DIGITAL EDUCATIONAL GAMES ON NUMERACY FOR 5-YEAR-OLDS: APPLICATION OF EYE TRACKING METHODOLOGY

Thesis submitted for the degree of

Doctor of Philosophy

at the University of Leicester

by

Dinna Mohd Nizam

Department of Informatics

University of Leicester

2019

DIGITAL EDUCATIONAL GAMES ON NUMERACY FOR 5-YEAR-OLDS: APPLICATION OF EYE-TRACKING METHODOLOGY

DINNA MOHD NIZAM

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 is the main motivation for my research work that contributed to the applied body of knowledge in Human-Computer Interaction, Game-Based Learning, and Educational Technology. With the use of state-of-the-art eye-tracking technology, especially its applicability for mobile devices, a better understanding of young children's focused attention and interaction strategies when learning with DEGs can be obtained. Methodologically, my research work demonstrates how eye-tracking methods can be applied with young users.

For the empirical studies of this research work, a dedicated DEG and its Cardboard version serving as control were designed with reference to the UK Early Years Foundation Stage Framework on numeracy. The DEG was developed based on the literature on game-based learning and young children's attention span. To validate the learning effect of both versions of the game, the between-subject experimental design was employed and a paper-based knowledge test was created. The research protocols and instruments were pilot tested and were improved, especially the eye-tracking calibration process. In analysing the eye-tracking data, *fixation duration* was used as proxy for focused attention, and *gaze sequences were* used to infer interaction strategies with a refined *Gaze Sub-sequence Marking Scheme*.

Overall, the empirical findings provided valuable groundwork and implication for future work. Specifically, a more child-friendly eye-tracker hardware design is required to facilitate data collection with children playing DEGs. Practical game design for 5-year-olds needs to consider different factors, including game duration, game difficulty, voice-based instructions and interactive non-player-character.

Acknowledgement

In the name of God, the Most Gracious and the Most Merciful, I am very blessed in every moment towards the end of my PhD journey.

First and foremost, I would like to express my deepest gratitude to my supervisor Professor Dr Effie Law who has endlessly guided and advised me along this PhD journey. Her countless motivation, dedication, knowledge and constructive comments had made this journey a great learning experience. I could never imagine having a better supervisor than her. I will miss her loads.

I would also like to express my sincere thank you to Professor Thomas Erlebach my second supervisor, Dr Neil Walkinshaw my former internal examiner and Dr Fer-Jan de Vries our postgraduate tutor for their patience and recommendations during the annual viva examination. Not to forget Mrs Karen J. Smith of the department for volunteering as the narrator for my digital game and Gilbert Laycock for his IT technical support. Also many thanks to my great colleagues Matthias Heintz, Kornchulee Khanana, Samaneh Soleimani, Maryam Al Hinai, Lana Gilli and Pamela Andrade Sevillano for their generous help and support.

My sincere thanks also go to all three head teachers involved in this study for giving me the permission to conduct my empirical studies in their lovely school. I also highly appreciate the generous support from all reception class teachers and staff that gave me help during my visits. Many thanks also go to all the parents and legal guardians of the children involved in this study and their children. Without their cooperation, my empirical study would not be possible.

I would also like to give a special thank you to my sponsor the Ministry of Education Malaysia (MOE) and my employer Universiti Malaysia Sabah (UMS) for the funding and opportunity to pursue my PhD overseas with the presence of my little family. A thank you note also dedicated to my Malaysian family (MASLEICS) and friends in Leicester who had helped not only academically but also in life. Thank you for the friendship and continuous support throughout the years in Leicester. It was one crazy life experience aboard.

To my beloved parents, Mohd Nizam Abdullah@ Harchan Singh and Maimona Salim, thank you from the bottom of my heart for your endless love and advice. Daddy, I know this has always been your dream. To my mother-in-law and my beloved brothers and sisters thank you too for the amazing support. And to my beloved adoring daughter Batrisyia Dayini you have always been my inspiration for this research study, from designing the non-player character of the game until understanding how young children learn was all about you.

Last but not least my ‘partner in crime’, my dearest husband, Mohd Ezainudin Jambol who has been very supportive through bad and good times throughout this PhD journey. I really owe him heaps for sacrificing a lot especially his career. I am really grateful for having him and my daughter by my side always. THANK YOU EVERYONE…

Table of Contents

List of Figures

List of Tables

Abbreviations

AOI	Area of Interest
DEG	Digital Educational Game
EYFS	Early Years Foundation Stage
GBL	Game-Based Learning
MDS	Mobile Device Stand

Chapter 1: Introduction

1.1 Motivation

Digital educational games (DEGs) are increasingly used to support young children to learn (Neumann & Neumann, 2014), thanks to their motivational influences and proliferation of digital gadgets that are becoming more affordable and usable (Shuler, Levine, & Ree, 2012). For instance, a study on the effects of using DEG "Endless Alphabet" and "Letter School" in enhancing literacy skills reported positive outcomes for children aged two to five years old (Neumann, 2018). Another study explored learning motivation on the use of a DEG "AZBUKA" on writing among kindergarteners and reported positive impacts (Duh, Koceska, & Koceski, 2017).

While these and a handful of other studies (Aladé, Lauricella, Beaudoin-Ryan, & Wartella, 2016; Burnett, 2010; Peirce & Centre, 2013) focus on *whether* DEGs could help improve the learning performance of young children, which is a very relevant research goal per se, very few attempts have been undertaken to understand *how* they interact and learn with such games. This understanding is critical as it can not only inform the design of DEGs to enhance their effectiveness but also gain insights into teaching strategies for this age group. Nonetheless, 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, Mathis, & Kupersmidt, 2013), making it challenging to utilise research methods like think-aloud or interview. A viable alternative approach to understanding young children's interactions with a DEG is eye-tracking.

The recent development of the eye-tracking technology, especially its applicability for portable devices such as tablets and smartphones, has enabled me to conduct a research study among young children with a dedicated DEG on a tablet. The study was aimed to understand via eye tracking data how learning strategies were performed by 5-year-olds when playing a DEG on numeracy. In order to understand the influences and interactive strategies on educational games, a dedicated DEG and its cardboard version were designed and developed for this research project. The game focused on numeracy from 1 to 20 and adapted the Early Years Foundation Stage (EYFS)[1] framework in the UK

[1] https://www.gov.uk/early-years-foundation-stage

educational system. Requirements for the DEG design were from multiple resources that included a Game Based Design Learning Model by Shi & Shih (2015) . Early game prototypes and experimental setups were evaluated in a preliminary test and pilot study with 5 and 31 children respectively. Another set of 94 children from different backgrounds were involved in the Main Study. To assess the learning effect, participants were asked to complete the same, knowledge test before and after playing the game, which were referred to as *pre-test* and *post-test*, respectively. The assumption was that children would yield a higher score in the post-test than in the pre-test as a result of playing the game. 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, Read, MacFarlane, & Hoysniemi, 2008). Due to this reason and the concern of prolonging an evaluation session with young children, who typically have a short attention span, led us to decide not to use such a questionnaire but to rely on observations and videos.

Theoretically, it is well-grounded that eye-tracking measures (fixation duration) are a good proxy of attention paid to objects of interest (Steichen, Wu, Toker, Conati, & Carenini, 2014). Practically, the recent advances in the eye-tracking technology, especially the improved usability, portability and affordability of the devices (Holmqvist et al., 2011), make it more applicable for a variety of research studies. Furthermore, increasingly sophisticated algorithms for gaze sequences analysis such as String-edit algorithm (Eraslan, Yesilada, & Harper, 2016; Le Meur & Baccino, 2013) and visualisation have been developed, contributing to the creation of software applications such as *eyePatterns* (West, Haake, Rozanski, & Karn, 2006). These encouraging software developments have facilitated this research project on DEG with young children using the eye-tracking methodology.

1.2 Research Goal and Research Questions

The goal of this study is to evaluate the learning effects of a dedicated DEG on numeracy and the applicability of eye-tracking methodology for gaining insights into interaction strategies of 5-year-old children. The following research questions (RQs) were derived to address the goal of this study:

- *RQ1:* To what extent does the learning effect of a dedicated DEG on numeracy differ from its cardboard version?

- *RQ2:* How can children's attributes and game experience help understand the learning effect induced by both game learning media?
- *RQ3:* How can children's focused attention help understand the learning effect induced by the two game learning media?
- *RQ4:* How is the achievement level of children related to their focused attention (i.e. fixation duration) to relevant and irrelevant objects in the games?
- *RQ5:* Are learning tasks drawing on recall more difficult than those on recognition for young children for both game learning media?
- *RQ6:* How are children's interaction strategies derived from gaze sequence analyses related to the learning effect induced by both game learning media?

1.3 Research Contributions

This research project presents a number of contributions drawn from the practical application of the eye-tracking methodology with young children, game development, experimental process and results of the empirical studies. Details of the overall contribution and related body of knowledge are discussed in Section 8.2. In summary, the contributions are as follows:

- Developing a deeper understanding of how children interact with DEGs and learn from such interactions;
- Bridging the gap of insufficient eye-tracking work with young children.
- Developing a modified calibration board for young children and eye-tracking application.
- Identifying limitations and strengths of using eye-tracking methodology with young children.
- Refining the Gaze Sub-Sequence Marking Scheme for studying young children's interaction strategies when interacting with DEGs.
- Validating the use of the existing Game-Based Learning Model for DEGs.
- Estimating the attention span of 5-year-old children for DEGs.

1.4 Outline of the Thesis

Chapter 2: Literature Review. This chapter describes the fundamental theoretical background of this research project, reviewing child development theories that help understand learning with DEGs, existing educational frameworks and game-based learning design requirements for the primary content of the DEG and Cardboard development. The literature review also addresses the gender preference for educational technology and attention spans of young children that are related to the game development. Towards the end of this chapter, the motivation of using eye-tracking in young children and the gap to be bridged by this research project are presented.

Chapter 3: Methodology. This chapter describes the research design, approaches and methods used to carry out the empirical work in this research project. It also presents the ethics consent and recruitment process, the initial game development for both digital and cardboard, and data collecting method device (eye-tracker) and instrument (pre and post-test).

Chapter 4: Preliminary-Test and Pilot Study. This chapter describes the preliminary test and pilot study that were conducted before the Main Study. The preliminary -test was conducted in a non-school environment while the pilot study was implemented in a formal school environment around Leicestershire. Requirements and issues in regards to the eye-tracker set-up, paper-based knowledge test, and game design were identified and discussed in this chapter

Chapter 5: Main Study. This chapter outlines the analysis methods in the main study of this research project. Methodological improvements from the preliminary -test and pilot study were applied. The analysis methods that are divided into three sections; Fundamental, Method 1 and Method 2 are presented in this chapter.

Chapter 6: Results and Discussion.

In this chapter the result from the three analysis sections are report and discuss. First the basic statistical analyses of the relations between Learning Effect and other factors, which did not involve any eye-tracking data, were reported. Further analyses that involved two different eye-tracking analysis methods were presented subsequently. Method 1 focused on how focused attention (fixation duration) on relevant and irrelevant UI objects in the games was related to the Learning Effect with children's Achievement Level being co-variates. Method 2 used gaze sequences instead to study how Learning Effect was related to Gender and Achievement Levels. Analysis of gaze sequences and visualisation data (Heatmaps) were also reported as the last part of this chapter.

Chapter 7: Lesson Learned. This chapter reflects the lesson learned through-out the whole project. Challenges encountered during the implementation of study and limitations that affected the results or restricted the process of the empirical research are discussed in this chapter.

Chapter 8: Conclusion. This final chapter reflects all six research questions in this thesis and concludes the findings for each RQ. A detailed contribution section of this research project is presented here. This thesis concludes with possible resolutions for future work direction.

Figure 1.1 gives an overview of all the chapters in this research projects and presents how they are connected from one to another.

1.5 Publication

A full paper entitled "In the Eyes of Young Children: A Study on Focused Attention to Digital Educational Games" was published and presented at the 32nd British Human Computer Interaction Conference 2018, 2-6 July, Belfast, Northern Ireland. The paper was about eye-tracking application issues learnt from the pilot studies and the focused attention (fixation duration) results of the Main Study. Parts of Chapter 3, 4, 5 and 6 were used in this paper.

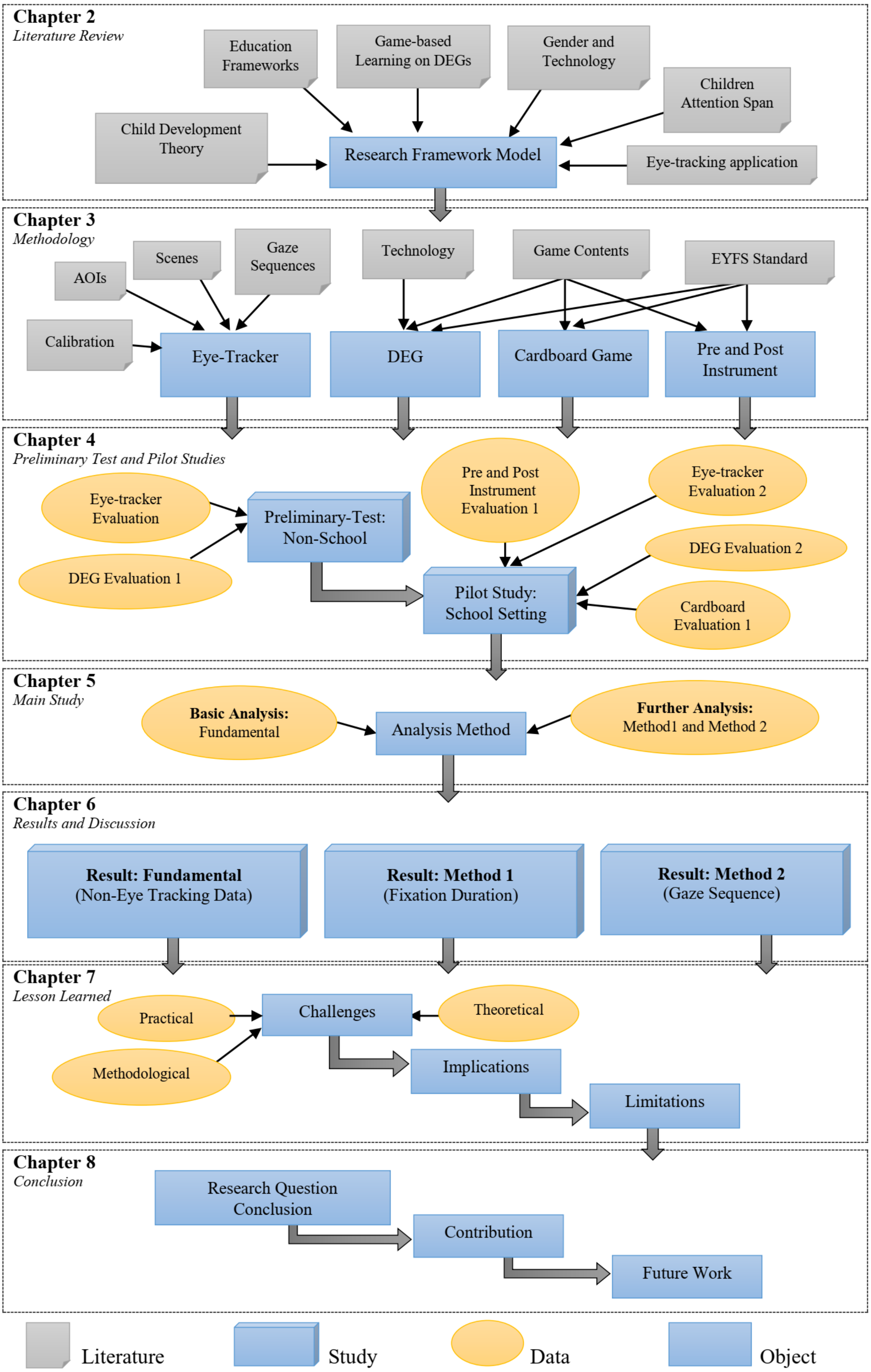


Figure 1.1: The Research Project Structure

Chapter 2: Literature Review

This chapter describes the fundamental theoretical background of this research project, reviewing child development theories that help understand learning with DEGs, existing educational frameworks, and game-based learning design requirements for the basic content of the game development. The literature review also addresses the gender preference for technology and attention spans of young children that are related to the game development. Towards the end of this chapter, motivation for the use of eye-tracking in young children and gaps that are aimed to be bridged by this research project are presented.

2.1 Child development theory

Among different theories of cognitive development, Vygotsky's theory on the Zone of Proximal Development (ZPD) in collaborative settings (Shabani, Khatib, & Ebadi, 2010) and Piaget's stage-based model contributed significantly to today's pedagogical approaches to teaching young children (Doherty & Hughes, 2009). As the DEG developed in this research project is a single-player game, the focus is then on the latter. Specifically, Piaget's work focuses on how children construct knowledge independently as they grow in five different stages (Piaget, 2000). Children of five years old – the target group of this research project - are in the preoperational stage when they start to recognise symbols and use language to understand the world, despite their capability of logical thinking is still limited (Johnston, 2005). Furthermore, in Piaget's theory, play is seen as an important vehicle for a child to explore the world around her and how a child plays is an indicator of her cognitive development. Accordingly, play can be delineated in three stages: functional play for developing sensorimotor skills (from birth up to 2 years old); symbolic and pretend play for acquiring experiences to build constructive concepts (between 2 and 7 years); and games with rules for enhancing social skills (7 years and older).

The constructivist learning theories exemplified by Vygotsky and Piaget became prominent in the mid-1980s in the field of educational psychology. Prior to that, the related work has largely been grounded in the Information Processing Model built upon a computer metaphor (Goswami, 2010). Both schools of thought have their strengths and weaknesses (Gosvāmī & Goswami, 2008). Although the Piagetian theory provided a particular framework for children development, it did not put emphasis on the role of social factor in a child learning development, a contrast to the Vygotskian theory where social interaction is the primary emphasis (DeVries, 2000). The fundamental differences between Piaget's and Vygotsky's theories are characterised by their underlying philosophy, social factor towards learning, the nature of child development, and the various cognitive processes of children (Lourenço, 2012).

Of particular relevance to this research is the Stage Theory in the Information Processing Model (Lutz & Huitt, 2003). Accordingly, there are three stages of memory - sensory, short-term and long-term. Short-term memory receives information before terminating (forgetting) or processing it into further memory stages known as long-term memory involving information elaboration and coding (Figure 2.1). Sensory perception relates to information gathered from the environment such as scent, sound and sights while short-term memory (working memory) serves as the part where new information is gathered temporarily before being stored in long-term memory or to be terminated completely (Huitt, 2003). Long-term memory is where information is stored and can be retrieved through the recognition and recall process.

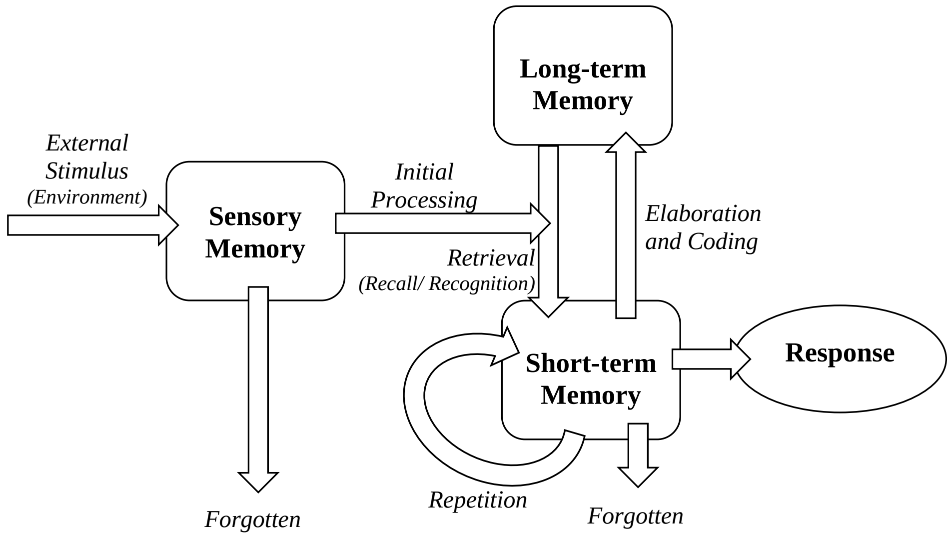


Figure 2.1: Atkinson and Shiffrin (1968) Information Processing Model cited in Lutz and Huitt, (2003).

