of HCI, it helped bridge the gap of the hitherto limited work on applying eye-tracking technology with young children. Specifically, we provided empirical evidence for the strengths and limitations of the existing eye-tracking methodology for 5-year-olds and identified improvement suggestions such as rendering the calibration board more child-friendly. For analysing eye-tracking data, we refined and augmented the Gaze-Subsequence marking scheme to gain deeper insights into children’s interaction strategies when interacting with the educational game. For DGBL, a better understanding of young children’s interaction strategy can inform the design of quality DEGs to support learning. Furthermore, the empirical findings of the gender effect can reduce or even eliminate stereotypical gender-biases for acquiring numeracy.

For our future work, we aim to conduct larger-scale studies with participants of different demographic backgrounds, enhancing the generalizability of our findings, and also with DEGs of different levels of complexity to support children with mixed abilities. When wearable eye-trackers become more affordable and child-friendly, we will utilise them to maximise the quality of eye-tracking data.

Furthermore, the main scope of the current study was to study the learning effect, we will attempt to evaluate game experience with DEGs through eye-tracking data. Some existing research use the pupil size derived from eye-tracking data to recognize a person’s emotion (Oliva and Anikin, 2018). However, it is an emerging research and no consistent findings or directions are concluded, implying more research is called forth. Furthermore, other scanpath measures such as ‘fixation sequence’ or longer sub-sequences (Eraslan et al., 2016) can be applied to study gaze sequences.

The main goal of this study to evaluate the learning effect of DEGs and the applicability of eye-tracking methodology for 5-year-olds has successfully been achieved. The empirical findings have laid some valuable groundwork for future research and practical work along this line of inquiry. This is particularly important as we are witnessing the increasing use of DEGs among young children, it is essential to understand how they can learn through DEGs and how DEGs can better be designed to support their young users.

# **References**

Adams, E., 2014. Fundamentals of Game Design, 3rd ed. New Riders Publishing, Thousand Oaks, CA, USA.

Al-Wabil, A., Alabdulqader, E., Al-Abdulkarim, L., Al-Twairesh, N., 2010. Measuring the user experience of digital books with children: An eyetracking study of interaction with digital libraries, in: 2010 International Conference for Internet Technology and Secured Transactions. pp. 1–7.

Aladé, F., Lauricella, A.R., Beaudoin-Ryan, L., Wartella, E., 2016. Measuring with Murray: Touchscreen technology and preschoolers’ STEM learning. Comput. Human Behav. 62, 433–441. https://doi.org/http://dx.doi.org/10.1016/j.chb.2016.03.080

Alemdag, E., Cagiltay, K., 2018. A systematic review of eye tracking research on multimedia learning. Comput. Educ. 125, 413–428. https://doi.org/10.1016/j.compedu.2018.06.023

Animate CC, 2017. ActionScript with Animate CC [WWW Document]. URL https://www.adobe.com/uk (accessed 2.5.18).

Arroyo, I., Burleson, W., Tai, M., Muldner, K., Woolf, B.P., 2013. Gender differences in the use and benefit of advanced learning technologies for mathematics. J. Educ. Psychol. 105, 957–969.

Arroyo, I., Royer, J.M., Woolf, B.P., 2011. Using an Intelligent Tutor and Math Fluency Training to Improve Math Performance. Int. J. Artif. Intell. Ed. 21, 135–152. https://doi.org/10.3233/JAI-2011-020

Bailey, R., Wise, K., Bolls, P., 2009. How avatar customizability affects children’s arousal and subjective presence during junk food-sponsored online video games. Cyberpsychology Behav. impact Internet, Multimed. virtual Real. Behav. Soc. 12, 277–283. https://doi.org/10.1089/cpb.2008.0292

Bedwell, W.L., Pavlas, D., Heyne, K., Lazzara, E.H., Salas, E., 2012. Toward a Taxonomy Linking Game Attributes to Learning: An Empirical Study. Simul. Gaming 43, 729–760. https://doi.org/10.1177/1046878112439444