In the retrieval process, recognition refers to the process of information comparison on encountered experiences of an object or event with the memory and normally involves a single decision process on perceived familiarity whereas recall is related to remembering objects, facts or events that need to be retrieved from the memory bank and normally involves a two-stage process - searching (retrieving) and decision-making (recognition) (Baddeley, 1997). Recognition and recall features are essential in the game development of this research project.

2.2 Educational Frameworks

Many nations have an early childhood learning framework in their education system. In Malaysia, the National Preschool Curriculum Standard (Kurikulum Standard Prasekolah Kebangsaan - KSPK) was introduced in 2010 to pre-school providers by the Ministry of Education Malaysia. The curriculum has six crucial main learning areas to help prepare young children before entering primary school (KementerianPelajaranMalaysia, 2010). The six learning areas are communication; spiritual, attitude and values; personal development; humanistic; science and technology; and physical development. In this learning framework, mathematics skills are grouped under the science and technology umbrella. While in the United Kingdom, the Department for Education initiated the “Early Years Foundation Stage – EYFS” which divides learning and development into prime and specific areas (Table 2.1) for pre-school or infant school providers. This learning framework is to ensure a strong foundation for young children before transferring into junior or primary school (Department for Education, 2014). Mathematics is one of the four specific learning areas of the EYFS.

However, the usage of DEGs in early childhood learning classrooms, particularly for Mathematics, remains uncertain. Guidelines for deploying DEGs in these educational frameworks remain limited. An effort in promoting DEGs in classrooms in Australia, the State Department of Queensland conducted iPad trials in traditional classrooms and provided an iPad Advice Guide for regular schools and iPad Usage Guideline for special education teachers and schools (Department of Education, 2012). However, these efforts remain low, and the need for more guidance on DEGs in educational frameworks is one of the motivating factors for this research project, particularly for the EYFS in the United Kingdom.

Table 2.1: Learning Areas of the EYFS Framework (Early Education, 2012).

<table>
<tr><th colspan="4">PRIME AREAS</th></tr>
<tr><td colspan="2">Personal, Social and Emotional Development
• Making relationships
• Self-confidence and self-awareness
• Managing feelings and behaviour</td><td>Physical Development
• Moving and handling
• Health and self-care</td><td>Communication and Language
• Listening and attention
• Understanding
• Speaking</td></tr>
<tr><th colspan="4">SPECIFIC AREAS</th></tr>
<tr><td>Literacy
• Reading
• Writing</td><td>Mathematics
• Numbers
• Shape, space and measure</td><td>Understanding the World
• People and communities
• The world
• Technology</td><td>Expressive Arts and Design
• Exploring and using media/materials
• Being imaginative</td></tr>
</table>

2.3 Game Based Learning on DEGs and Requirements

Piaget's model on play and the concepts of recognition and recall from the Information Processing model can inform game-based learning (GBL) in general (Frost, Wortham, & Reifel, 2008). Games, given their strong motivational power in engaging players, have increasingly been used for supporting learning. A study on the use of a digital educational application "Martha Speaks: Dog Party and Super Why" reported positive outcomes on learning vocabulary for children between age three to seven years old (Chiong & Shuler, 2010). Another preliminary study adopted DEGs in kindergarten classrooms to explore the learning effectiveness of DEGs as compared with that of traditional teaching methods, and reported that the use of the DEGs resulted in better learning outcomes (Zaranis, Kalogiannakis, & Papadakis, 2013).

A study by Furió et al. (2015) investigated the learning efficiency and satisfaction between an educational game on mobiles and traditional classroom learning. The study reported both methods showed positive outcomes in learning the water cycle however motivational effect was found better on the educational game and suggested educational games could be a useful learning tool for eight to ten-year-old children. A study by Papastergiou (2009) also studied the motivation and effectiveness of DEG among high school students through a computer memory concept game and reported positive outcomes on motivation and effectiveness regardless of gender. A study by Sung and

Hwang (2013) studied the application of DEGs in a collaborative learning environment among sixth grade science learners and reported that the student's learning attitude and motivation presented promising outcomes. Additionally, the student's achievement and self-efficacy improved by the collaborative educational game. Moreover a study by Fu-Hsing et al. (2012) studied other factors that influenced the learning efficiency of DEGs for sixth-grade students and reported that factors such as learning ability, playing skills, prior knowledge and game experience could influence the effectiveness of knowledge on DEGs. A study by Oskar et al. (2014) however studied the student's confidence level and performance in learning mathematics using a DEG with fourth-graders. The study reported that both high and low performing students significantly improved their confidence level towards learning mathematics as compared to the paper-based learning approach.

Furthermore, a study by Hung et al. (2014) studied the influence of applying an interactive educational game with different difficulty levels in learning geographical concepts for third-grade students. The study reported that a significant improvement in learning was found and moderate difficulty levels contributed to better performance. Erhel and Jamet (2013) investigated the impact of different instructions types of a DEG known as ASTRA among university students and reported that learning instructions presented in an educational module can promote deeper learning as compared to entertainment instructions that are presented in a game dimension. Likewise, the study suggested that educational game can promote learning and motivation by providing frequent feedback to the students in the gameplay. Since many studies reported positive outcomes from the use of DEGs, some studies had reported negative effects of using DEGs among children. A study by Moser et al. (2015) examined the level of transferring knowledge deficit among children between 2.5 to 3 years with video and touchscreen puzzles. The study reported that both groups showed some level of deficit and the younger group showed a greater transfer deficit. While a study by Barr (2013) studied how infants use their memory from books, television and touchscreen. However, the study reported memory constraints among infants while learning using 2D media. Therefore, considering the inconsistent findings of DEGs this study intends to study the learning effect of learning media.

Prensky (2001) discussed six basic features of a game: rules, goals, feedback, challenge, interaction, and story. With the advent of new technologies such as touchscreen, the way games are played has changed. A useful design guideline on touchscreen applications for young children highlights several features such as navigation support, layout orientation, design placement and suitable multimedia objects (SesameWorkshop, 2012). Apart from taking into account these features, the DEG in this research project was primarily developed and designed based on the Shi and Shih's (2015) GBL model (Figure 2.2).

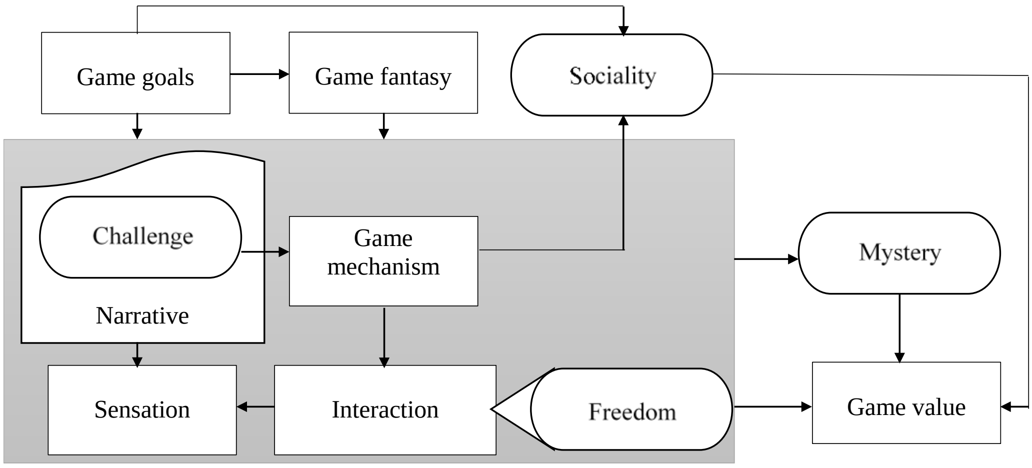


Figure 2.2: Shi and Shih (2015) Game-based Learning design model.

The GBL Model has eleven game design factors which a designer can accommodate into their game: game goals, game fantasy, sociality, challenge, narrative, game mechanism, sensation, interaction, freedom, game value and mystery. Depending on the nature of the game, not all factors are compulsory. For instance, a single player game or non-online game does not have the sociality factor because no player participation is presented. Explanation on how the game of this project was derived and guided by this GBL model is described in Section 3.2.4.

2.4 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, Burleson, Tai, Muldner, & Woolf, 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, Lindberg, Linn, Ellis, & Williams, 2008) and the latter in girls (Voyer & Voyer, 2014). However, recent 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 & 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 2nd-grade. They reported that at such 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 the interview method with kindergarteners is questionable. Also there was no objective pre-test to provide a baseline for comparing the post-test performance. Nor were there any post-intervention interviews to identify any changes of the attitudes inferred from the earlier interviews. For drawing more solid conclusions, more work beyond the pilot study reported in Sullivan and Bers (2016) is required.

In contrast to Arroyo et al.(2013) and Sullivan and Bers (2016), in studying the potential gender effect in benefiting from DEGs on computer science for high school students, Papastergiou (2009) reported no significant differences in learning gains 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 (A. 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, Ihmeideh, & Al-Khawaldeh, 2018; Sullivan & Bers, 2016) and their findings are inconsistent. As to address these issues, the gender effect in learning with technology (DEG) was explored in the Main Study (Section 6.1.4).

2.5 Attention Span in Young Children

Attention is an age-old complex concept that has been researched in psychology for more than a century. It can be defined as "a state of focused awareness on a subset of the available perceptual information" [2], comprising the perceptual, cognitive, neurophysiological and behavioural aspects. The multi-component view of attention accounts for different types of (in)attention and their related cognitive as well as behavioural issues (Ruff, Capozzoli, & Weissberg, 1998). Elaboration of such types and issues, however, is beyond the scope of this research project.

Young children are known to have difficulty in maintaining attention on assigned tasks or objects and are easily distracted (Richter & Courage, 2017). A number of studies have measured and studied attention in different ways. A study by Oakes, Ross-Sheehy and Kannass (2004) observed attention using the concept of attentional inertia and attentional state on infants between 6.5 to 9 months. Ruff et al. (1998) and Lander (2002) used the idea of sustained attention in their studies among preschool children. Another study by Ruff and Capozzoli (2003) classified attention into three types: focused attention, settled attention and casual attention while observing the development of attention in early childhood. Casual attention and settled attention differ in terms of the intensity an object

[2] http://www.apa.org/research/action/glossary.aspx

is being looked at with the former being less intense than the latter; the intensity can be inferred by a person's verbal as well as behavioural responses. However, neither casual nor settled attention is strong enough to build any engagement with the object, which is the case for focused attention (Ruff et al., 1998; Ruff & Lawson, 1990).

In the Main Study (Chapter 5), attention span is referred to the focused attention duration which increases as children age. Focused attention duration is the period when concentration on a specific task happens, involving minimal body movements, intensive facial expressions and a body posture that shows an interest (Ruff & Capozzoli, 2003). According to Ruff and Capozzoli (2003), the average focused attention span of 47 months (~ 4 years old) was about 260 seconds (4.3 minutes) as derived from their empirical study with young children playing with construction (problem solving) toys. One might extrapolate their finding to assume that 5-years-old's would be about 5 to 6 minutes[3]. However, bearing in mind that such extrapolation can be speculative. As Ruff et al. (1998) work was published more than 20 years ago, the finding might no longer be valid because children's attention span is known to have been attenuated due to the use of technology (Patel, 2017; Perles, 2013). Nonetheless, there has been limited recent empirical evidence for the attention span of 5-year-old, apart from some grey literature[4,5] suggesting that attention span could be 10-25 minutes, which may include all three types of attention - casual, settled and focused –discussed above.

2.6 Eye Tracking Application on Children and Measurements

In the field of HCI, eye-tracking has become a standard research approach (Majaranta & 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, Duchowski, & Majaranta, 2011). Capturing participants' eye movements in natural conversations and behaviours is the primary benefit of mobile eye-tracking (Bulling & 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.,

[3] The focused attention span was shorter for free play with a mix of construction and symbolic toys, which 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 & Lawson, 1990).

[4] https://www.dealwithautism.com/how-important-is-attention-span-for-children-with-adhd-and-autism/

[5] http://www.parentingpress.com/media/is-this-a-phase_excerpt2.html

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 & Roggen, 2011; Rayner, 2009).

A number of eye-tracking studies through fixation or paid attention have been conducted to investigate cognitive processes. A study by Ariasi and Mason (2011) applied eye-tracking in examining cognitive process variations from reading debatable and non-debatable text among university students and reported that visual attention predicted learning on debatable text. Furthermore a study by Liu and Chuang (2011) employed eye-tracking method to investigate college student's cognitive process while viewing multimedia elements on a webpage. The study found that attention was paid more to text elements and concluded that eye-tracking technology could promote insights into cognitive process. In addition, a study by Mason et al. (2013) utilised eye-tracking technology to investigate the cognitive process of eleventh-grade students between concrete and abstract pictures in explaining a text. The study reported that abstract pictures presented a better understanding of the text as shown from the eye-tracking results. Moreover, a study by She and Chen (2009) applied eye-tracking to examine the cognitive process of middle-school children between different interaction and sensory modes. The eye-tracking results revealed that more visual attention was paid to animation with narrated multimedia materials and a relation between fixation duration and depth of learning was found.

Additionally, eye-tracking studies through gaze strategies and behaviour patterns also contributed to insights of the cognitive process. A study by Ponce and Mayer (2014) applied eye-tracking in investigating the different gaze strategies used between a normal note taking approach and a graphic organizer approach while learning. The eye-tracking result revealed two type of strategies, namely linear and generative, used by the college students. Mason, Tornatora and Pluchino (2013) applied eye-tracking in investigating the visual behaviour patterns while reading illustrative science text among fourth graders in school. The study reported three visual behaviour patterns applied by the children while reading and a significant relation between eye-tracking data on reading comprehension and prior knowledge was found. Furthermore, a study by Tsai et al. (2012) utilised eye-tracking to examine the scan patterns used while solving multiple-choice science questions among university students and revealed successful solvers focused on relevant objects while unsuccessful solvers seek around and experienced difficulties. The study also suggested that eye-tracking technique could promote a deeper level of understanding the cognitive process in learning.

However, the number of studies involving young children with eye-tracking and multimedia learning remains limited, despite the growing number of young multimedia applications 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 & Cagiltay, 2018). Therefore, this study aims to apply the eye-tracking methodology to investigate the learning process as well as the learning effect for young children interacting with digital educational games.

Among well-known eye-tracking indicators are fixation, saccade (Lai et al., 2013), 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 & Baccino, 2013). Heat maps visualise fixations gathered over a specific group of participants to indicate the distribution of focused attention at a time (Rösler, 2012) whereas scanpath is the combination of both static and dynamic aspects of eye movements that create a complete sequence (Poole & Ball, 2006). These indicators are categorised as *diachronic* as it takes time into account in the measurements and are the least studied category of eye-tracking studies (Le Meur & Baccino, 2013).

Fixation duration (*synchronic*) 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) whereas gaze sequences (*diachronic*) that are derived from scanpaths are known to understand the human cognitive processes (West et al., 2006). These indicator and with numbers of gaze sequence techniques such as *position-based weighted model* (Eraslan et al., 2016) and software applications such as *eyePatterns* (West et al., 2006) are used in the research study to investigate the use of eye-tracking methodology with young children on DEGs.

2.7 Chapter Summary

Today's pedagogical approaches to teaching young children with or without the support of educational technology are often referenced to Piaget's stage-based model (Section 2.1). Specifically, digital educational games (DEGs) have been increasingly used to support learning. While many of the studies on DEGs have reported positive outcomes, some have reported certain negative effects in terms of knowledge transfer (Section 2.3). There has been limited work on studying the learning effect of DEGs for children as young as five years old. The current development of eye-tracking technology, especially its applicability for portable devices such as tablets and smartphones enable empirical studies to be conducted on DEGs with young children. Gaze measures are a good proxy of focused attention (Section 2.5). Moreover, gaze sequence analyses derived from AOI (Area of Interest) visits are utilised to infer children's interaction strategies when interacting with DEGs. Overall, the research questions examined in this thesis (Section 1.2 and Chapter 5) were aimed to develop a better understanding of how young children interact with DEGs and learn through such interactions.

Chapter 3: Methodology

This chapter describes the research design, methods and tools used to carry out the empirical work in this research project. It also presents the ethics consent and recruitment process, the initial game development for both the digital and cardboard version of the game, and data collecting device (eye-tracker) and measurement instrument (pre and post-test).

3.1 Research Design Terminology

This research project consisted of a preliminary -test, pilot study and one main study. Throughout this thesis, the term "research project" refers to the whole PhD thesis project while the empirical work refers to the preliminary test, pilot study and the main study (Figure 3.1). The preliminary -test was implemented in a small scale of participant while the pilot study was implemented in a larger scale that followed a formal experimental protocol as shown Figure 4.5. The main study involved two different eye-tracking analysis methods, designated as Method 1 – focused attention inferring from fixation duration (Section 6.2) and Method 2 – interaction strategy inferring from gaze sequences (Section 6.3).

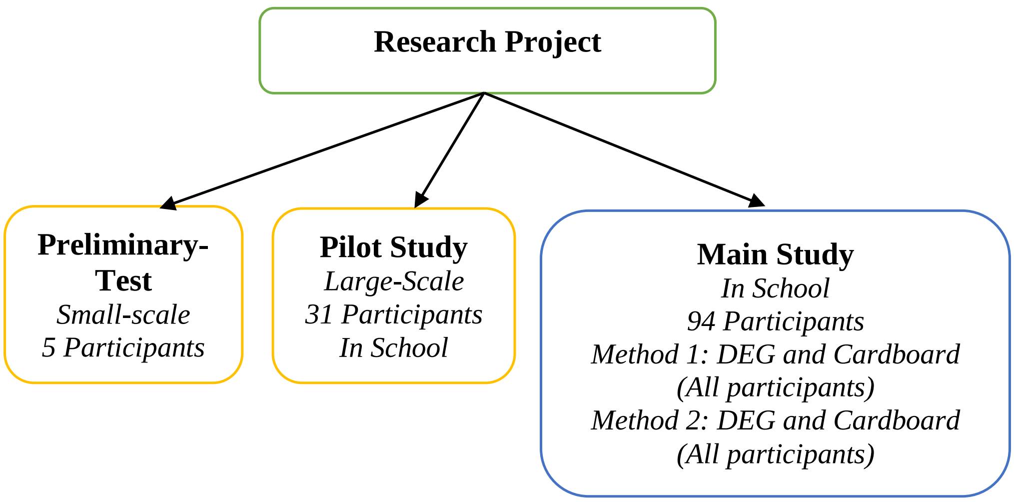


Figure 3.1: Research design of this thesis.

3.1.1 Ethics approval and recruitment

The main goal of this research project was to gain insights via eye-tracking data on how 5-year-olds performed learning activities with a dedicated DEG addressing numeracy from 1 to 20 with reference to the EYFS standard (Department for Education, 2013) of the UK educational system. Participants involved in the empirical studies of the research project were 5-year-old children attending reception (or foundation) years in the county Leicestershire in the UK. Altogether 36 children were involved in the preliminary -test and pilot study (Chapter 4) while another 94 children involved in the Main Study (Chapter 5). Empirical work for the Pilot Study and Main Study were held in participating schools with approval and informed consent forms from head-teachers and parents. Throughout the recruitment process, intention letters, consent forms and study flyers (Appendix A.2 to A.6) were sent via postal mails and emails to more than 100 infant and community schools in Leicestershire. Upon the approval of participating head-teachers, the researcher worked in the school voluntarily a few days before the experimental sessions. This volunteering 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 that involved human participants are required to go through an ethical approval review by the University's ethics committee. Approval must be completed before approaching any participants. Before submitting for an ethical application, a prerequisite Disclosure and Barring Service (DBS) check is required from the UK Government body to ensure that a researcher or individual is suitable and eligible to work with vulnerable people in a community. As this research project involved young participants, a DBS check was made and granted with a certificate number: 001500710472 and followed by an ethical approval, reference number: 2907-dmn9-computerscience from the University of Leicester Ethics Review Committee. A copy of the correspondent email for approval can be seen in Appendix A.1.

3.1.2 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 having the same design and content. In the experiment, every participant had to complete a session, with the goal of studying whether a digital educational game or its cardboard version could lead to better learning effect (dependent variable- DV). Each participant was randomly selected from each reception classroom in the participating infant and community schools to attend an individual experimental session. Figure 3.2 demonstrates the between-subject experimental design involving both intervention groups who were asked to complete a pre- and post-test before and after playing the game.

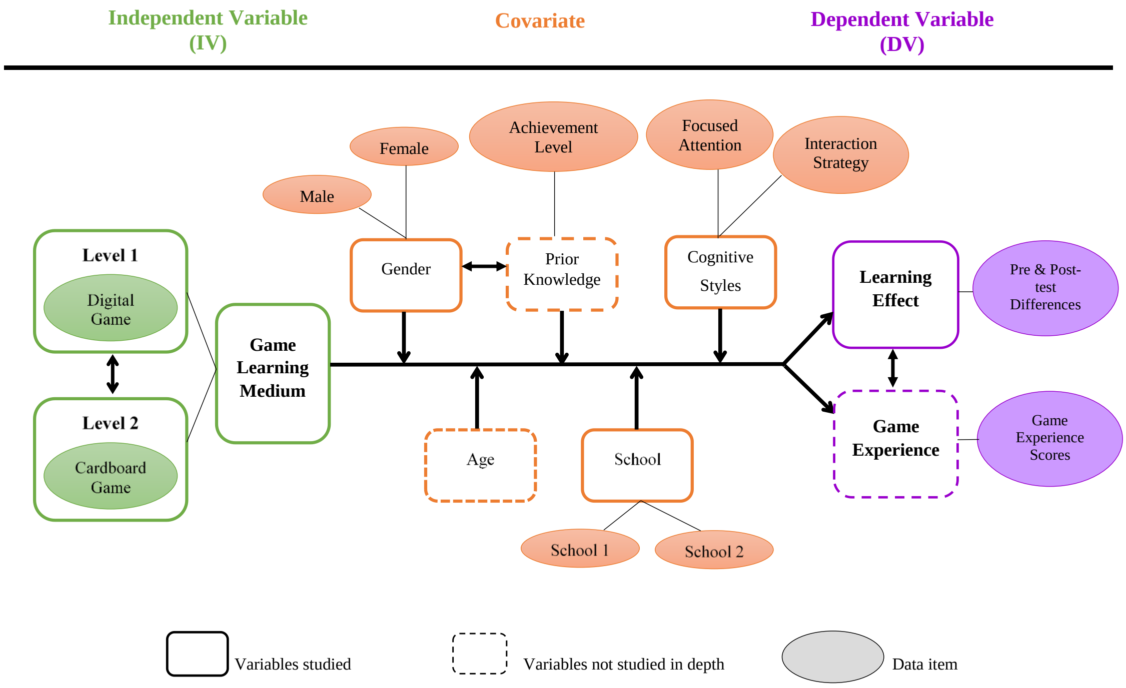


Figure 3.2: The research model for this study.

Gender, Prior Knowledge, Cognitive Style, Age and School were covariates that characterise participants and were beyond the control of this research project. However, Age was irrelevant since all participants were from the same age group, 5-year-old. Learning effect (DV) was derived from the pre and post-test score differences of both game learning media. Cognitive Style was derived from gaze path and heat map data captured by the eye-tracker. Prior Knowledge was measured through the achievement level of each participant. While game experience is generally relevant to game-based learning research, it is not that the main focus of this research project addressing primarily the learning effect of the DEG. Nonetheless, game experience was evaluated by facial expressions of individual children recorded by the eye tracker.

3.1.3 Researcher Role

The main role of the researcher (i.e. the author of this thesis) in this research project was manifold, including (i) designing and developing the DEG and its cardboard version, research protocol, measuring instruments; (ii) collecting data in the Preliminary test, Pilot study and Main study according to the research protocol; (iii) and (iii) implementing the analyses using the basic as well as multivariate statistical methods and specific eye-tracking methods such as gaze sub-sequences. Data collection included gathering ideas for the game design and the paper-based instrument (Section 3.2.2 and Section 3.4). A potential bias from the researcher with a background of teaching children could have impact on the game design. Data collected from the empirical studies also included children's game experience (Section 5.3). A second coder with an IT engineering background assisted in the game experience data analysis.

The researcher was also responsible for analysing eye tracking measurement by observing every participant's scanpaths. Each gaze plot on AOI visits was observed and coded (Section 3.3.6). Furthermore, a personal journal to record the researcher's own opinions and judgements throughout the research project was used in reporting the results and discussion in this research. Overall, the researcher played an indispensable role in this research project from the inception of the initial ideas till its conclusion.

3.2 Game Design and Development

The game contents of this research project was designed by referring to existing learning materials, the Shi & Shih (2015) game-based learning design model and the UK Early Years Foundation Stage Framework on numeracy. The development of the game content also took into consideration of 5-year-old children attention span in the design. Figure 3.3 illustrates the resources of the game content and design of this research project. Details of how these resources inform the game design are discussed in Section 3.2.1, Section 3.2.2 and Section 3.2.4.

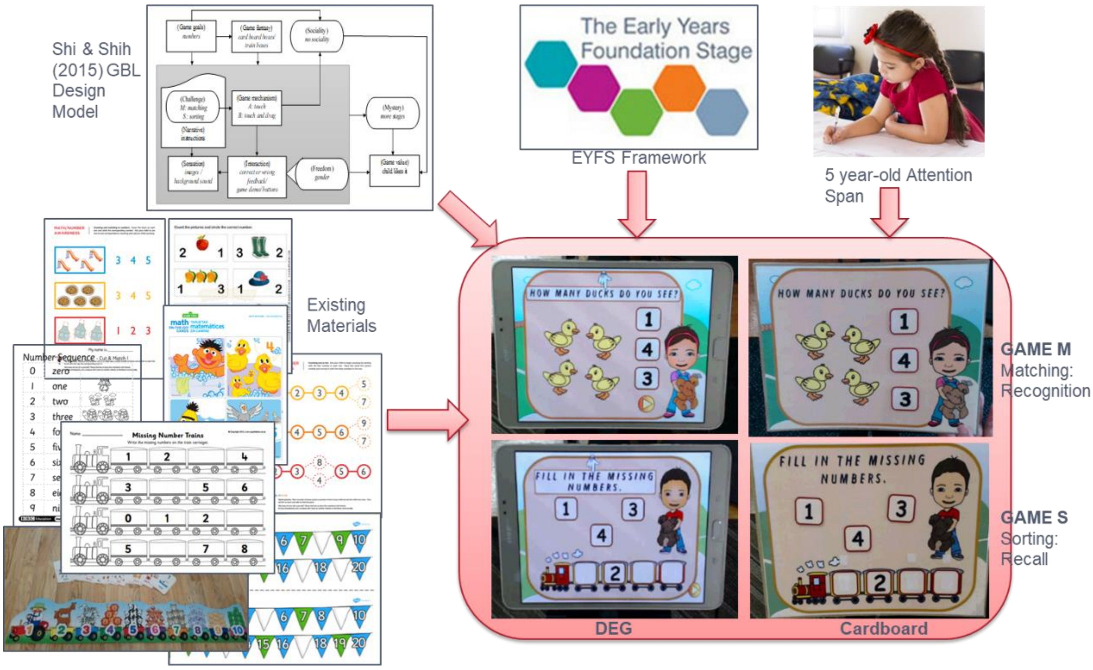


Figure 3.3: Resources of the game content and design.

3.2.1 United Kingdom EYFS Standards

The 5-year-olds involved in this study were children attending reception or foundation years in their school. Infant schools in the UK must follow a curriculum known as Early Years Foundation Stage Framework (EYFS) in their teaching, learning and care (Department for Education, 2014). The framework consists of three main prime areas and four specific areas (Table 2.1). In this research project, mathematics was selected to be the subject content of the game as this is the basics of mathematics that children should learn with fun, building their affinity for maths later on in their learning stages. According to the EYFS outcome evaluation, a child between 40 to 60 months and over should be capable of counting numbers between 1 to 20 and sort them correctly in order (Department for Education, 2013). Therefore, the numbers involved in the game design were between 1 and 20.

3.2.2 Game Content

The educational game for this research project was developed to support children's ability to recognise and recall numeric symbols from 1 to 20, which was also compatible with the recommendation of the UK national curriculum. One of the researcher's roles before the experiment was to collect ideas for the game. The researcher's own 4-year-old child at home and a couple of children of the same age from personal contacts in their home were observed informally to understand and gain insights about existing learning materials. Observations such as the design of toys, especially age-appropriateness and potential educational values were noted. It was observed that parents would normally choose educational toys that are age- appropriate and safe to use.

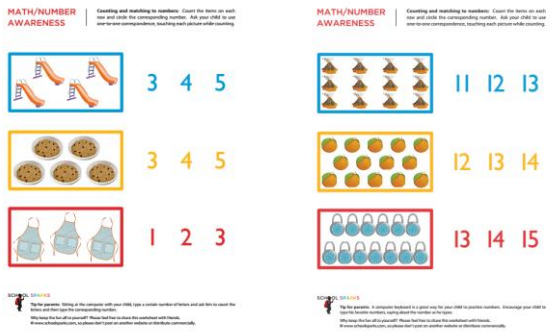
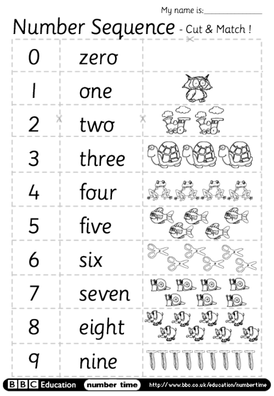
Some children in the kindergarten where the researcher had volunteered as a teaching assistant also contributed to the game design. The observation informed how learning numbers are typically introduced to young children at this age in the school context and what kind of toys as well as non-digital games children play that may help them build the concepts of numeracy. According to observations, the children were introduced to numbers with physical objects from everyday applications and to the concept of grouping things according to assigned numbers. Figure 3.4 (left) illustrates the matching and counting leaves activity while Figure 3.4 (right) illustrates the placing numbers activity that were carried out in classroom.

Figure 3.4: Counting Leaves (left) and Placing numbers activity (right).

Overall, the observations proved useful to inform the game design. Moreover, the guidelines from the EYFS framework for 5-year-old children focusing on the capability and expectation in mathematics were taken into account in the design. Other resources that inspired the game design are described in Table 3.1 and

Table 3.2.

Table 3.1: Learning materials that were used to design Game *M* (where *M* = Matching).

RESOURCE	ADAPTED INTO GAME '*M*'
School Sparks (Source: www.schoolsparks.com/kindergarten-worksheets)  	- Position of counting objects on the left side and number choices on the right side were used in the game. - The number box options of this resource inspired the use of 3 number box options () of Game *M*.
Sesame Street (Source:www.sesamestreet.org/toolkits)  	- The big style images () and layout from this resource were adapted into the game to attract children's attention. - The space in between images was also taken into account to give children ample of space for touching gestures and counting.
BBC Education (Source: http://www.bbc.co.uk/schools/numbertime)  	- The 'cut and match' task in this worksheet inspired the use of number box options of both game versions (), where in the digital game children could 'touch and drag' their answer while in the cardboard game children could pull-out their answer after counting the object images.
Turtle Diary (Source: www.turtlediary.com)	- This resource inspired the use of clear and understandable images

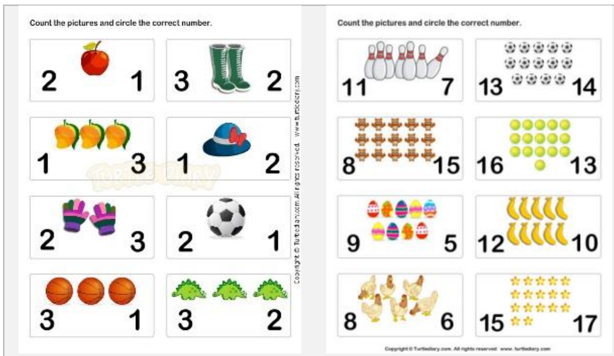
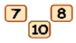
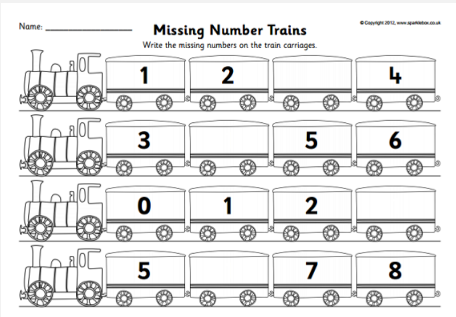
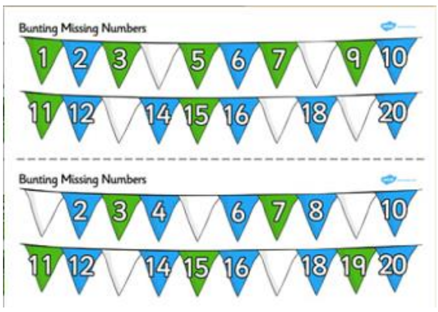
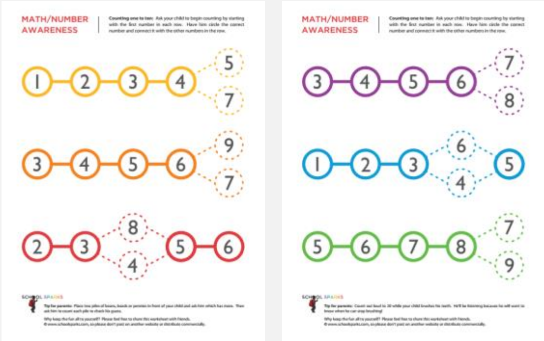
	(, ) that could be understood by young children.

Table 3.2: Learning materials that were used to design Game *S* (where *S* = Sorting).

RESOURCE	**PARTS ADAPTED INTO GAME 'S'**
Universe Of Imagination: Numbers and Words Learning Pack (ToysRUs) 	- The number choices (1, 2, 3, 4) that were given to the player when solving the tractor and trailer puzzle in this resource inspired to number box options () in Game *S*.
SparkleBox Teacher Resources Limited (Source: www.sparklebox.co.uk) 	- The locomotive engine and train compartment of this resource inspired the train engine and compartment design () in Game *S*. - The number of train compartment per task was also inspired this resource.
Twinkl Foundation (Source: www.twinkl.co.uk) 	- The number clues in-between the empty triangle banner in this resource inspired the number clues and empty train compartment () in Game *S*.

School Sparks (Source: www.schoolsparks.com/kindergarten-worksheets)  	- The number choices and number shapes in this resource also helped inspire the number box choices design for the game. Instead of using circles, squares ( ) were used to fit into the train compartment.

The game consisted of two parts: Matching (Game *M*) and Sorting (Game *S*) with recognition- and recall-based tasks, respectively. Both games progressed from easy to hard in four levels. This was to build children's engagement with the game, which nevertheless 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 (Figure 3.5: left) whereas in Game *S*, the task was to sort the numbers correctly into the empty train compartments at the bottom of the page (Figure 3.5: right). Interactions with the digital and cardboard games are different and discussed in Section 3.2.4 and Section 3.2.5, respectively. Throughout the game, a non-player character (NPC) accompanied the player until the end. 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). A complete display of all the game pages for both Game *M* and Game S can be seen in Appendix A.9 and A.10.

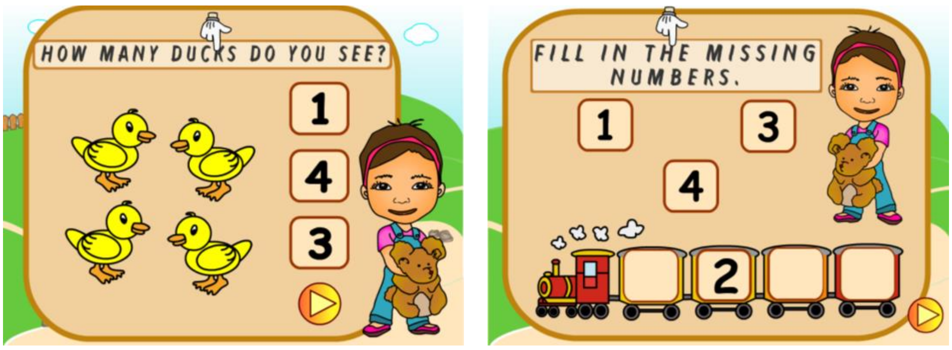


Figure 3.5: Page layout of level 1 (Page1) for Game *M* (left) and Game *S* (right).

3.2.3 Technology used in the DEG Development

The software used to develop the digital educational game was Adobe Animate CC which was previously known as Adobe Flash Professional. The software provides a platform for game designers and web developers to create powerful graphic and rich animation content. Designers can easily sketch and design their objects or characters with the user-friendly authoring tool and animate them using the frame-by-frame timeline feature in the software. The two non-player characters created in the game for this study were sketched manually, then scanned and digitally sketch using Animate CC. Other objects in the design were mostly created digitally.

To allow interactivity into the digital game design 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 scripts were coded onto graphical objects that needed control or actions to give interactivity in the game. Among graphical objects that were scripted in the game for this research project were the buttons, instruction boxes and the train compartments in Game *S*.

Once the digital game was complete, it was exported and published using Adobe AIR to package related files of the game into a single file so the game application could be used on other devices such as tablets and smartphones. Adobe AIR is a cross-operating runtime system that combines different file formats used in the Animate CC software into a package and delivers them on multiple portable devices (Adobe System Incorporated, 2018). Since the device used in this research was a tablet using an Android operating system. The Adobe AIR was set for Android during publishing.

3.2.4 Concepts for Digital Educational Game Development

The DEG in this research project was primarily designed and developed based on the Shi & Shih (2015) game-based learning (GBL) design model as shown in Figure 3.6. Other issues such as children's short attention span and non-complex user interfaces for young children (Peirce & Centre, 2013; SesameWorkshop, 2012) were also taken into account in the digital game design.

The GBL model has been used to guide developers by defining the game design factors of a DEG. For this research project, the *game goal* was to teach the players to recognise and recall numeric symbols from 1 to 20 by introducing simple puzzle tasks. The game adapted a cardboard *game fantasy* which was common learning material for young children in a school environment. Since the DEG was a single-player game, no *social* component exists because no communication was possible between players. However, the children had the *freedom* to choose their preferred non-player character (NPC) of interest at the beginning of the game on the main page.

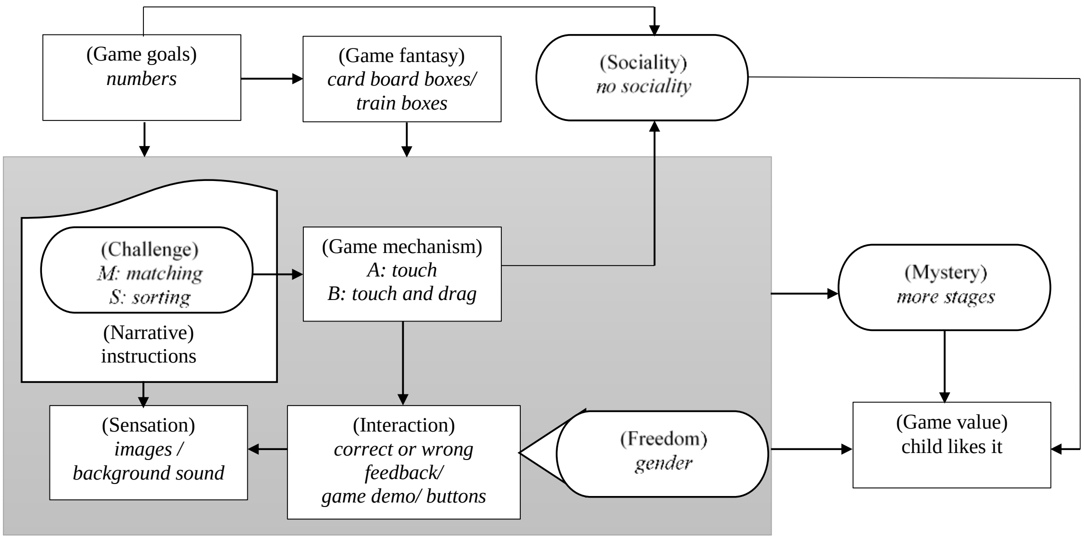


Figure 3.6: The DEG game design factors based on Shi & Shih (2015) GBL model.