Birkett, S., Galpin, A., Cassidy, S., Marrow, L., Norgate, S., 2011. How Revealing Are Eye-Movements for Understanding Web Engagement in Young Children, in: CHI ’11 Extended Abstracts on Human Factors in Computing Systems, CHI EA ’11. Association for Computing Machinery, New York, NY, USA, pp. 2251–2256. https://doi.org/10.1145/1979742.1979900

Bulling, A., Duchowski, A.T., Majaranta, P., 2011. PETMEI 2011: the 1st international workshop on pervasive eye tracking and mobile eye-based interaction. Proc. 13th Int. Conf. Ubiquitous Comput. https://doi.org/10.1145/2030112.2030248

Bulling, A., Gellersen, H., 2010. Toward Mobile Eye-Based Human-Computer Interaction. Pervasive Comput. IEEE 9, 8–12. https://doi.org/10.1109/MPRV.2010.86

Bulling, A., Roggen, D., 2011. Recognition of visual memory recall processes using eye movement analysis. Proc. 13th Int. Conf. Ubiquitous Comput. https://doi.org/10.1145/2030112.2030172

Burnett, C., 2010. Technology and literacy in early childhood educational settings: A review of research. J. Early Child. Lit. 10, 247–270. https://doi.org/10.1177/1468798410372154

Catsambis, S., 2004. The Gender Gap in Mathematics: Merely a Step Function ?, in: Gallagher, A.M., Kaufman, J.C. (Eds.), Gender Differences in Mathematics : An Integrative Psychological Approach. pp. 220–245.

Choi, Y.S., Mosley, A.D., Stark, L.W., 1995. String editing analysis of human visual search. Optom. Vis. Sci. Off. Publ. Am. Acad. Optom. 72, 439–451.

Department for Education, 2014. Statutory framework for the early years foundation stage, DfE publications. Department for Education. https://doi.org/00266-2008BKT-EN

Doherty, J., Hughes, M., 2009. Child development: theory and practice 0-11. Longman, Harlow.

Duh, E.S., Koceska, N., Koceski, S., 2017. Game-based learning: educational game Azbuka to help young children learn writing Cyrillic letters. Multimed. Tools Appl. 76, 14091–14105. https://doi.org/10.1007/s11042-016-3829-9

Eraslan, S., Yesilada, Y., Harper, S., 2016. Eye tracking scanpath analysis techniques on web pages: A survey, evaluation and comparison. J. Eye Mov. Res. 9(1), 1–19. https://doi.org/10.16910/jemr.9.1.2

Erhel, S., Jamet, E., 2013. Digital game-based learning: Impact of instructions and feedback on motivation and learning effectiveness. Comput. Educ. 67, 156–167. https://doi.org/http://dx.doi.org/10.1016/j.compedu.2013.02.019

Essa, E.L., 2012. Introduction to early childhood education. Cengage Learning.

Feng, J., Spence, I., Pratt, J., 2007. Playing an Action Video Game Reduces Gender Differences in Spatial Cognition. Psychol. Sci. 18, 850–855. https://doi.org/10.1111/j.1467-9280.2007.01990.x

Fullerton, T., 2008. Documentary games: Putting the player in the path of history.

Furió, D., Juan, M.C., Seguí, I., Vivó, R., 2015. Mobile learning vs. traditional classroom lessons: a comparative study. J. Comput. Assist. Learn. 31, 189–201. https://doi.org/10.1111/jcal.12071

Glenn, N.M., Knight, C.J., Holt, N.L., Spence, J.C., 2013. Meanings of play among children. Childhood 20, 185–199. https://doi.org/10.1177/0907568212454751

Heintz, S., Law, E.L.C., 2018. Digital Educational Games: Methodologies for Evaluating the Impact of Game Type. ACM Trans. Comput. Interact. 25, 8:1--8:47. https://doi.org/10.1145/3177881

Heintz, S., Law, E.L.C., 2015. Game Elements-Attributes Model: a First Step towards a Structured Comparison of Educational Games, in: DiGRA Conference.