The game had two types of *challenges*: matching (Game *M*) and sorting (Game *S*) which progressed from easy to hard in four levels. Every page had a *narrator* reading the instructions to assist the young children on what to do especially for those who were unable to read at this age. The *game mechanism* was simple for the young children to understand: to mainly touch the correct answer or to 'touch and drag' the answer. Apart from presenting suitable images and background sounds in creating the game *sensation*, the game provided different *interactions* such as the interactive feedback (star or sad face) for every correct and wrong answer, the animated hand on the instruction bar and the demonstration page at the beginning of Game *S*. Evaluation and improvement of the DEG was done through a preliminary test and pilot study discussed in Section 4.1 and Section 4.2.

3.2.5 Cardboard Game Development

A cardboard version of the educational game was designed 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. In this version, children could easily pull-out objects (as in Game *M*) and stick them back on the cardboard (as in Game *S*) with the help of Blu Tack or Velcro strips similar to some conventional learning materials. The Blu Tack also assisted the game from slipping away since the cardboard game was played on the slant recording area that had limited space situated under the camera device and above the eye tracker which was supported by the MDS (Figure 4.10).

Figure 3.7: Level 1 of cardboard Game *M* (left) and cardboard Game *S* (right).

During the gameplay experiment, the researcher's presence played a significant role. The researcher had to read out instructions and provide verbal and non-verbal feedback to every answer pulled-out and returned by the child. The researcher was also responsible for switching the cardboard pages throughout the game. Figure 3.7 shows Level 1 of both cardboard versions of Game *M* and Game *S*. A complete display of all the cardboard game pages can be seen in Appendix A.10.

3.3 Eye Tracker

3.3.1 Basic concepts of eye-tracking

Eye tracking is a method used for measuring eye movements of a person such as studying what or where they are looking at or the attention given to a specific spot. The process involves five crucial elements: the eye-tracker, invisible near-infrared light or illuminators, camera, image processing algorithms and gaze mapping algorithms (Tobii Eye Tracker, 2015). The near-infrared light is created by the eye tracker and reflects it to the participant's cornea and pupil as seen in Figure 3.8. The high definition camera then records images of the eye reflection patterns and calculates the position of the eye using the image processing algorithm and gaze mapping algorithm. Metrics produced from eye-tracking can be statistical (i.e., fixation duration, fixation count) or visualisation (i.e., heat maps and gaze plots).

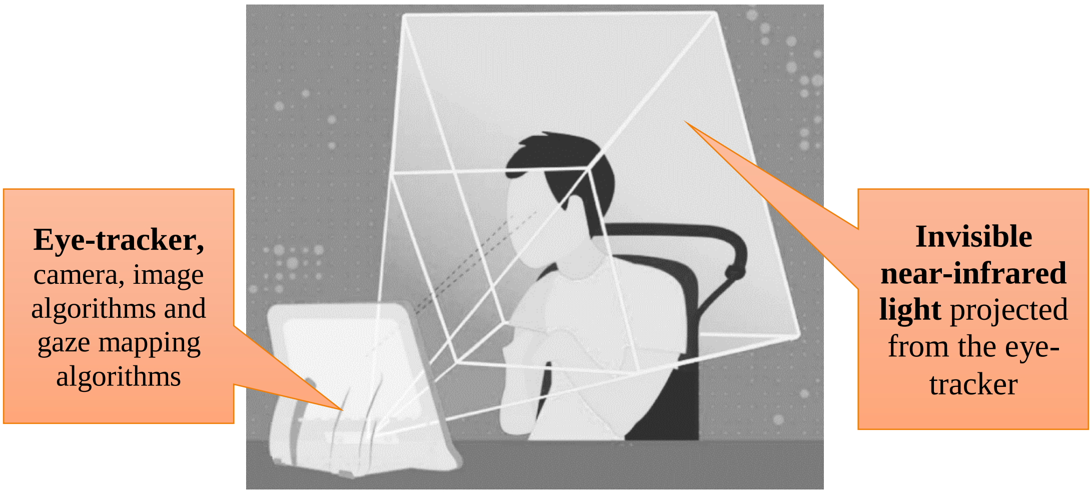


Figure 3.8: How eye-tracking works with a screen base eye tracker. Image source from (Tobii Dynavox, 2017)

There are various terms in eye-tracking methodology. In this research project, eye-tracking terms that are used are explained as follows:

Recording duration. Recording duration refers to the total time of when the simulation recording takes place. Beginning from when the eye-tracker starts recording until when the eye-tracker terminates the recording. The duration format is generally in milliseconds (Tobii AB, 2014).

Fixation. Fixation refers to the static position of an eye on a specific area being viewed (Schall & Bergstrom, 2014). Fixation are measured as fixation duration or fixation count.

Fixation count. Fixation count refers to the number of occasions an individual fixates a specifically defined area known as AOI. The numbers include returning fixation visits to the AOI. Statistically, fixation count is usually calculated with fixation duration to measure descriptive statistics such as the mean fixation duration.

Fixation duration. Fixation duration refers to the duration of each fixation within an AOI. The duration of fixation includes all returning fixation to the same AOI. Fixation duration is generally calculated as the total sum or mean fixation duration when calculated with fixation count. It was previously known as fixation length (Tobii AB, 2014).

Saccades. Saccades are quick movements of the eye from one fixation to another in a sequence (Schall & Bergstrom, 2014). In other words, fast jumps from one fixation to another fixation are called saccades.

Area of Interest (AOI). AOI is used to link eye tracking measurements (e.g. fixation duration) to specific areas of the stimulus displayed. By using the eye-tracking software (i.e. Tobii Studio), the AOI is drawn on parts of the stimulus according to the objective of a study. The eye-tracking software can then calculate the measurements needed for the study. The positioning of AOIs for the Main Study is described in Section 3.3.5.

Heat maps. Heat maps are visualisation of fixations accumulated over a specific group of participants to indicate the distribution of focused attention at a time (Rösler, 2012). They are represented by colours and depends on how long an area is looked at. Red usually shows the longest time taken on an area or highest number of fixation while green is the least time taken on a certain area (Tobii AB, 2014). Yellow normally indicated the average number of fixation.

Gaze plots. Gaze plots are series of fixations and saccades that are presented by circles and lines in the eye-tracking recording (Rösler, 2012). The circles in Figure 3.9 refers to fixation duration and lines are saccades. Gaze plot mainly displays the location, sequence and time spent looking on a section. The usage of gaze plots in the Main Study of this research project is described in Section 3.3.6.

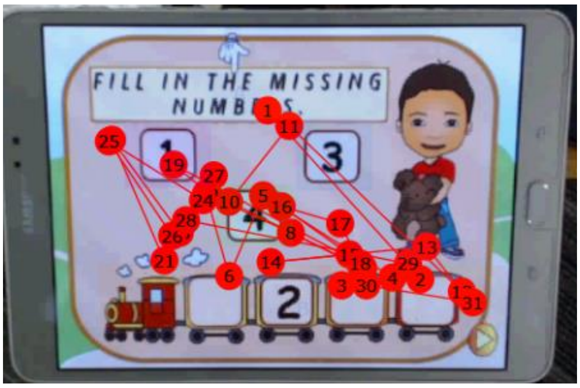


Figure 3.9: Gaze plots example presented by circles (fixations) and lines (saccades).

3.3.2 Hardware and software

Methodologically, it is challenging to work with children aged five years old, given their low verbal ability. Typically it is hard for these young children to describe on which objects they focus in a learning environment and explain why they do so, given their immature verbal ability (Parker et al., 2013). A practical alternative to gain insight into how children learn is to derive attention and gaze behavioural patterns from eye-tracking data. The recent development of eye-tracking technology, especially its applicability on portable devices such as tablets and smartphones, has enabled studies on DEG to be conducted with young children.

The eye tracker used in this research project was a Tobii X2-30 (Figure 3.10: left) incorporating with a mobile device stand (MDS). The MDS is a frame that supports the tablet with the installed DEG, eye tracker and camera device together (Figure 3.10: right). The Tobii X2-30 device that records the gaze data is placed below the participant, allowing her or him to play in a normal way with the tablet. Technically, the MDS has eight standard configurations for Tobii X2-30 which can be applied subject to the device type (tablet) used and the participant's height (Tobii AB, 2016). In this research project, configuration C2 (i.e. eye-tracker angle -5,2°; eye-tracker distance -8cm; device angle 0°; height difference between device and eye-tracker 5,7cm) was used since the participants in this research were young children who are typically short.

Figure 3.10: The Tobii X2-30 device (left) and Mobile Device Stand (right).

For the eye tracking recording to be effective, a calibration process is necessary to get a proper eye tracking position and allow the retrieval of good data quality. The distance between the participant and the eye tracker surface (Figure 3.11) has to be in the range from 60 to 65cm allowing a maximum gaze angle of 31º (Tobii AB, 2016). The distance is indicated via the tracking status window on the Tobii Studio 3.3.1 software that operates the eye tracker. After the correct range has been detected, the calibration process then continues by going through the calibration points which are indicated with numbers from 1 to 5 on the calibration plate.

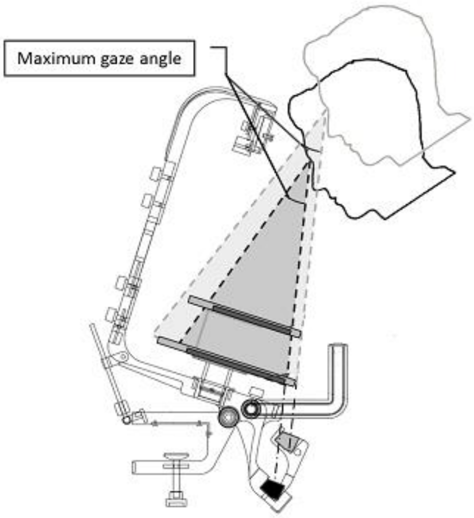


Figure 3.11: The gaze angle and range in the study represented by the dark grey area. Image source from Tobii Pro (Tobii AB, 2016)

The Tobii Studio software also retrieves metrics such as recording duration, fixation duration and fixation count which can be exported into other analysis software. In this research project, two software applications were used for analysis: IBM SPSS Software for advanced statistical analysis (v.24) and eyePatterns for gaze sequence analysis. Fundamental processes such as segmenting recording scenes, defining AOIs and plotting gaze sequences were done before analysing the data. A process flow of how data was obtained and extracted from Tobii Studio can be seen in Figure 3.12.

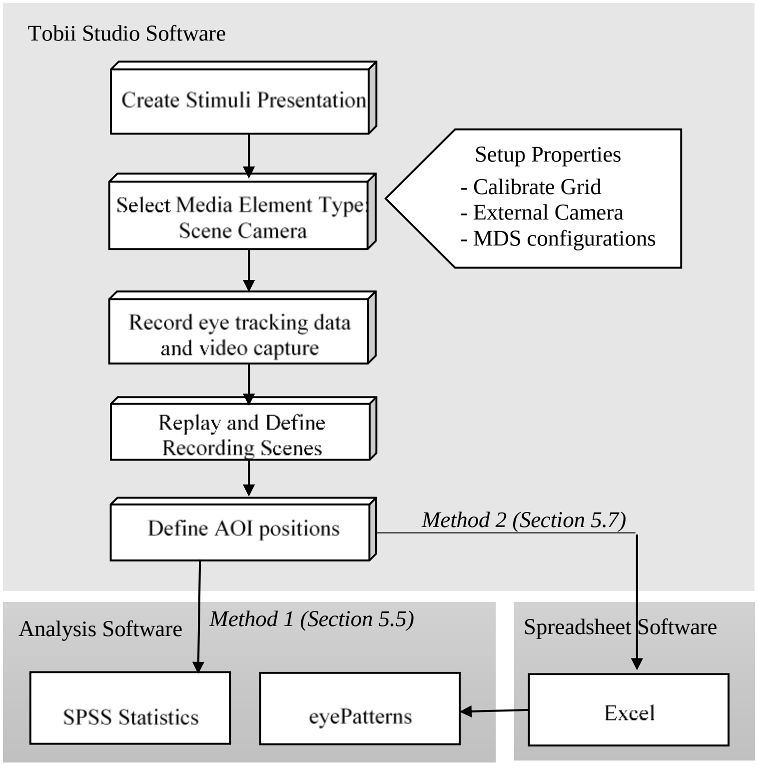


Figure 3.12: The processes of how data was obtained and extracted from Tobii Studio in this study.

At the beginning of the process flow, the stimuli presentation was created in the Tobii Studio workspace by selecting the related media element type from the media toolbars in the software. Among media types that can be associated with Tobii Studio for analysis include images, webpage, movies, scene camera and pdf elements. For this research project, the scene camera media type was used because it involved an object (tablet) being recorded with an external camera via the Tobii Studio software. This media element type enables video and gaze data to be captured simultaneously to provide saccades and gaze plots. After selecting the media type, other technical properties such as the calibration grid (Section 3.3.3), external camera and MDS configuration was set-up before the recording session. Once complete, the recordings were divided into scenes by defining the start and end point of a section that contained relevant gaze data (Section 3.3.4). The scenes were then applied to the defined area of interest (Section 3.3.5) to calculate statistical data (Method 1) and to create visualisation data (Method 2).

3.3.3 Calibration

Calibration is a process that measures the characteristics of a participant's eye by using an internal 3D eye model within the Tobii Studio software. Information gathered such as light refraction, eye reflection and eye shapes were inputs to calculate the gaze data (Tobii AB, 2014). Throughout the calibration process, participants were asked to look at the calibration points (number indicators) on the calibration plate (Figure 3.13) while the eye information is being collected and calculated.

The default number of calibration points in Tobii is five (1 to 5). However, depending on the participants and research design these points can be set into 2 or 9 points. In this research project, the default number of calibration points were maintained, but the design surface of the calibration board was modified to accommodate the participants' characteristics. Modification of the calibration board is discussed in Section 4.1.1.

Figure 3.13: The original calibration plate before being modified for this study.

3.3.4 Segmenting Recording Scenes

Segmenting and creating shorter scenes of the eye tracking recordings enables statistical and visualisation data to be produced. In this research project, each participant's recording was divided into short scenes based on the pages of the games. In total there were 13 scene pages, which included 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 significant amount of data, only the scene pages of Game *M* (*M1* to *M4*) and Game *S* (*S1* to *S4*) were used for most of the analysis in this research project. Scenes *M1* to *M4* were pages associated with the matching activities (i.e. recognition) whereas *S1* to *S4* were scenes related to the sorting activities (i.e. recall).

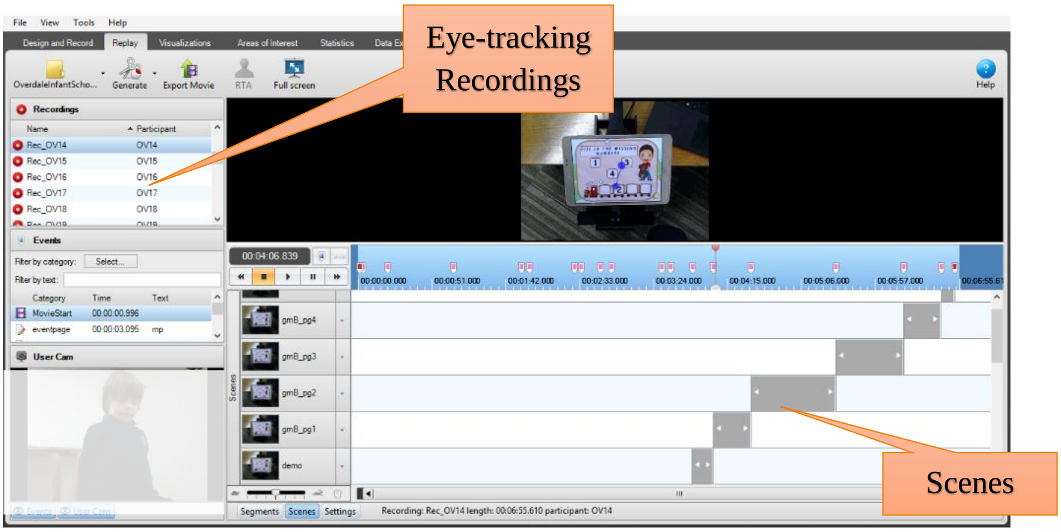


Figure 3.14: Segmenting Scenes from a participant's eye-tracking recording.

Figure 3.14 shows the process of dividing the eye-tracking recording into smaller scenes of a participant. At the bottom left side of the screen, one can view a video capture of the participant during the session (it is made blurry for the sake of anonymity), which helped identify unrecorded or missing eye-tracking data.

3.3.5 **Positioning Area of Interest (AOIs)**

In the eye-tracking methodology, areas of interest (AOIs) are used to link eye tracking measurements (e.g. fixation duration) to specific areas of the stimulus displayed. The AOI statistics facilitates the interpretation of eye-tracking data (Hessels, Kemner, van den Boomen, & Hooge, 2016). According to Orquin, Ashby and Clarke (2016), AOIs are 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 & Reingold, 2015). Due to the experimental stimuli design in this research project, where overlapping of AOIs may happen, a smaller AOI margin (≈0º margin) was used to balance the proportion of fixations (Orquin et al., 2016). AOIs were defined based on relevant (e.g. counting objects) and irrelevant images (e.g. non-player character).

Six AOIs were positioned for every selected page of Game *M* and six AOIs for every selected page of Game *S* as shown in Figure 3.15. In Game *M*, five user interface (UI) objects were identified as relevant: instruction *A*, counting objects *B*, answer box *C*, answer box *D* and answer box *E* whereas there was only one irrelevant UI object non-player character (NPC) *F*. A slightly different structure was used for Game *S*, given its different design: instruction *A*, train *B*, answer box *D*, answer box *E* and answer box *F* were identified as relevant UI objects whereas NPC *C* was the only irrelevant UI object. By defining the AOIs and scenes beforehand, statistical data can then be calculated using the Tobii Studio statistical tool.

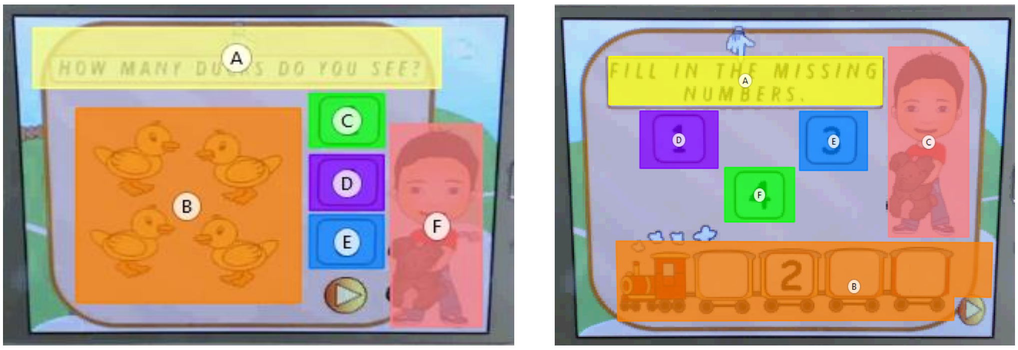


Figure 3.15: Area of Interest (AOIs) of Game *M* (left) and Game *S* (right).

3.3.6 **Gaze Sequence (EyePatterns)**

Gaze sequence analysis is a specific type of scanpath analysis that analyse eye-tracking data through visualisation. For the analysis of the Main Study, 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 (Figure 3.16). The fixation sequences were then fed into *eyePatterns*[6], a software program that identifies sequence patterns and similarity between fixation sequences (West et al., 2006).

[6] https://sourceforge.net/projects/eyepatterns/

Figure 3.16: Extracting gaze plots into fixation sequence in Excel.

However, for the Main Study, *eyePatterns* was only used 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, this 40-AOI fixation sequence (also known as extended sequence): *"BABBBBCBBBFBCDBCBBBCBBAAAACBCABBCBBBBBDB"* was converted into a 24-AOI gaze sequence (also known as collapsed sequence) "*BABCBFBCDBCBCBACBCABCBDB* with the *eyePatterns* software. In this collapsing process, consecutive fixations were gathered together such as "BBBB" to "B" and "AAAA" to "A" forming 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 (West et al., 2006) shown by the repeated fixations. The repeated fixation in the fixation sequence (Steichen et al., 2014) also indicates the approximate duration taken by each individual. A detailed gaze sequence analysis is discussed in Section 6.1.

3.4 Paper-based Pre- and Post-Test Design

For this research project, a paper-based pre and post-test were designed based on multiple EYFS materials described in Table 3.1 and Table 3.2. These resources were also materials that were used among teachers of the participating schools of this research. Once the paper-based test was designed two teachers of the participating school were given a copy to validate the difficulty level of the task for 5-year-olds. Both teachers agreed with the difficulty level, and gave advice on constructing the instruction of both tasks in the paper-based test.

The pre- and post-test were to evaluate the learning effect of both learning media, digital and cardboard games. The paper test was divided into two parts: matching task and sorting task. Initially, the matching task had five questions with a maximum score of 5 while the sorting task had four questions with a maximum score of 12 as shown in Figure 3.17. However, changes were made after the pilot study discussed in Section 4.2.1 especially on the question layout. Both tasks were designed to be similar to the DEG and cardboard game activities to evaluate the learning effect. In the experiment, each child was required to complete the pre and post-test before and after the game-play session. The technique on how the questions were revealed to the young participants was also evaluated and improved for the main study discussed Section 4.2.1.

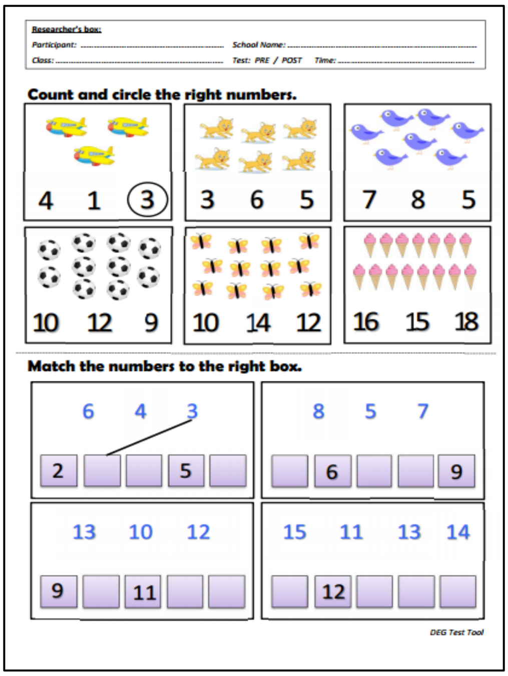


Figure 3.17: The initial Paper-based Pre and Post Test Design.

3.5 Chapter Summary

This chapter mainly describes the research model, game development, methods and tools used to carry out the empirical work in this research project. The research design applied in this research is based on a between-subject experimental design. The learning media (i.e., game) developed for this study was dedicated to 5-year-old children on numeracy. The game contents were referenced to multiple resources, including existing learning materials, the UK Early Years Foundation Stage Framework, existing Game-Based Learning Design Model and the literature on the attention span of 5-year-old children.

Two versions of the game were developed for the experiment, the Digital Educational Game and its Cardboard version. An eye-tracker was used to collect data of the children's eye movements and a paper-based knowledge test was used to evaluate the learning effect. Ethical consent was applied before reaching any participants.

Chapter 4: Preliminary Test and Pilot Study

A preliminary test and a pilot study were conducted to test the research protocol and instruments being used on a small scale to identify any shortcomings and address them before implementing it on a larger scale (Leon, Davis, & Kraemer, 2012). This chapter describes the preliminary test and pilot study that were conducted before the Main Study. The preliminary test was implemented in a non-school environment while the pilot-study was implemented in a formal school setting (Figure 3.1). Issues that were identified and subsequently improved included problems with the eye-tracker set-up (e.g. room lighting, body features and calibration plate), the paper-based test readability and game design flaws (e.g. navigation support, interaction and feedback)**.**

4.1 Preliminary Test: Non-School Setting

The preliminary test involved five 5-year-old children in a non-school environment. In this small-scale preliminary test, the primary focus was on the eye tracking set-up with young children and their acceptance towards the device. The game prototype was also evaluated to capture requirements for refining the prototype. According to the work of Read et al. (2016) and Walsh et al. (2013) on Participatory Design (PD) with young children, gathering feedback on a preliminary design of an artefact is considered a viable approach to eliciting user requirements from this target group.

4.1.1 Initial Eye Tracker Set-Up and Evaluation

Initially, the type of the eye-tracker used in this research was used on adults. However, in this research, participants consisted of young children that had smaller body frames than adults. Therefore, the device had to be examined and altered to fit the smaller body frames of young children. This section discusses issues that were identified and subsequently improved and were not limited to the device set-up only but also the surroundings of the set-up.

Lighting. In the experiment room, reflections from ceiling lights prevented the eye tracker device from scanning and recording the participant's pupil (eye). To enable smooth recording of the gameplay, the eye tracker device had to be placed away from ceiling lights above to avoid reflections beaming into the eye tracker. In addition to that, dark areas in the room were also avoided to enable the camera to capture explicit facial expressions of the participants.

Body features. In some cases, a five-year-old may have smaller than average body frame. One of the participants was unable to reach the tablet (DEG) around the mobile device stand (MDS) handle due to the participant's petite body frame. A solution was to remove the handles of the MDS (Figure 3.10) to allow the participant to be within the reach of the tablet. Unfortunately, this particular participant experienced another problem, his body was covering the eye tracker's surface after removing the MDS handles. Consequently, the participant had to be eliminated from the pre-test. However, for the subsequent pilot study, the children were informed not to cover the eye-tracker surface during the gameplay session. Unfortunately, there was a risk of disengaging the children from the gameplay.

Calibration plate. The original calibration plate that came with the MDS (Figure 3.13) was not feasible for young participants. The small number indicators (1, 2, 3, 4, and 5) on the plate were not visible and attractive for young participants to look upon. Consequently, the lack of constant attention to the calibration plate made the calibration process more difficult, causing insufficient calibration data being gathered by the eye-tracking software (Tobii Studio) to measure and evaluate the participant's pupil (eye) features. Modification of the calibration plate was done for the subsequent pilot study by replacing the small number indicators with coloured stars (i.e., red, blue, yellow, green, black) and glued on to the calibration plate (Figure 4.1). Colour indicators were preferable since 5-year-old children are more familiar with colours. As to further smoothen the calibration process, children were asked to identify the colours in advance to make sure they recognised each star colour.

Calibration duration. The calibration process was a time-consuming process dependent on individual participants' eye features and cooperation. The process starts from measuring the distance of the eye-tracker surface and the participant's eyes indicated by the Tobii Studio tracking window. Next, the participants were asked to look and focus on the calibration points in a sequence. This step was error-prone when the calibration points (number indicators) could not draw the young participant's attention. When calibration data was insignificant, the whole calibration process had to be repeatedly done. This repeated process caused the young participant to lose interest in the task, given their short attention span. Nevertheless, this process could not be dismissed since it was an essential component in eye-tracking methodology. However, the improved calibration plate had slightly reduced the calibration duration in the subsequent pilot study.

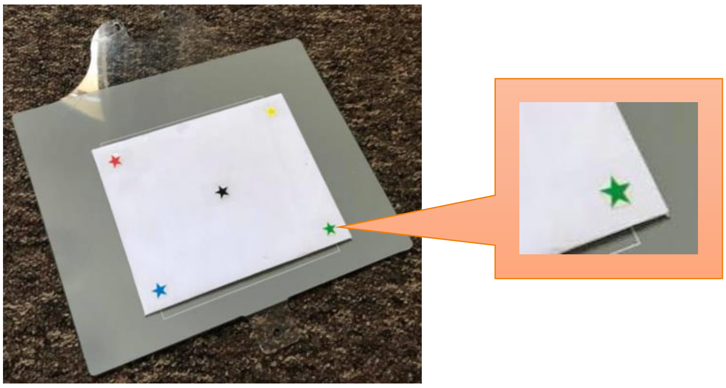


Figure 4.1: The modified calibration plate with coloured stars indicators.

4.1.2 Initial DEG Prototype

According to Peirce (2013), young children's attention span is very short; however, the exact duration was not given. Children tend to lose focus and interest when a game becomes too long. For this reason, the DEG developed for this study was designed to be short within ~ 2 – 7 minutes (cf. the focused attention span in Section 2.5), varying with the child's characteristics. This short duration was to avoid fatigue in the participants, though it was still grounded in the Shi and Shih's (2015) GBL design model. The first phase of the DEG design had a few flaws which were identified and improved for the following pilot study. The flaws included the navigation support, interaction and feedback.

Navigation support. As discussed earlier in Section 3.2.4 the game had two tasks known as Game *M* and Game *S*, which are both directed to different pages of the game. On the main page, players were asked to choose their preferred character of interest. Unfortunately, the instruction sound was ignored by the participant, and with the given limited reading ability, the child was unable to navigate and sought help from the researcher during the experiment. To improve the game, a speaker icon was added to every instruction text in the game as shown in Figure 4.2.

Figure 4.2: Speaker icons on the main page (above) and other pages (below).

Interaction. In the Game *S* sorting challenge, the aim was to arrange the given numbers into a correct order by dragging the number boxes on the upper side of the page into the target boxes in the train below. In this pre-test phase, participants had difficulty in placing the numbers onto the target boxes of the train which sometimes confused the children. This stickiness function of placing the number onto the target boxes was due to the small pixel range that was coded for the target box in Action Script. Therefore, the initial pixel range of the target box was widened from 15px to 25px to allow more target space for the participant to place the number boxes and consequently give more flexibility (Figure 4.3).

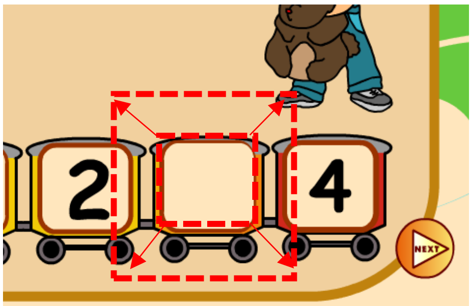


Figure 4.3: The target box range was widen to 25px from 15px

Feedback. Another issue identified with the DEG was the feedback interaction in Game *S*. Initially, the only feedback interaction given to the participant was a feedback voice responding "Well done" for every correct answer dragged into the target box in the train.

Unfortunately, this feedback voice was unnoticed and ignored by the participants. Instead, the young participant behaved as if they were waiting for a visual response on the screen after every correct answer. To overcome this problem and to notify the participant that they had made a correct answer a smiley star image was added together with the feedback voice. The smiley star image would appear above the targeted train compartment concurrently with the feedback voice "Well done" (Figure 4.4).

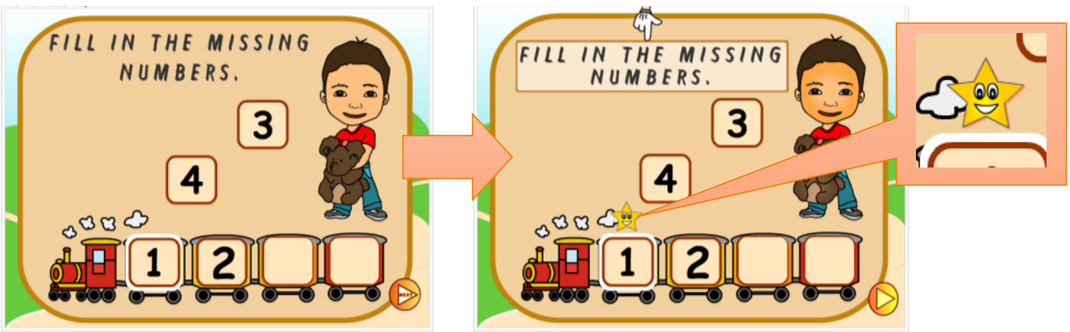


Figure 4.4: Smiley star images as an additional feedback to the sound feedback.

4.2 Pilot Study: School Setting

The pilot study involved 31 children (18 females and 13 male) from a local school in Leicestershire and provided more comprehensive input to the main study (Chapter 5). The study was conducted in a dedicated reading room of the participating school, being separate from the classroom to minimise distractions such as concurrent classroom activities.

The initial study protocol (Figure 4.5) started by allowing the child to sit in front of the eye-tracker. Next the external camera was turned on and the calibration process was performed. The usage of an external camera was originally to capture a better quality video of the child. The calibration process proved to be challenging for some cases as it is in general not easy for a child to listen to instructions. Repeated calibrations had to be done when not enough calibration data was collected. The frequency this process needed to be repeated depended on individual participant's eye and bodily features as well as on cooperativeness (Section 4.1.1). Once enough data was collected the child proceed by answering a paper-based pre-test while being in front of the MDS (Mobile Device Stand). The child then played either the DEG or cardboard game until the game ends. Finally, the child answered another paper-based post-test before ending the calibration process and was sent back to the classroom.

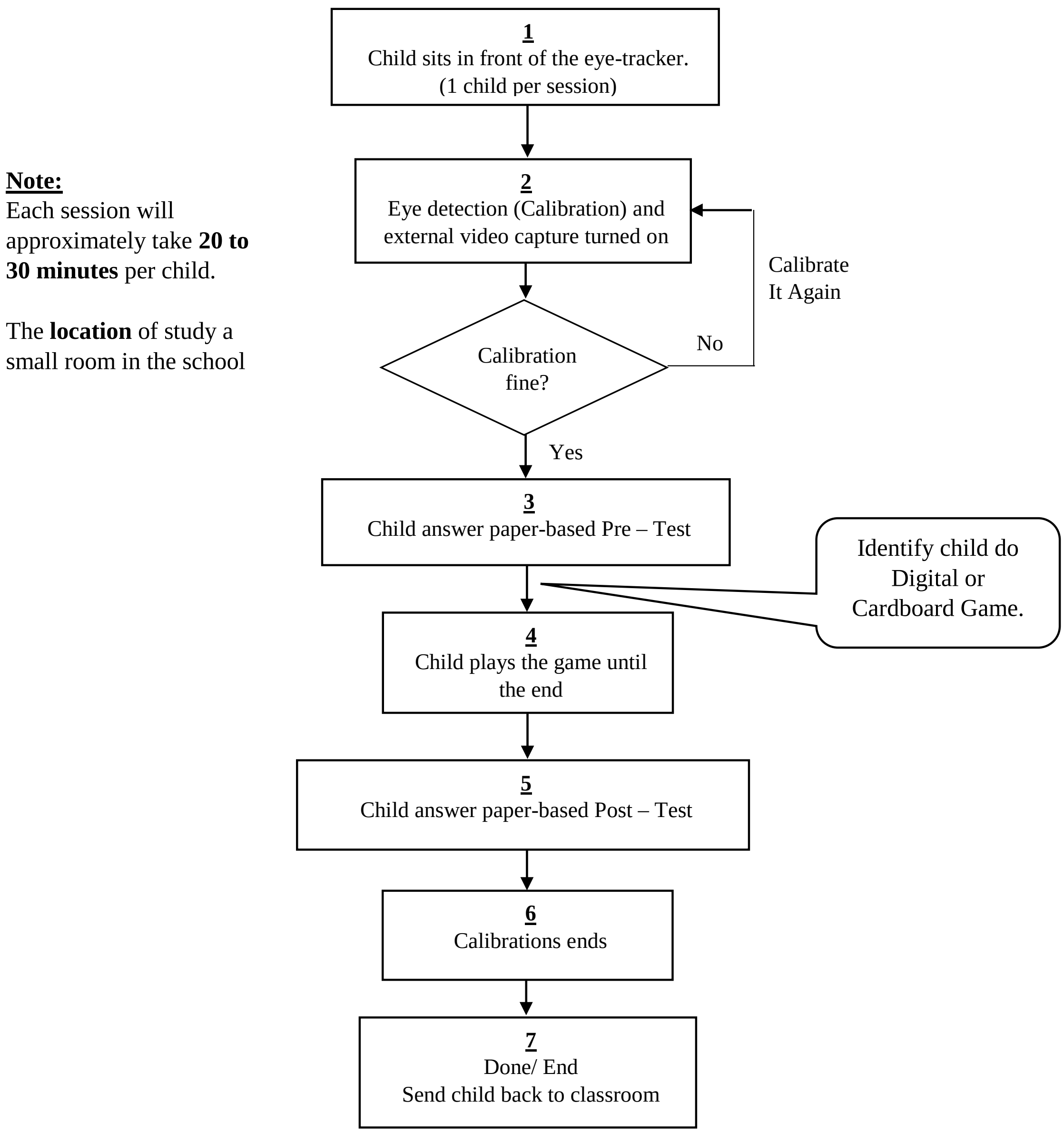


Figure 4.5: Initial Study Protocol (flow)

The experimenter was present in the room throughout the session, to observe the child's interaction behaviour and provided help when necessary. For the duration of the cardboard game session, the experimenter's presence was essential as she played a role in the gameplay by delivering feedback to the child.

The initial protocol was improved for the main study due the calibration phase and length, external camera usage (Section 4.2.2) and paper-based test area. Initially the calibration phase began before the paper-based pre-test and ended after the completion of the post-test. However, the length for the calibration process was to lengthy which soon out-range the calibration points because of too much distractions. Therefore, in the main study the calibration phase began after the paper-based pre-test and ends after the child's gameplay session. As for the paper-based test area that was initially in front of the MDS was located to a dedicated table beside the eye-tracking area to minimize distraction of the child while completing the test. An improved study protocol for the main study is discussed in Section 5.2.

4.2.1 **Paper-based Pre-Test and Post-Test Evaluation**

As discussed earlier in Section 3.4 the paper-based test question had two parts: matching and sorting. Initially, the questions were arranged and squeezed into a single page layout to minimise paper usage. The above layout area had the matching task questions while the lower part had the sorting task questions. However, this layout confused the young children who could not focus on one question at a time. Therefore the layout of the questions was rearranged, displaying every question horizontally and distributing the question over two pages with appropriate spacing (Figure 4.6). The initial number of sorting task questions was also added and improved from four to five questions while decreasing the maximum score from 12 to 7. The decreased score was due to reducing the number of empty boxes per question to allow for 5-year-old children's capability. To assist children to be more focused on each question at a time, each question was revealed one at a time vertically downwards (arrow in Figure 4.6) while covering other questions with a piece of paper.

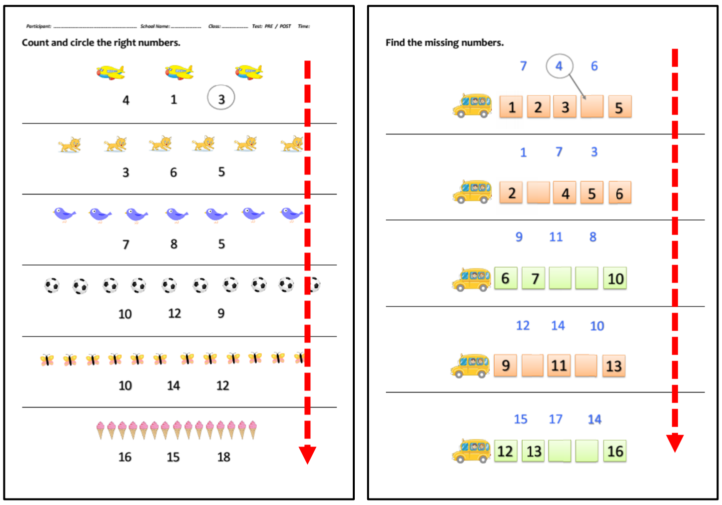


Figure 4.6: Revised paper-based test layout and direction of revealing the questions.

4.2.2 Improved Eye Tracker Set-up and Evaluation

The improved and modified calibration plate was tested in the pilot study and proved effective in attracting children's attention to the calibration points. This improvement slightly eased the calibration process. However, occasionally, due to a child's cooperation and eye feature, the repeated process could not be avoided.