Henderson, J., Hollingworth, A., 2003. Eye movements and visual memory: Detecting changes to saccade targets in scenesHenderson, J., & Hollingworth, A. (2003). Eye movements and visual memory: Detecting changes to saccade targets in scenes. Perception & Psychophysics. Retrieved from http://lin. Percept. Psychophys. 65, 58–71.

Hessels, R.S., Kemner, C., van den Boomen, C., Hooge, I.T.C., 2016. The area-of-interest problem in eyetracking research: A noise-robust solution for face and sparse stimuli. Behav. Res. Methods 48, 1694–1712.

Hoaglin, D.C., Iglewicz, B., 1987. Fine-Tuning Some Resistant Rules for Outlier Labeling. J. Am. Stat. Assoc. 82, 1147–1149. https://doi.org/10.2307/2289392

Holmqvist, K., Nyström, M., Andersson, R., Dewhurst, R., Jarodzka, H., Van de Weijer, J., 2011. Eye tracking: A comprehensive guide to methods and measures. OUP Oxford.

Hung, C.Y., Kuo, F.O., Sun, J.C.Y., Yu, P.T., 2014. An Interactive Game Approach for Improving Students’ Learning Performance in Multi-Touch Game-Based Learning. IEEE Trans. Learn. Technol. 7, 31–37. https://doi.org/10.1109/TLT.2013.2294806

Hyde, J.S., Lindberg, S.M., Linn, M.C., Ellis, A.B., Williams, C.C., 2008. Gender Similarities Characterize Math Performance. Science (80-. ). 321, 494 LP – 495.

Isen, A.M., Daubman, K.A., Nowicki, G.P., 1987. Positive affect facilitates creative problem solving. J. Pers. Soc. Psychol. https://doi.org/10.1037/0022-3514.52.6.1122

Jackson, L.A., Zhao, Y., Kolenic III, A., Fitzgerald, H.E., Harold, R., Von Eye, A., 2008. Race, gender, and information technology use: The new digital divide. CyberPsychology Behav. 11, 437–442.

Jarodzka, H., Holmqvist, K., Gruber, H., 2017. Eye tracking in Educational Science: Theoretical frameworks and research agendas 10. https://doi.org/10.16910/jemr.10.1.3

Jaswal, V.K., Wayne, A., Golino, H., 2020. Eye-tracking reveals agency in assisted autistic communication. Sci. Rep. 10, 1–10. https://doi.org/10.1038/s41598-020-64553-9

Johnston, J., 2005. Early Explorations in Science 2nd Edition. McGraw-Hill, Inc.

Kampmann Walther, B., 2003. Playing and Gaming - Reflections and Classifications. Game Stud. 3.

Katsini, C., Abdrabou, Y., Raptis, G.E., Khamis, M., Alt, F., 2020. The Role of Eye Gaze in Security and Privacy Applications: Survey and Future HCI Research Directions, in: Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems, CHI ’20. Association for Computing Machinery, New York, NY, USA, pp. 1–21. https://doi.org/10.1145/3313831.3376840

Ku, O., Y Chen, S., H Wu, D., C C Lao, A., Chan, T.W., 2014. The Effects of Game-Based Learning on Mathematical Confidence and Performance: High Ability vs. Low Ability. J. Educ. Technol. Soc. 17, 65–78.

Kuperman, V., Bertram, R., Baayen, R.H., 2008. Morphological dynamics in compound processing. Lang. Cogn. Process. 23, 1089–1132. https://doi.org/10.1080/01690960802193688

Lai, M.L., Tsai, M.J., Yang, F.Y., Hsu, C.Y., Liu, T.C., Lee, S.W.Y., Lee, M.H., Chiou, G.L., Liang, J.C., Tsai, C.C., 2013. A review of using eye-tracking technology in exploring learning from 2000 to 2012. Educ. Res. Rev. 10, 90–115. https://doi.org/http://dx.doi.org/10.1016/j.edurev.2013.10.001

Le Meur, O., Baccino, T., 2013. Methods for comparing scanpaths and saliency maps: Strengths and weaknesses. Behav. Res. Methods 45, 251–266. https://doi.org/10.3758/s13428-012-0226-9

Leon, A.C., Davis, L.L., Kraemer, H.C., 2012. Role and Interpretation of Pilot Studies in Clinical Research. J Psychiatr Res 45, 626–629. https://doi.org/10.1016/j.jpsychires.2010.10.008.The

Liu, H.-C., Chuang, H.-H., 2011. An examination of cognitive processing of multimedia information based on viewers’ eye movements. Interact. Learn. Environ. 19, 503–517. https://doi.org/10.1080/10494820903520123

LoBue, V., Baker, L., Thrasher, C., 2017. Through the eyes of a child: preschoolers’ identification of emotional expressions from the child affective facial expression (CAFE) set. Cogn. Emot. 32, 1122–1130. https://doi.org/10.1080/02699931.2017.1365046

Majaranta, P., Bulling, A., 2014. Eye tracking and eye-based human–computer interaction., in: Advances in Physiological Computing. Springer London, pp. 39–65. https://doi.org/10.1007/978-1-4471-6392-3

Malone, T., Lepper, M., 1987. Making learning fun: A taxonomy of intrinsic motivations for learning, Aptitude,. ed, Making Learning Fun: A Taxonomy of Intrinsic Motivations for Learning.

Manly, T., Robertson, I.H., Anderson, V., Nimmo-Smith, I., 1999. The test of everyday attention for children (TEA-CH). Bury St Edmunds, UK Thames Val. Test Co.

Markopoulos, P., Read, J.C., MacFarlane, S., Hoysniemi, J., 2008. Evaluating children’s interactive products: principles and practices for interaction designers. Morgan Kaufmann.

Mason, L., Pluchino, P., Tornatora, M.C., Ariasi, N., 2013a. An Eye-Tracking Study of Learning From Science Text With Concrete and Abstract Illustrations. J. Exp. Educ. 81, 356–384. https://doi.org/10.1080/00220973.2012.727885

Mason, L., Tornatora, M.C., Pluchino, P., 2013b. Do fourth graders integrate text and picture in processing and learning from an illustrated science text? Evidence from eye-movement patterns. Comput. Educ. 60, 95–109. https://doi.org/http://dx.doi.org/10.1016/j.compedu.2012.07.011

Mayer, R.E., 2017. Using multimedia for e-learning. J. Comput. Assist. Learn. 33, 403–423. https://doi.org/10.1111/jcal.12197

Miles, C.A.L., Wood, G., Vine, S.J., Vickers, J.N., Wilson, M.R., 2017. Quiet eye training aids the long-term learning of throwing and catching in children: Preliminary evidence for a predictive control strategy. Eur. J. Sport Sci. 17, 100–108. https://doi.org/10.1080/17461391.2015.1122093

Mohd Nizam, D.N., Law, E.L.C., 2018. In the eyes of young children: A study on focused attention to digital educational games, in: Proceedings of the 32nd International BCS Human Computer Interaction Conference, HCI 2018. https://doi.org/10.14236/ewic/HCI2018.20

Moser, A., Zimmermann, L., Dickerson, K., Grenell, A., Barr, R., Gerhardstein, P., 2015. They can interact, but can they learn? Toddlers’ transfer learning from touchscreens and television. J. Exp. Child Psychol. 137, 137–155. https://doi.org/10.1016/j.jecp.2015.04.002

Neumann, M.M., 2018. Using tablets and apps to enhance emergent literacy skills in young children. Early Child. Res. Q. 42, 239–246. https://doi.org/10.1016/j.ecresq.2017.10.006