Synchronising video capture. Another problem identified in the pilot study was with eye-tracker set-up. Initially, the video capture process was performed through an external camera situated next to the MDS (Figure 4.7: left). Unfortunately, the eye-tracker software (Tobii Studio) did not support external video files to be imported into the software. This non-support feature created difficulty in synchronising the video capture with the eye tracking recording. The high-resolution video capture was necessary to provide more explicit supplementary data such as children's facial expressions. As an alternative to this problem, the external camera was replaced with an embedded laptop camera which allowed video files and eye-tracker data to be recorded simultaneously into Tobii Studio (Figure 4.7: right), thereby avoiding data inconsistency. Nonetheless, this benefit of synchronicity is gained at the expense of the flexibility of the external camera's angles.

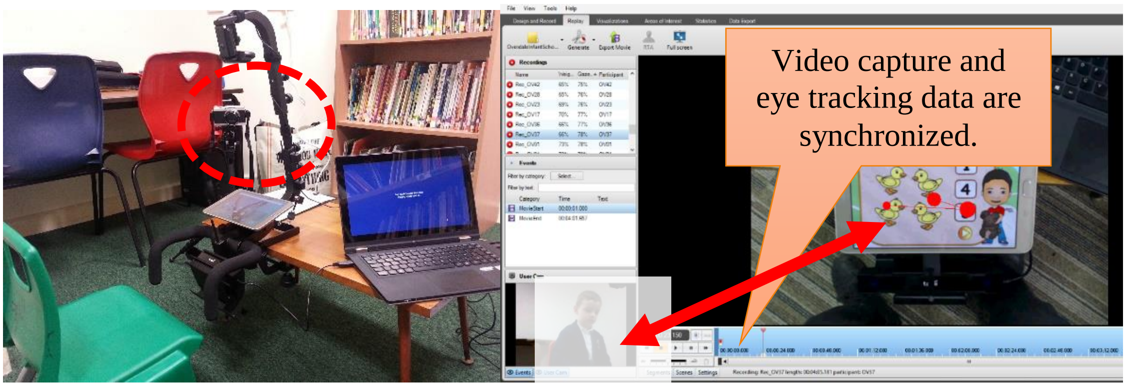


Figure 4.7: The initial external camera was replaced to enable data synchronicity.

4.2.3 Improved DEG Prototype

The improved DEG version was tested in the pilot study. However, the navigation support such as the speaker icon which activates the narrator (i.e. vocal sound that reads the instruction) continued to be overlooked by the children, creating further confusion and misleading. Therefore, the speaker icon was removed and replaced with an animated pointing hand directing to where the child should touch to activate the navigation narrator. In addition to that, the arrows on the signboard of the main page were also animated to inform and direct participants where to start in the beginning of the game (Figure 4.8).

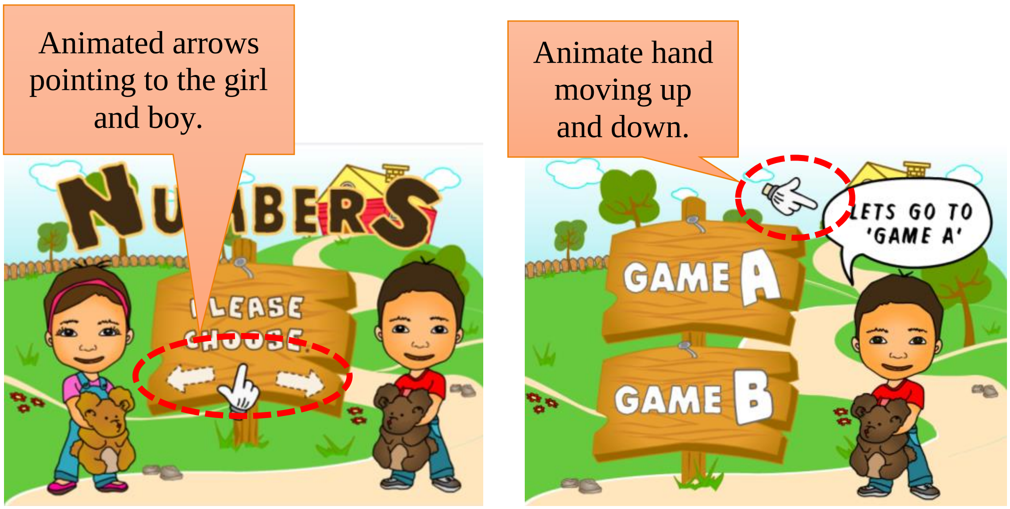


Figure 4.8: Animated pointing hands and arrows replaced confusing speaker icons.

Demo Page. The pilot study identified an additional problem in Game *S*. A number of participants were clueless when they reached Game *S* and had to seek assistance from the researcher. Occasionally, participants would touch the screen and waited for something to appear. Therefore, a demo page on how to play Game *S* was created and added before entering Game *S*. The flow of game pages from the selection page until the first level page including the demo page is demonstrated in Figure 4.9.

Figure 4.9: Placement of demo page between selection page and level 1 page.

4.2.4 Educational Cardboard Game Evaluation

Apart from the DEG design, the cardboard was also tested. For the cardboard game, instructions were manually given by the researcher and had a standard script of instructions similar to the DEG narrator and feedback sounds that was prepared in advance. Since the cardboard game experiment was played on a slant recording area that had limited space, situated under the camera device and above the eye tracker while being supported by the MDS frame (Figure 4.10), adjustment were made so that the eye-tracker could record data smoothly.

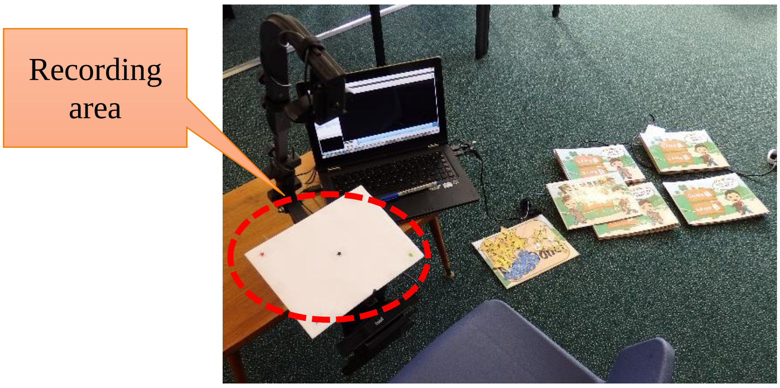


Figure 4.10: The cardboard game recording area situated under the camera device and above the eye tracker with the support of the MDS frame.

For instance, the number boxes had to be attached with Blu Tack to allow the boxes to stay and stick on the cardboard while the children played the game. Without the Blu Tack, the cardboard boxes would fall off from the recording area and distract the child's focus. If the child gets distracted and looks outside the recording area, no eye-tracking data can be recorded, contributing to missing data.

4.3 Chapter Summary

In summary, both the preliminary -test and pilot study provided comprehensive input for the main study. The initial calibration process that was too lengthy was improved for the main study protocol. The key issue with the eye-tracking set-up, particularly the calibration plate, was improved by adding recognisable symbols (i.e. colourful stars) for young children for both DEG-based and cardboard-game-based experiments. The room conditions for the eye-tracking experiment were also checked in advance, so no ceiling light reflected into the sensors. As for the paper-based test, the entire structure was modified to avoid confusion. Multiple game design flaws on both DEG and cardboard game were resolved for the next phase of the research project.

Chapter 5: Main Study (Fundamental, Method 1 and Method 2)

This chapter outlines the Main Study of this research project. Methodological improvements derived from the findings of the pilot study (Section 4.2) were implemented for the empirical work of the Main Study. Different schools and participants were recruited for this study. First, the basic statistical analyses of the relations between Learning Effect, which was the key variable of the study, and other factors, including Game Experience, were reported (Section 6.1); for these analyses no eye-tracking data were involved. Further analyses involved two different eye-tracking analysis methods (Figure 5.1). Method 1 used data from both DEG and cardboard game group and focused on how attention (fixation duration) on relevant and irrelevant UI objects in the games was related to the learning effect. Method 2 also involved the same datasets but used gaze sequences. Each of the two methods addresses different research questions (Section 1.2) with the ultimate aim to enrich the applied body of knowledge in deploying eye-tracking methods with young children.

Chapter 5
Main Study

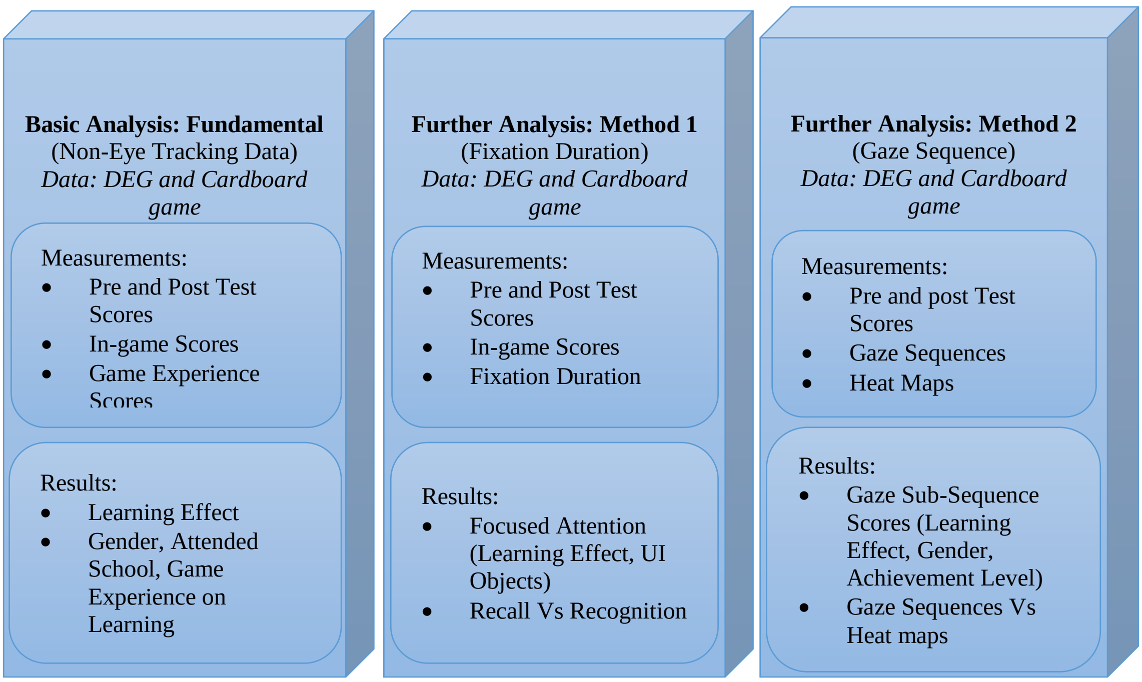


Figure 5.1: Basic (Fundamental) and Further Analysis (Method 1 and Method 2).

5.1 Participants

The Main Study was conducted in two local infant schools in Leicestershire for seven weeks. The experimental sessions were carried out on an individual basis, and the schedule of each session was bound by the 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.

5.2 Study Protocol

The improved study protocol (Figure 5.2) for the main study began by answering a pre-test on a table near the eye-tracking setup. Next, the child proceeds to the eye-tracking set-up area, identified either playing the DEG or cardboard game in front of the MDS (Mobile Device Stand). The eye-tracking session then commenced by performing the calibration process which also proved challenging as faced during the pilot study. Repeated calibration was done to collect enough data. At the same time the built-in camera of the laptop capturing data into the Tobii Studio software was turned on and recording began as the child plays the game. Similar to the initial study protocol, throughout the session, the experimenter was present in the room to observe the child's interaction behaviour and provided help when necessary. As duration of the cardboard game, the experimenter's presence was essential as she played a role in the gameplay by delivering feedback to the child (Section 4.1.1). After the child completed the gameplay the calibration recording ends. Ending the experiment, the child answers another paper-based post-test on the table before being sent back to their classroom.

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 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-years-old (Ruff et al., 1998).

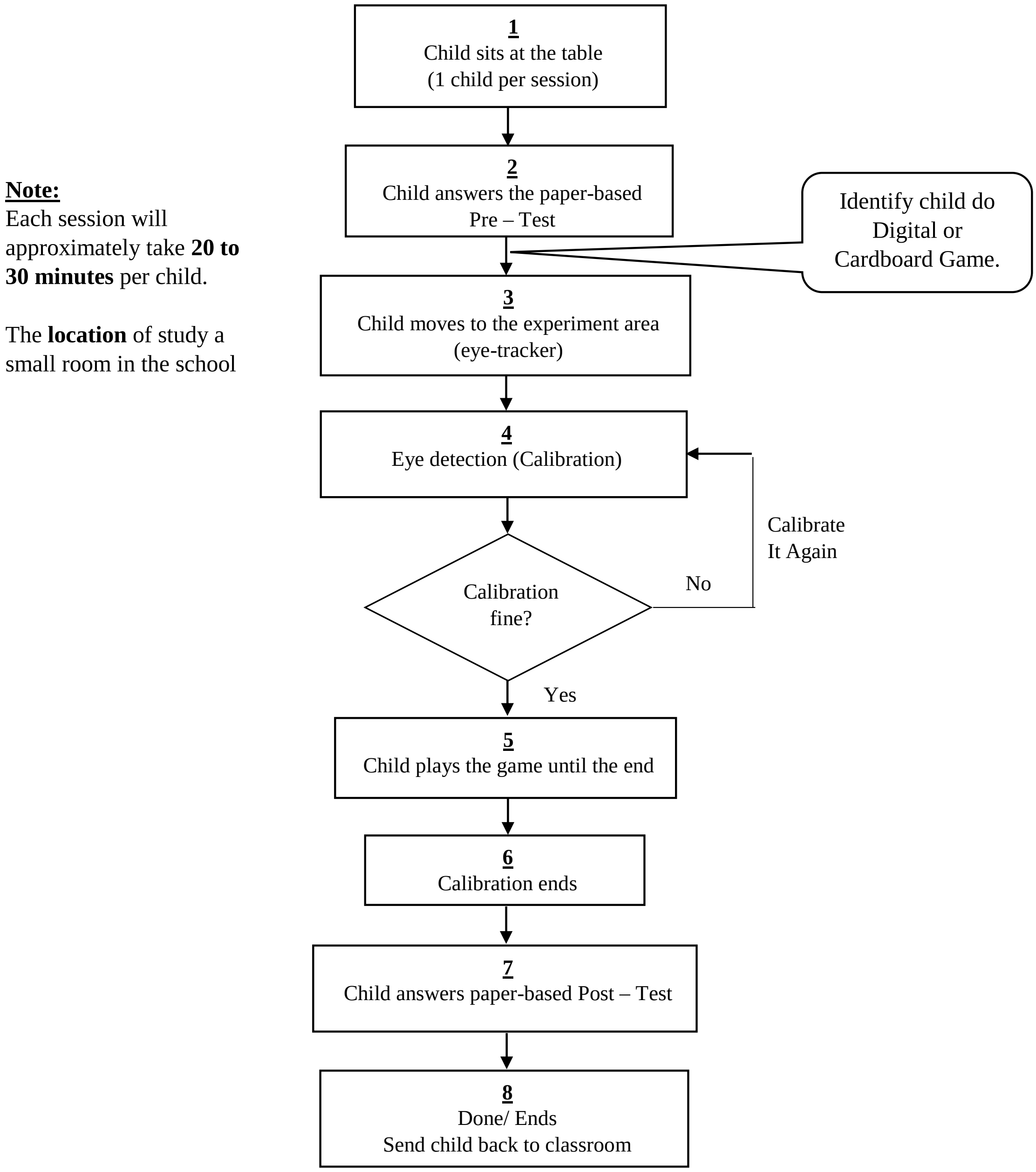


Figure 5.2: Improved study protocol for the main study

Figure 5.3 shows the research model of the Main Study. The basic analysis (Fundamental) and two further analyses (Method 1 and Method 2) mainly studied the learning effect (dependent variable) of both game learning media (independent variable), the DEG and Cardboard game, while other factors (covariates) including children's attributes (i.e. Gender, Prior Knowledge, School, and Cognitive Style) were also considered. The effect of age was not studied since all children in this study were of the same age. The Prior Knowledge could not be measured independently apart from the pre-test score, because of the lack of formal assessment at such a young age (reception or foundation classroom) in the UK system. Additionally, the relation between Game Experience and Learning Effect was studied too. The fundamental analysis that does not involve eye-tracking data addresses RQ1 and RQ2. Method 1 applies the fixation duration data and addresses RQ3, RQ4 and RQ5 while Method 2 applies gaze sequence data and addresses RQ6. The related hypotheses for each research question are discussed in Section 5.3, Section 5.4 and Section 6.1.

In order to avoid confusion with the terms used in reporting the results in this chapter, definitions of the terms are given in Table 5.1.

Table 5.1: Explanation of the terms used in this chapter.

Terms		**Description**
Learning Effect	:	Is referred to the pre and post-test score differences.
Performance Level	:	Are the categories of the "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 undertaking the experiment. For example, in this study "Achievement Level" is one of the prior knowledge components.
Focused Attention	:	Derived from the fixation duration for relevant and irrelevant objects; one aspect of Cognitive Styles.
Interaction Strategy	:	Is referred to the approach to playing the game as derived from gaze sequences; one aspect of Cognitive Styles.

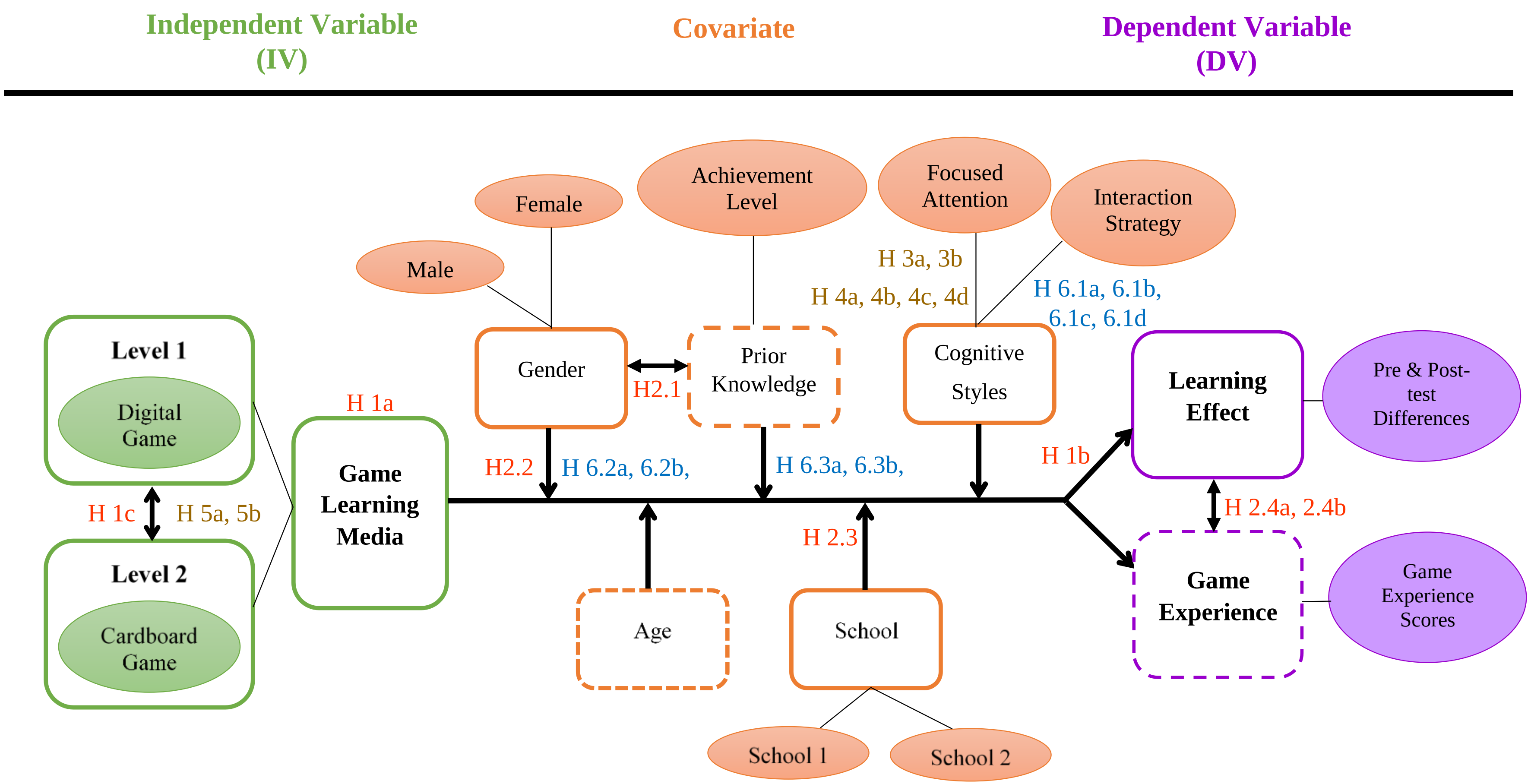


Figure 5.3: The research model and hypotheses with their related variables. Red labels are hypotheses related to the non-eye tracking analysis on Learning Effect. Brown labels are hypotheses related to Method 1 and Blue labels are hypotheses related to Method 2.