Neumann, M.M., Neumann, D.L., 2014. Touch Screen Tablets and Emergent Literacy. Early Child. Educ. J. 42, 231–239. https://doi.org/10.1007/s10643-013-0608-3

Oliemat, E., Ihmeideh, F., Al-Khawaldeh, M., 2018. The use of touch-screen tablets in early childhood: Children’s knowledge, skills, and attitudes towards tablet technology, Children and Youth Services Review. Elsevier Ltd. https://doi.org/10.1016/J.CHILDYOUTH.2018.03.028

Oliva, M., Anikin, A., 2018. Pupil dilation reflects the time course of emotion recognition in human vocalizations. Sci. Rep. 8, 1–10. https://doi.org/10.1038/s41598-018-23265-x

Orquin, J.L., Ashby, N.J.S., Clarke, A.D.F., 2016. Areas of Interest as a Signal Detection Problem in Behavioral Eye-Tracking Research. J. Behav. Decis. Mak. 29, 103–115. https://doi.org/10.1002/bdm.1867

Papastergiou, M., 2009. Digital Game-Based Learning in high school Computer Science education: Impact on educational effectiveness and student motivation. Comput. Educ. 52, 1–12. https://doi.org/http://dx.doi.org/10.1016/j.compedu.2008.06.004

Papavlasopoulou, S., Sharma, K., Giannakos, M.N., 2018. How do you feel about learning to code? Investigating the effect of children’s attitudes towards coding using eye-tracking. Int. J. Child-Computer Interact. 17, 50–60. https://doi.org/10.1016/j.ijcci.2018.01.004

Parker, A.E., Mathis, E.T., Kupersmidt, J.B., 2013. How Is This Child Feeling? Preschool-Aged Children’s Ability to Recognize Emotion in Faces and Body Poses. Early Educ. Dev. 24, 188.

Peirce, N., Centre, L., 2013. Digital Game-based Learning for Early Childhood. A state of the art report. Dublin, Ireland: Learnovate Centre.

Piaget, J., 1971. The theory of stages in cognitive development., in: Measurement and Piaget. McGraw-Hill, New York, NY, US, pp. ix, 283–ix, 283.

Ponce, H.R., Mayer, R.E., 2014. Qualitatively different cognitive processing during online reading primed by different study activities. Comput. Human Behav. 30, 121–130. https://doi.org/http://dx.doi.org/10.1016/j.chb.2013.07.054

Poole, A., Ball, L.J., 2006. Eye tracking in HCI and usability research. Encycl. Hum. Comput. Interact. 1, 211–219.

Prensky, M., 2001. Digital game-based learning. McGraw-Hill, New York;London;

Rayner, K., 2009. Eye movements and attention in reading, scene perception, and visual search. Q. J. Exp. Psychol. 62, 1457–1506. https://doi.org/10.1080/17470210902816461

Rayner, K., Reingold, E.M., 2015. Evidence for direct cognitive control of fixation durations during reading. Curr. Opin. Behav. Sci. 1, 107–112. https://doi.org/http://dx.doi.org/10.1016/j.cobeha.2014.10.008

Read, J.C., Fitton, D., Sim, G., Horton, M., 2016. How Ideas Make It Through to Designs: Process and Practice, in: Proceedings of the 9th Nordic Conference on Human-Computer Interaction, NordiCHI ’16. ACM, New York, NY, USA, pp. 16:1--16:10. https://doi.org/10.1145/2971485.2971560

Read, J.C., MacFarlane, S., 2006. Using the fun toolkit and other survey methods to gather opinions in child computer interaction, in: Proceedings of the 2006 Conference on Interaction Design and Children. ACM, pp. 81–88.

Reingold, E., Charness, N., Pomplun, M., Stampe, D.M., 2001. Visual Span in Expert Chess Players: Evidence From Eye Movements. Psychol. Sci. 12, 48–55. https://doi.org/10.1111/1467-9280.00309

Rösler, A., 2012. Using the Tobii Mobile Device Stand in Usability Testing on Mobile Devices.

Ruff, H.A., Capozzoli, M., Weissberg, R., 1998. Age, individuality, and context as factors in sustained visual attention during the preschool years. Dev. Psychol. 34, 454–464. https://doi.org/10.1037/0012-1649.34.3.454

Ruff, H.A., Capozzoli, M.C., 2003. Development of attention and distractibility in the first 4 years of life. Dev. Psychol. 39, 877–890. https://doi.org/10.1037/0012-1649.39.5.877

Ruff, H.A., Lawson, K.R., 1990. Development of sustained, focused attention in young children during free play. Dev. Psychol. 26, 85–93. https://doi.org/10.1037/0012-1649.26.1.85

Salen, K., Zimmerman, E., 2004. Rules of Play : Game Design Fundamentals / K. Salen, E. Zimmerman ; pról. de Frank Lantz.

Schall, A., Bergstrom, J., 2014. Introduction to Eye Tracking, in: Schall, J.R.B.J. (Ed.), Eye Tracking in User Experience Design. Morgan Kaufmann, Boston, pp. 3–26. https://doi.org/http://dx.doi.org/10.1016/B978-0-12-408138-3.00001-7

Schöbel, S., Janson, A., Mishra, A.N., 2019. A configurational view on avatar design - The role of emotional attachment, satisfaction, and cognitive load in digital learning. 40th Int. Conf. Inf. Syst. ICIS 2019 1–17.

Segaran, K., Ali, A.Z.M., Tan, W.H., 2019. Mediating Effects of Avatar Realism on Motivation in Game Based Learning - Review. Online J. TVET Pract. 4.

SesameWorkshop, 2012. Best Practices : Designing Touch Tablet Experiences for Preschoolers [WWW Document]. URL http://www.sesameworkshop.org/wp\_install/wp-content/uploads/2013/04/Best-Practices-Document-11-26-12.pdf

She, H.C., Chen, Y.Z., 2009. The impact of multimedia effect on science learning: Evidence from eye movements. Comput. Educ. 53, 1297–1307. https://doi.org/http://dx.doi.org/10.1016/j.compedu.2009.06.012

Shi, Y.R., Shih, J.L., 2015. Game factors and game-based learning design model. Int. J. Comput. Games Technol. 2015, 11.

Shuler, C., Levine, Z., Ree, J., 2012. iLearn II: An analysis of the education category of Apple’s app store, New York: The Joan Ganz Cooney Center at Sesame Workshop.

Snell, S., Snell-siddle, C., 2013. Mobile Learning : The Effects of Gender and Age on Perceptions of the Use of Mobile Tools. Sdiwc 274–281.

Steichen, B., Wu, M.M.A., Toker, D., Conati, C., Carenini, G., 2014. Te,Te,Hi,Hi: Eye Gaze Sequence Analysis for Informing User-Adaptive Information Visualizations, in: Dimitrova, V., Kuflik, T., Chin, D., Ricci, F., Dolog, P., Houben, G.-J. (Eds.), User Modeling, Adaptation, and Personalization: 22nd International Conference, UMAP 2014, Aalborg, Denmark, July 7-11, 2014. Proceedings. Springer International Publishing, Cham, pp. 183–194. https://doi.org/10.1007/978-3-319-08786-3\_16

Sullivan, A., Bers, M.U., 2016. Girls, boys, and bots: Gender differences in young children’s performance on robotics and programming tasks. J. Inf. Technol. Educ. Innov. Pract. 15, 145–165.

Sung, H.Y., Hwang, G.J., 2013. A collaborative game-based learning approach to improving students’ learning performance in science courses. Comput. Educ. 63, 43–51. https://doi.org/http://dx.doi.org/10.1016/j.compedu.2012.11.019

Sutcliffe, A., Namoun, A., 2012. Predicting user attention in complex web pages. Behav. Inf. Technol. 31, 679–695. https://doi.org/10.1080/0144929X.2012.692101

Sutton-Smith, B., 2006. Play and ambiguity. game Des. Read. A rules Play Anthol. 296–313.