5.3 Fundamental: Analysis of Learning Effect and Other Factors

In this section, the relations between Learning Effect and other factors, including Gender, Achievement Level, and Game Experience, were analysed. Learning Effect and Game Experience were measured as the results of playing the DEG and Cardboard game whereas Gender and Achievement Level were children's attributes.

Learning Effect: Many studies have reported positive outcomes from the use of DEGs for improving learning in young children (Aladé et al., 2016; Burnett, 2010; Neumann, 2018; Peirce & Centre, 2013; Zaranis et al., 2013). However, some studies reported negative effects of using DEGs for children with regards to transfer of knowledge (Moser et al., 2015) and memory constraints (Barr, 2013). Therefore, considering the inconsistent findings of these previous studies, the following hypotheses involving learning effect and game learning media were formulated to answer research question 1 (**RQ1**):

> ***Hypothesis 1a****: There is a statistically significant correlation between (i) the DEG in-game scores and pre-test; (ii) the DEG in-game and post-test scores; (iii) the Cardboard game in-game and pre-test scores; (iv) the Cardboard game in-game and post-test scores.*
>
> ***Hypothesis 1b****: There is a statistically significant learning effect of (i) the DEG and (ii) the Cardboard game.*
>
> ***Hypothesis 1c:*** *There is a statistically significant difference in the learning effect between the DEG and Cardboard game.*

Materials: Evaluation material in the study consisted of 3 parts: paper-based pre-test evaluation, paper-based post-test evaluation and in-game evaluation. The improved paper-based test had the maximum score of 5 for Game *M* and 7 for Game *S*, which was reduced from 12 of the earlier version (cf. Section 3.4), having adjusted to the new layout structure (Section 4.2.1). The pre and post-test evaluation were calculated by summing the correct answers given, 1 score for each correct answer and 0 for each wrong answer. As for the in-game evaluation, the scores were counted by analysing the video and eye-tracking recordings in the Tobii Studio. A correct answer was given 1 score and 0 for a wrong answer made by a child. The maximum in-game score for Game *M* was 4, and for Game *S* was 12.

Learning Effect based on Gender, Attended School and Game Experience: The limited number of research studies related to gender and the learning effect on educational technologies with young children (Oliemat, Ihmeideh, & Al-Khawaldeh, 2018) was the motivation to study gender as a learning effect factor. Gender was examined further to compare its role in the learning effect of both game learning media. The variable School was also look into account since both schools involved in this study were from different achievement and deprivation level profile (Sabol, Bohlmann, & Downer, 2018). Another variable that was also observed was the children's game experience. Game experience may contribute valuable information with regards to the quality and motivation of the game design (Drachen et al., 2010). Therefore, the following hypotheses were formed to answer research question 2 (**RQ2**):

> ***Hypothesis 2.1****: There is a statistically significant correlation between Gender and Achievement Level for (i) the DEG participants; (ii) the Cardboard game participants.*
>
> ***Hypothesis 2.2****: There is a statistically significant difference in the learning effect between Genders of (i) the DEG participants; (ii) the Cardboard game participants.*
>
> ***Hypothesis 2.3****: There is a statistically significant difference in the learning effect between both schools of (i) the DEG group; (ii) the Cardboard group.*
>
> ***Hypothesis 2.4a****: For the DEG, there is a statistically significant correlation between the learning effect and Game Experience for (i) Game M; (ii) Game S.*
>
> ***Hypothesis 2.4b****: For the Cardboard game, there is a statistically significant correlation between the learning effect and Game Experience for (i) Game M; (ii) Game S.*

Materials: The evaluation material used in this study consisted of the paper-based pre-test and post-test evaluation, which was actually the same test administered before and after the gameplay. The maximum score of the test for Game *M* was 5 and 7 points for Game *S*. The scores were calculated by summing the correct answers the child made, 1 score for each correct answer and 0 for each wrong answer.

Game Experience: The game experience context in this study refers to the children's facial expression during the gameplay session. The facial expressions were retrieved from observing the video recording captured by the eye-tracking software. This observational approach was used in this study because interviewing the child may take up some time which was not possible due to the school's strict regulation in taking out a child from their classroom individually (25 to 30 minutes). Stretching beyond the 30 minutes may cause the child to miss out a certain lesson or task in their classrooms. Whereas applying questionnaire are not suitable for 5-year-old children due to their limited reading ability (McIntyre et al., 2006). Therefore, the observational approach was the ideal method for this study.

Before the facial expression could be coded and assessed according to the coding scheme, the video recording was divided into four sections; Selection scene, Game *M* scene, Game *S* scene and the End scene, all covering the beginning until the end of the gameplay session. The coding scheme with three emotional categories as seen in Figure 5.4 was used -Negative (e.g., Sad, Bored), Neutral, and Positive (Happy, Excited) - and each category was associated with a numerical value of 0, 1 or 2, respectively (Table 6.4).

Two raters coded the children facial expressions with the coding scheme. The first rater was the researcher of this study. The second rater was a male that had an IT engineering background and was not familiar with the experiment. Coding these facial expressions was a new task for both raters; therefore, both raters read related articles on coding facial expression emotions (i.e. LoBue, Baker & Thrasher, (2017) and Widen & A Russell, (2010)) and practiced with some existing data from the literature before undertaking the real coding task. Initially, the Cohen's Kappa was 0.49, which was not satisfactory. For the discrepant ratings, the two coders discussed the differences until a consensus was reached. With the agreed values, the improved Cohen's Kappa was 0.71. The values were then used for the statistical analysis discussed in Section 6.1.4.

Happy/Excited Neutral Sad/Bored

Figure 5.4: Examples of children facial expression adapted from LoBue et al. (2017)

5.4 Method 1: Validating Learning Effect based on Focused Attention

Data Analysis Method 1 studied how Focused Attention (fixation duration) was related to the learning effect and the children's prior knowledge (Achievement Level). Further analysis related to the learning effect such as children attributes was also performed and the effect of the difficulty level of individual tasks was observed. All 94 participants' data from both the DEG and Cardboard game were used in the analysis.

Focused Attention on Learning Effect: Focused attention is related to cognitive and behavioural aspects (Ruff et al., 1998). With recent advances in the eye-tracking technology (Holmqvist et al., 2011), it stimulates more opportunities for different research studies. The use of the eye-tracking technology enables research on focused attention and learning effect of a particular intervention by utilising the fixation duration measurement. With this, the following hypotheses involving fixation duration on each of the four pages of both Game M and Game S, corresponding to the four levels of difficulty in ascending order, on both game learning media were derived to answer the research question 3 (**RQ3**).

> ***Hypothesis 3a****: For the DEG, focused attention is a statistically significant predictor of the learning effect for each of the four difficulty levels of (i) Game M; (ii) Game S*

> ***Hypothesis 3b****: For the Cardboard game, focused attention is a statistically significant predictor of the learning effect for each of the four levels of (i) Game M and (ii) Game S.*

Focused Attention on Relevant and Irrelevant objects: Theoretically, gaze measures are a good proxy of attention paid to objects of interest (Steichen et al., 2014). However, maintaining attention to assigned tasks or objects is difficult for young children, who are easily distracted (Richter & Courage, 2017). Therefore, from these statements the following hypotheses on focused attention given to relevant and irrelevant object between high and low achievers were derived to answer the research question 4 (**RQ4**).

> ***Hypothesis 4a:*** *There is a statistically significant difference between high and low achievers in focused attention given to the relevant objects of the DEG Game M (i) level 1; (ii) level 2; (iii) level 3; (iv) level 4.*

> ***Hypothesis 4b****: There is a statistically significant difference between high and low achievers in focused attention given to the relevant objects of the DEG Game S (i) level 1; (ii) level 2; (iii) level 3; (iv) level 4.*

> ***Hypothesis 4c****: There is a statistically significant difference between high and low achievers in focused attention given to the relevant objects of the Cardboard Game M (i) level 1; (ii) level 2; (iii) level 3; (iv) level 4.*

> ***Hypothesis 4d****: There is a statistically significant difference between high and low achievers in focused attention given to the relevant objects of the Cardboard Game S (i) level 1; (ii) level 2; (iii) level 3; (iv) level 4.*

Fixation duration was used to study the attention given by the children to the game learning medium. Fixation duration data were obtained through the Tobii Studio software. The mean fixation duration of each AOI that was defined (Section 3.3.5) on all pages of Game *M* and Game *S* was exported to SPSS for analysis. The total fixation duration on each page of Game *M* and Game *S* was also exported for analysis.

Learning Task Difficulty: Recall and recognition are both memory accessing process to the long-term memory (Huitt, 2003). However, recall is a more complex process that involves searching and decision making while recognition involves comparing information derived from an encountered experience of an object or event with the memory (Baddeley, 1997). Based on these theoretical assumptions, to validate whether recall tasks are more difficult than recognition tasks, the following hypotheses were derived to answer the research question 5 (**RQ5**). Game *M* represented the recognition-based game tasks while Game *S* represented the recall-based tasks for both the digital and cardboard game learning media:

> ***Hypothesis 5a****: There is a statistically significant difference in the in-game scores between the easy level tasks of Recognition and Recall for both the DEG and Cardboard game.*

> ***Hypothesis 5b****: There is a statistically significant difference in the in-game scores between the hard level tasks of Recognition and Recall for both the DEG and Cardboard game.*

5.5 Method 2: Validating Learning Effect based on Gaze Sequence Analysis

Data Analysis Method 2 addressed how gaze sequences (interaction strategy) were related to the Learning Effect, Gender and Achievement Level on both game learning media. Further analyses of gaze sequence results and visualisation data (heat maps) based on relevant and irrelevant objects were also observed and studied towards the end of the chapter. All 94 participants' data from both the DEG and Cardboard game were used in the analyses.

Gaze Sequence (Interaction Strategy) based on Gender, Achievement Level and Learning Effect: 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 in empirical studies (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 3.3.6. 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).

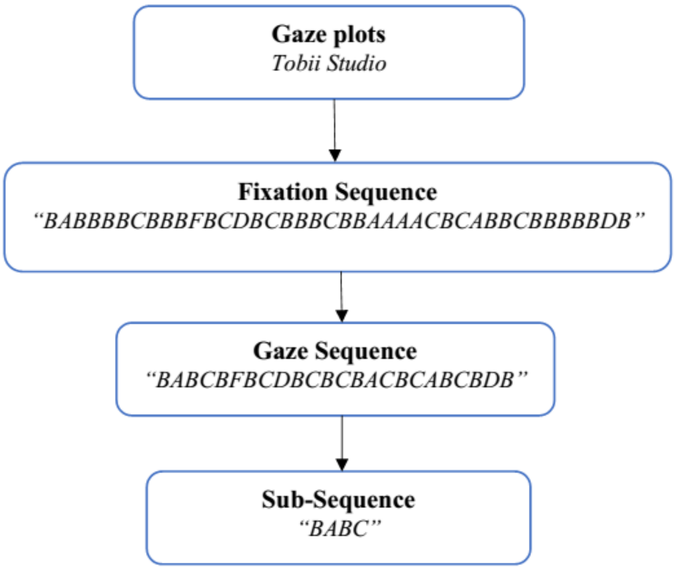


Figure 5.5: Converting Sequence process

Figure 5.5 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 (Section 6.3.2), Gender (Section 6.3.3) and Achievement Level (Section 6.3.4) 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 (Game *M* and Game *S*) by sorting, colour-coding, clustering and calculating the gaze sub-sequence scores (Figure 5.8).

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). The scoring scheme used in this study is 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 ideal one (i.e., the shortest and accurate gaze sub-sequence for answering the game task) and applied a scoring scheme based on the AOI position in the sub-sequence. In this study, the gaze sub-sequences (e.g. *BABC*) which were converted from the gaze sequences (e.g. *BABCBFBCDBCBCBACBCABCBDB*) as shown in Figure 5.5 were scored and described subsequently.

In Game *M*, the ideal gaze sub-sequence involved only the first 3 of the four AOIs fixated for a gaze sub-sequence, this is because a single task can efficiently and effectively be completed by visiting 3 AOIs. Figure 5.6 shows how the scores were given to each participant in level 1 of Game *M*. For example, the participant OV42's gaze sub-sequence was "ABA" and the ideal one was "ABD". Therefore, only position 1 and position 2 of this gaze sub-sequence were given points while position 3 was not given any point because "A" did not match with the 3rd position of the ideal sub-sequence.

No	Partcode id	Sub-sequence	Position 1 Max: 1	Position 2 Max: 1	Position 3 Max:1	Total Score Max: 3
		Ideal sequence:	A	B	D	
1	OV42	ABA	1	1	0	2
2	UP54	ABC	1	1	0	2
3	UP51	ACD	1	0	1	2
4	OV31	BAB	0.5	0.5	0	1
5	UP56	BAB	0.5	0.5	0	1
6	OV14	BEC	0	0.5	0	0.5
7	OV11	CAB	0.5	0.5	0	1
8	OV12	DFD	0	0	0.5	0.5
9	OV46	EAB	0.5	0.5	0	1
10	OV18	FHA	0.5	0	0	0.5

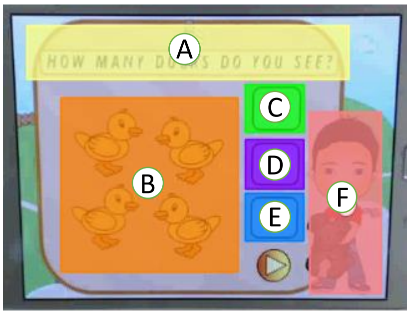


Figure 5.6: Scoring the gaze sub-sequence for Page 1 in Game *M* (left). Game *M* design (right).

Another example is the participant OV31's gaze sub-sequence for the same task. With the gaze subsequence "BAB", position 1 and position 2 were given 0.5 point each; although "B" and "A" were not in the ideal position they were still among the first three positions of the ideal gaze sub-sequence. While position 3 of OV31's sub-sequence was given 0, because no repetition was accepted in calculating the scores.

As for Game S, given its different design the game scores were calculated slightly different but still adapted the position-based weighted model approach (Sutcliff & Namoun, 2012; Eraslan et al., 2016). In Game *S*, the ideal gaze sub-sequence involved fixating the first 4 AOIs to complete a single task (i.e. selecting and dragging one of the number boxes into the train). Figure 5.7 shows how the scores were given to each participant in level (or page) 3 of Game *S*. For example, the participant OV14's gaze sub-sequence was "DBFA" and the ideal one for Game *S* was "AB(X)B". The X in the ideal 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.

		Ideal sequence:	A	B	D or E or F	B	
No	Partcode id	Sub-sequence	Position 1 Max: 1	Position 2 Max: 1	Position 3 Max:1	Position 4 Max:1	Total Score Max: 4
40	OV13	CEFD	0	0	1	0	1
41	OV45	DABF	0	0	0	0	0
42	OV14	DBFA	0	1	1	0	2
43	OV11	DBFB	0	1	1	1	3
44	OV05	DBFB	0	1	1	1	3
45	OV02	EBEB	0	1	1	1	3
46	UP53	EBFA	0	1	1	0	2
47	OV21	EFBE	0	0	0	0	0
48	OV24	FBCD	0	1	0	0	1
49	UP52	FBDE	0	1	1	0	2

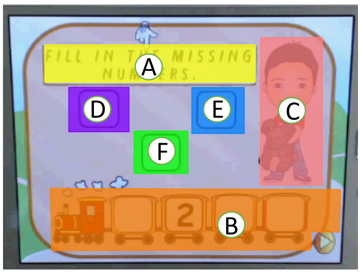


Figure 5.7: Scoring the gaze sub-sequence in Game *S* (left). Game *S* design (right).

To simplify the matter, the ideal gaze sub-sequence does not take the correctness of the move into account, focusing on the relevance of the related AOI. Therefore, for OV14 only position 2 and position 3 were given points while position 1 and position 4 were not given any points. Another example is the participant OV02's gaze sub-sequence, which was "EBEB", and the ideal one was still "AB(X)B". 1 point was given to position 2, 3 and 4 each while no point was given for position 1, giving a total of 3. Nonetheless, in Game *S* repetitions were accepted but no scores were given for 'not in the ideal position' within the first 4 AOIs fixated, in contrast to the scoring scheme used in Game *M*.

Therefore, from these studies and analysis methods, the following hypotheses involving gaze sub-sequences were proposed to answer research question 6 (**RQ6**):

> ***Hypothesis 6.1a****: For the DEG, there is a statistically significant difference in the Gaze Sub-Sequence Score between the high and low performers in terms of Learning Effect for (i) Game M; (ii) Game S.*

> ***Hypothesis 6.1b****: For the Cardboard game, there is a statistically significant difference in the Gaze Sub-Sequence Score between the high and low performers in terms of Learning Effect for (i) Game M; (ii) Game S.*

> ***Hypothesis 6.1c****: For the DEG, there is a statistically significant correlation between the learning effect and Gaze Sub-Sequence Score for (i) Game M; (ii) Game S.*

***Hypothesis 6.1d**: For the Cardboard game, there is a statistically significant correlation between the learning effect and Gaze Sub-Sequence Score for (i) Game M; (ii) Game S.*

***Hypothesis 6.2a**: For the DEG, there is a statistically significant difference in the Gaze Sub-Sequence Score between Genders for (i) Game M; (ii) Game S.*

***Hypothesis 6.2b**: For the Cardboard game, there is a statistically significant difference in the Gaze Sub-Sequence Score between Genders for (i) Game M; (ii) Game S.*

***Hypothesis 6.3a**: For the DEG, there is a statistically significant difference in the Gaze Sub-Sequence Score between the two groups with high and low Achievement Level (Prior Knowledge) for (i) Game M; (ii) Game S.*

***Hypothesis 6.3b**: For the Cardboard game, there is a statistically significant difference in the Gaze Sub-Sequence Score between the two groups with high and low Achievement Level (Prior Knowledge) for (i) Game M; (ii) Game S.*

Gaze Sub-Sequence Analysis: The interaction strategy used by the children at the beginning of each level of the game was analysed by using the gaze sub-sequences and the gaze sub-sequence percentage (cf. Relevant and Irrelevant AOIs). The gaze sub-sequences focus on where individual children fixated in the beginning of the game across the 4 levels while the sub-sequence percentage (%) focus on the proportion of children fixated Relevant and Irrelevant AOI objects.

Clustering the Sub-sequences and calculating the sub-sequence percentage: In studying the potential interaction strategy applied by the child when entering each level of the game task, 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 Figure 5.8). Boxes coded in red represent Irrelevant AOIs objects while other colours such as yellow, orange, purple, blue and green are Relevant AOI objects. How the UI objects were defined as Relevant and Irrelevant objects was discussed earlier in Section 3.3.5. The numbers of Relevant and Irrelevant 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 relevant AOI than an irrelevant one.