Sutton-Smith, B., 1966. Piaget on play: A critique. Psychol. Rev. https://doi.org/10.1037/h0022601

Tabachnick, B.G., Fidell, L.S., 2007. Using multivariate statistics, 5th ed., Using multivariate statistics, 5th ed. Allyn & Bacon/Pearson Education, Boston, MA.

Tan, C.T., Bakkes, S., Pisan, Y., 2014. Inferring player experiences using facial expressions analysis. ACM Int. Conf. Proceeding Ser. 02-03-Dece. https://doi.org/10.1145/2677758.2677765

Taufik Akbar, M., Nasrul Ilmi, M., Rumayar, I. V., Moniaga, J., Chen, T.K., Chowanda, A., 2019. Enhancing game experience with facial expression recognition as dynamic balancing. Procedia Comput. Sci. 157, 388–395. https://doi.org/10.1016/j.procs.2019.08.230

Tobii AB, 2016. Tobii Pro Mobile Device Stand User ’ s Manual.

Tsai, F.H., Yu, K.C., Hsiao, H.S., 2012. Exploring the Factors Influencing Learning Effectiveness in Digital Gamebased Learning. J. Educ. Technol. Soc. 15, 240–250.

Turner, P., 2017. A Psychology of User Experience: Involvement, Affect and Aesthetics. Springer. https://doi.org/10.1007/978-3-319-70653-5

Vickers, J.N., 2011. Mind over muscle: The role of gaze control, spatial cognition, and the quiet eye in motor expertise. Cogn. Process. 12, 219–222. https://doi.org/10.1007/s10339-011-0411-2

Voyer, D., Voyer, S.D., 2014. Gender differences in scholastic achievement: A meta-analysis. Psychol. Bull. 140, 1174–1204. https://doi.org/10.1037/a0036620

Walsh, G., Foss, E., Yip, J., Druin, A., 2013. FACIT PD: A Framework for Analysis and Creation of Intergenerational Techniques for Participatory Design, in: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI ’13. ACM, New York, NY, USA, pp. 2893–2902. https://doi.org/10.1145/2470654.2481400

Wang, J., Antonenko, P., Celepkolu, M., Jimenez, Y., Fieldman, E., Fieldman, A., 2019. Exploring Relationships Between Eye Tracking and Traditional Usability Testing Data. Int. J. Human–Computer Interact. 35, 483–494. https://doi.org/10.1080/10447318.2018.1464776

Warner, R.M., 2008. Applied statistics: From bivariate through multivariate techniques., Applied statistics: From bivariate through multivariate techniques. Sage Publications, Inc, Thousand Oaks, CA, US.

West, J., Haake, A., Rozanski, E., Karn, K., 2006. EyePatterns: Software for identifying patterns and similarities across fixation sequences. https://doi.org/10.1145/1117309.1117360

Westerman, D., Tamborini, R., Bowman, N.D., 2015. The effects of static avatars on impression formation across different contexts on social networking sites. Comput. Human Behav. 53, 111–117. https://doi.org/10.1016/j.chb.2015.06.026

Wexler, B.E., Iseli, M., Leon, S., Zaggle, W., Rush, C., Goodman, A., Esat Imal, A., Bo, E., 2016. Cognitive Priming and Cognitive Training: Immediate and Far Transfer to Academic Skills in Children. Sci. Rep. 6, 1–9. https://doi.org/10.1038/srep32859

Widen, S., A Russell, J., 2010. Differentiation in Preschooler’s Categories of Emotion. Emotion 10, 651–661.

Wilson, K., Bedwell, W., Lazzara, E., Salas, E., Burke, S., Estock, J., Orvis, K., Conkey, C., 2008. Relationships Between Game Attributes and Learning Outcomes: Review and Research Proposals. Simul. Gaming - Simulat Gaming 40, 217–266. https://doi.org/10.1177/1046878108321866

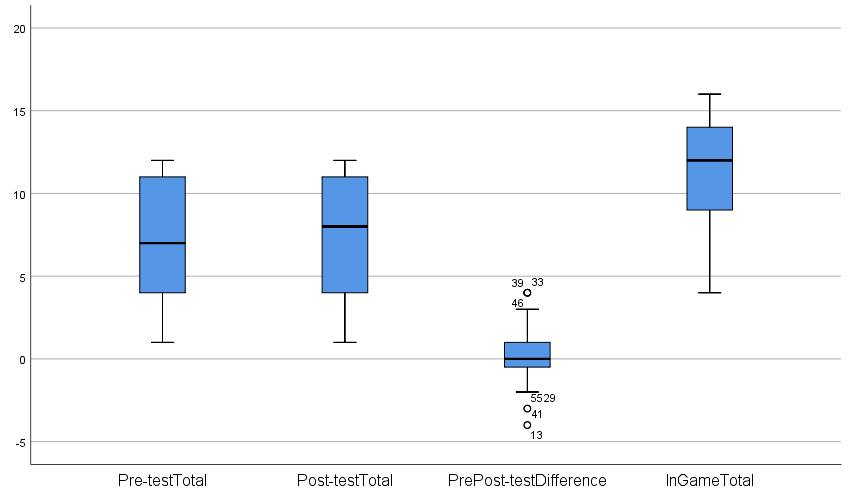
# **Appendix A.**

*Children’s facial expression developed by LoBue and colleagues (2017) that was used to encode game experience.*

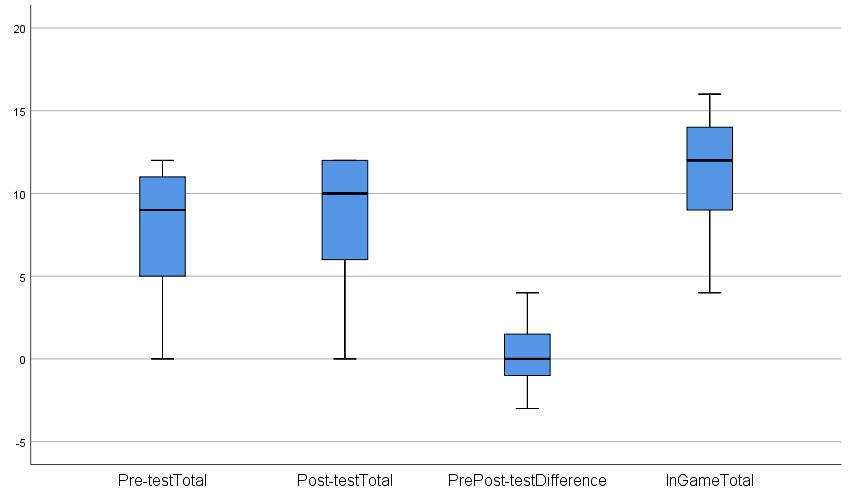
|  |  |  |
| --- | --- | --- |
|  |  |  |
| Happy/Excited  *Value = 2* | Neutral  *Value =1* | Sad/Bored  *Value =0* |

# **Appendix B.**

*Outlier examination for DEG (Boxplot – before transformation).*



*Outlier examination of Cardboard game data (Boxplot – before transformation).*



1. The overlapping text between this and our conference paper is the hypotheses on the learning effect and associated results (Section 5.2). It is necessary to include them, though it is a small % of the entire paper, because the learning game medium (digital vs. cardboard) was the main intervention. In, Mohd Nizam and Law, (2018) we mainly reported on using fixation durations for estimating focused attention (which we do *not* include here), an analysis method distinct from using scanpaths for inferring cognitive strategies that we report in this paper. [↑](#footnote-ref-1)