For example (Figure 5.8), the Male-Low cluster playing the Cardboard game, 5 out of the 7 children fixated first on Relevant AOIs and 2 Irrelevant 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:

$$\boldsymbol{Relevant}\ AOIs\colon \frac{5}{7}\ divided\ by\ \frac{5}{6} = 0.86 = \frac{0.86}{(0.86 + 1.71)} = 33.3\%$$

$$\boldsymbol{Irrelevant}\ AOIs\colon \frac{2}{7}\ divided\ by\ \frac{1}{6} = 1.71 = \frac{1.71}{(0.86 + 1.71)} = 66.7\%$$

Similarly, the same approach was applied to the Male -High, Female-Low and Female-High clusters and the results are discussed in the sub-sequence percentage (%) section (Section 6.3.5). The weighted averages are converted into percentages, allowing easier comparisons across the clusters

5.6 Chapter Summary

The Main Study was implemented in two local infant schools in Leicestershire, involving altogether 94 5-year-old children. The analysis methods consisted of three types: Fundamental, Method 1 and Method 2. The Fundamental analysis method employed basic statistical analyses that did not use any eye-tracking data and addressed primarily. RQ1 and RQ2. Both Method 1 and Method 2 involved eye-tracking data but differed in terms of eye-tracking metrics and analytical techniques. Method 1 focused on fixation duration analyses to address the research questions RQ3, RQ4 and RQ5 pertaining to focused attention and other variables. Method 2 focused on gaze sequences analyses to address RQ6. The related results and discussions of the analyses are presented in Chapter 6.

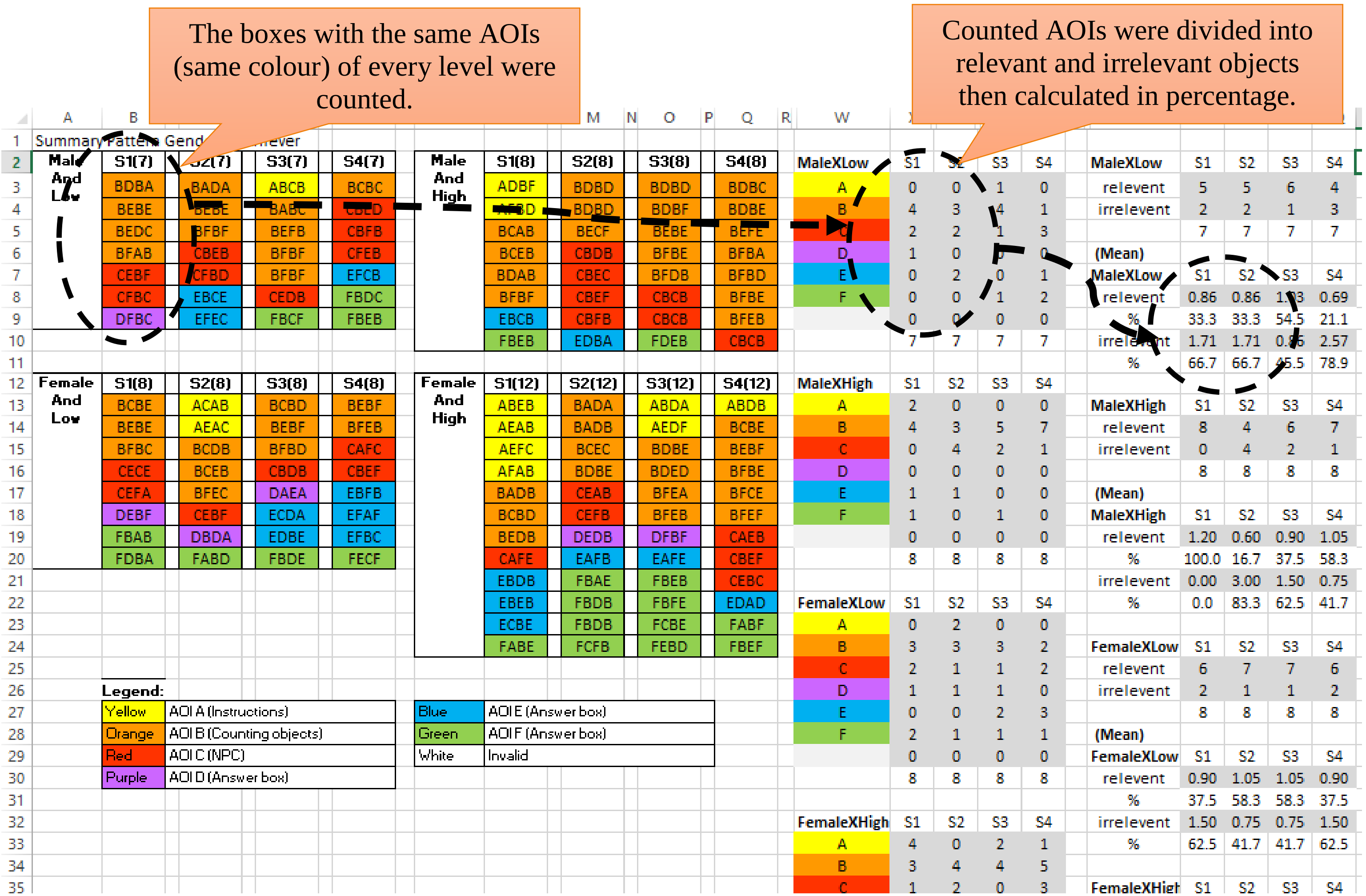


Figure 5.8: Calculating the Sub-sequence Percentage from Game *S* of the Cardboard Game. The conversion of the weighted averages was also to help compare the results of sub-sequence with those of the heat map visualisation analysed by the Tobii Studio software (Section 6.3.5)

Chapter 6: Results and Discussion

This chapter presents and discusses the results for each analysis method (Fundamental, Method 1 and Method 2) discussed in Chapter 5.

6.1 Fundamental

6.1.1 Preliminary Analysis Approaches

For each of the subsequent statistical analyses, a preliminary analysis was performed to make sure that no missing data existed, 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.

Outlier: By using the boxplots, the variables (pre-test, post-test and in-game variables) were examined and observed for outliers. Cases with extreme outliers (values more than three IQRs (Interquartile Range) 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. The rationale of transforming the score rather than deleting such legitimate cases was to retain the nature of the outlier without them biasing the results

Normality: A Shapiro-Wilk test was used to study the data distribution to identify which statistical method to apply. Based on the data distribution of the variables, parametric (normal distribution, $p>.05$) or non-parametric (non-normal distribution, $p<.05$) statistical methods were applied accordingly.

6.1.2 Learning Measurement Consistency

In this section the relationships among the measures (pre-test, post-test and in-game scores) taken in the Main Study are examined. This examination is to make sure the participating children applied their numeracy knowledge consistently across the experimental conditions. The maximum score for the pre-/ post-test (NB: pre-test and post-test were the same paper-based tests administered before and after playing the respective game) for Game *M* was 5 and Game *S* was 12, while the maximum in-game score for Game *M* was 4 and Game *S* was 12.

Main Analysis: Non-parametric Spearman correlation tests were performed to assess the relationships between the Pre- and Post-tests and In-game performance for both the DEG and cardboard game. For the DEG there were highly statistically significant positive correlations ($N = 59$, $p<.01$) between the In-game scores and Pre-test scores (Game *M*: $r = 0.67$; Game *S*: $r = 0.71$) and between the In-game scores and Post-test scores (Game *M*: $r = 0.46$; Game *S*: $r = 0.69$). The same statistically significant positive correlations ($N = 35$, $p<.01$) were also found for the Cardboard game between the In-game scores and Pre-test scores (Game *M*: $r = 0.47$; Game *S*: $r = 0.64$) and between the In-game scores and Post-test scores (Game *M*: $r = 0.29$; Game *S*: $r = 0.63$). Based on these results, the knowledge on numeracy among the participating children was applied consistently across the evaluation conditions and thus ***Hypothesis 1a*** **(i) - (iv)** was accepted.

6.1.3 Learning Effect

This section evaluates the Learning Effect between both game learning media, DEG (***Hypothesis 1b (i)***) and Cardboard (***Hypothesis 1b (ii)***). The total scores for the pre-/post-test were calculated by summing up those of Game *M* and Game *S*. The maximum score for both pre and post-test was 17.

Main Analysis: Non-parametric Wilcoxon signed-ranks tests were used to study the Learning Effect of both the DEG and Cardboard game. Statistical results for the DEG ($Mdn_{pre}=7.00$, Range=4.00-11.00; $Mdn_{post}=8.00$, Range=4.00-11.00) indicated that no statistically significant improvement in learning was achieved ($Z=1.10$, $p=0.27$), thus rejecting ***Hypothesis 1b (i)***. However, for the Cardboard game ($Mdn_{pre}=9.00$, Range=5.00-11.00; $Mdn_{post}=10.00$, Range=6.00-12.00), results indicated that there was a statistically significant improvement in learning ($Z=1.90$, $p<.05$, $r=0.23$), thus accepting ***Hypothesis 1b (ii)***. 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 manipulation of 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 ($U = 948.5$, $Z = 0.67$, $p = .50$) between both game learning media ($Mdn_{media} = 0.00$, Range= -1.00-1.00). This result could be attributed to the difference in sample size between both game learning media, DEG: 59 and Cardboard Game: 35. Nevertheless, recruitment of participants was beyond control and subject to the school's (head teachers) and parents' consent. When opportunities arose, evaluation with the DEG was prioritized because it was the alternative educational medium of which the effectiveness was of particular research and practical interest. From the Mann Whitney result, ***Hypothesis 1c*** was rejected.

6.1.4 Gender, Attended School and Game Experience on the Learning Effect

This section first reports the results on the relationship between Gender and Achievement Level of both game learning media (***Hypothesis 2.1)***, and then the influence of Gender on Learning Effect for both DEG (***Hypothesis 2.2 (i))*** and Cardboard game (***Hypothesis 2.2 (ii)***). Further, the relationships between Attended School and Learning Effect of both game learning media are presented (***Hypothesis 2.3)***. Finally, the relationships between Game Experience and Learning Effect on both DEG (***Hypothesis 2.4a)*** and Cardboard game (***Hypothesis 2.4b)*** are reported. The schools involved in this study had different deprivation levels. Deprivation levels in this context is inversely proportional to the level of socioeconomic background of individual families. Children in School 1 came from the least deprivation level and had an achievement rate of 86% for level 4 and above in reading and maths test. School 2 came from the medium-worst deprivation level and had an achieving rate of 74% for level 4 and above in reading and math test. In other words, children attending School 1 came from better well-off families as compared to School 2.

Main Analysis: A Chi-square correlation test was performed between Gender and Achievement Level to assess their independence. The result showed that no statistically significant relation between Gender and Achievement Level was found for both the DEG ($\chi^2 = 0.02$, $p = .89$) and Cardboard game ($\chi^2 = 0.16$, $p = .69$). Therefore, ***Hypothesis 2.1(i)*** and ***Hypothesis 2.1 (ii)*** were rejected. The distribution of the two variables is shown in Table 6.1.

Table 6.1: Distribution of High and Low achievers between genders for both DEG and Cardboard participants.

	DEG			Cardboard		
Achiever / Gender	Girls	Boys	Total	Girls	Boys	Total
High Achiever	15	14	29	12	8	20
Low Achiever	15	15	30	8	7	15

Mann Whitney tests were used to evaluate the difference in the Learning Effect between Genders (Table 6.2) on both game learning media. The result showed that there was no statistically significant Gender difference for the DEG group (U= 383.5, Z= .81, p= .42, p>.05). As for the Cardboard game, the results of the Mann Whitney test (U = 94.5, Z = 1.89, p<.05, r=0.32) showed that there was a statistically significant gender difference in the Learning Effect. Therefore, ***Hypothesis 2.2 (i)*** in relation to the DEG group was rejected and ***Hypothesis 2.2 (ii)*** for the Cardboard group was accepted.

Table 6.2: Medians of pre and post-test scores on gender for both DEG and Cardboard Game.

	Gender	Pre-Test	Post-Test	Mann U
DEG	Boys (n=29)	7.00	7.00	U = 383.5, Z = .81, p = .42
	Girls (n=30)	7.50	9.00	
Cardboard	Boys (n=15)	9.00	9.00	U = 94.5, Z=1.89, p=.05*
	Girls (n=20)	9.50	11.00	

*p<.05

The Mann Whitney tests were also used to evaluate the difference in the learning effect between both schools attended by the children on both game learning media (Table 6.3). The results showed no statistically significant difference between School 1 and School 2 on the DEG (U =281, Z=0.02, p=.98). However, for the Cardboard game a statistically significant difference was found between both schools in the learning effect (U = 94.5, Z = 1.89, p<.05, r=0.32). These results could be attributed to the fact that more than half of the children for the Cardboard group were from School 2 attended by children from low-medium income family which may not have access to game learning media. Therefore, ***Hypothesis 2.3(i)*** with regard to the learning effect between both schools on the DEG was rejected while ***Hypothesis 2.3(ii)*** on the Cardboard game was accepted.

Table 6.3: Medians of pre- and post-test scores on school for both DEG and Cardboard Game.

	Schools	Pre-Test	Post-Test	Mann U
DEG	School 1 (n=47)	6.00	7.00	U = 281.0, Z = .02, p = .98
	School 2 (n=12)	10.50	11.00	
Cardboard	School 1 (n=15)	5.00	4.00	U = 94.5, Z=1.89, p=.05*
	School 2 (n=20)	10.50	11.50	

*p<.05

Spearman's rho correlations were computed to assess the relationships between Learning Effect and Game Experience (Table 6.4) for both the DEG and Cardboard game. For the DEG, the result showed that no statistically significant correlation was found for Game *M* (N = 59, r=0.18 p=.17) between Learning Effect and Game Experience. However, there was a statistically significant positive correlation for Game *S* (N = 59, r=0.29 p=.03) between the two variables. As for the Cardboard game, contrariwise, a statistically significant positive correlation for Game *M* (N = 35, r=0.37 p=.03) between Learning Effect and Game Experience was found. However, no statistically significant correlation was found for Game *S* (N = 35, r=0.07 p=.69).

This indicated that Game *M* task (recognition-based task) was preferably played through a physical environment while Game *S* task (recall-based task) was preferably played through a digital game-based learning environment possibly because of the game attractiveness and flexibility. Based on the result for the DEG, children seemed to enjoy game tasks that involved multiple interactions (Game *S*) such as choosing and dragging the number boxes as compared to a single interaction task (Game *M*) for instance tapping the answer directly without dragging or swiping on the tablet. For the Cardboard game, the different correlation between Game *M* and Game *S* showed that children were excited at the beginning of the game but tended to get exhausted towards the end of the game session. However, the groups were not split into two groups, allowing one group to start with Game *M* and the other with Game *S* to control the order effect. Having all children played Game *M* first and then Game *S* was to follow the game principle design to move from the easy to hard level. On top of that, asking the children to begin straightaway with the hard-level game tasks may demotivate the child and make them disengaged from the experiment.

Table 6.4: Game Experience scores coded by the raters. The complete list in Appendix A.9.

		Coding 1						Coding 2						Agreed Coding				
partid	**partcode**	**Selection**	**Game M**	**Game S**	**End**	**Total Score**		**Selection**	**Game M**	**Game S**	**End**	**Total Score**		**Selection**	**Game M**	**Game S**	**End**	**Total Score**
		2	2	2	2	8		2	2	2	2	8		2	2	2	2	8
1	OV01	2	2	2	2	8		1	1	1	1	4		2	2	2	2	8
2	OV02	1	1	1	1	4		1	1	1	1	4		1	1	1	1	4
3	OV03	1	0	1	1	3		1	0	1	1	3		1	0	1	1	3
4	OV04	2	2	2	2	8		2	2	2	2	8		2	2	2	2	8
5	OV05	1	1	1	1	4		1	1	1	1	4		1	1	1	1	4
6	OV06	1	2	2	2	7		1	1	1	2	5		1	2	2	2	7
7	OV07	2	2	2	2	8		1	1	1	2	5		2	2	2	2	8
8	OV08	2	2	2	1	7		2	2	2	0	6		2	2	2	1	7
9	OV09	1	2	1	2	6		1	2	1	2	6		1	2	1	2	6
10	OV10	1	2	2	2	7		1	2	2	2	7		1	2	2	2	7
11	OV11	2	2	1	2	7		2	2	1	2	7		2	2	1	2	7
12	OV12	1	2	1	2	6		1	2	1	2	6		1	2	1	2	6
13	OV13	2	1	1	1	5		2	1	1	1	5		2	1	1	1	5
14	OV14	1	1	1	1	4		1	1	1	1	4		1	1	1	1	4
15	OV15	0	1	1	0	2		1	0	1	0	2		0	1	1	0	2

Another possible explanation for the Cardboard game was that Game *S* children were more occupied with extraneous influences such as pulling and sticking (Velcro and Blu Tack) the cardboard number boxes in the correct places. Some were concerned whether the cardboard boxes would fall off or were not in the proper position. Therefore, ***Hypothesis 2.4a*** and ***Hypothesis 2.4b*** were both partially accepted.

6.2 Method 1

6.2.1 Preliminary Analysis Approaches

A preliminary analysis was performed to make sure that no missing data existed, 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.

Outlier: The variables (pre-test, post-test and in-game variables) were examined and observed for outliers. Extreme outliers (values more than three IQRs 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. For multiple regression analysis (Section 6.2.2), further screening metrics- Mahalanobis (df =4, x^2

=9.488) and Cook's (greater than 1) distance - were used to remove remaining outliers, reducing the total number of cases for some groups.

Normality: A Shapiro-Wilk test was used to study the data distribution to identify which statistical method to apply.

6.2.2 Focused Attention on Learning Effect

In this section, the results of the Learning Effect based on the Focused Attention duration on each page for both DEG (***Hypothesis 3a***) and Cardboard game (***Hypothesis 3b***) are reported.

Main Analysis: A multiple linear regression analysis was performed to predict the Learning Effect based on Focused Attention (i.e. fixation durations) for all pages (*M1*, *M2*, *M3* and *M4*) of the DEG Game *M* (Table 6.5). A statistically significant regression equation was found ($F(4, 52) = 2.839$, $p<.05$, Cohen's effect size $f^2 = 0.22$) with an R^2 of 0.179 after controlling the other fixation durations. The R^2 value indicates 17.9% of the variation of the Learning Effect can be explained by Focused Attention on each page of Game *M* of the DEG. There was a statistically significant predictive relationship between Focused Attention and Learning Effect for the game page *M*2 ($p<.01$) and M3 ($p<.05$). There was also statistically significant correlation for game page *M2* ($p<.05$). The predicted Learning Effect model for Game *M* equals to: -0.310 + 0.018(*M1*) + 0.102(*M2*) -0.078(*M3*) + 0.026(*M4*). As *M1* and *M4* were statistically insignificant predictors, the multiple regression analysis was re-run with only *M2* and *M3* as predictors ($F(2, 54) = 4.596$, $p=.01$, $f^2=0.20$, $R^2=0.15$), resulting in a new regression model: -0.143 + 0.105(*M2*) - 0.057(*M3*). From this model, the Learning Effect increased 0.105 points for each second attended to page *M2* and decrease 0.057 points for each second attended to page *M3*. This may be because the children were in the 'exploratory phase' for the first page (*M3*) of the hard task of Game *M* (cf. the tasks presented in *M1* and *M2* were relatively easy); the longer time they attended to *M3* was for making sense of the more challenging task presented there. The scatterplot of standardised predicted values versus standardised residual met the assumptions of homogeneity of variance, and the residuals were normally distributed ($p=.44$).

Table 6.5: Statistics, correlations and regression analysis summary for the Focused Attention (FA) Duration (in second) on the Learning Effect of DEG Game *M* and Game *S*.

Descriptive Statistics (DEG) M: (F (4, 52) = 2.839, p=.03, R^2=0.179) – normally distributed S: (F (4, 45) = 2.426, p=.06, $R^{2=}$ =0.177) - normally distributed					Multiple Regression	
	Variables	Mean	SD	Correlation with Learning Effect	b	β
Game M	Learning Effect Game *M*	.16	1.21		-.310	
	FA duration *M1*	7.77	4.81	.078	.018	.070
	FA duration *M2*	8.76	5.18	.287*	.102**	.439
	FA duration *M3*	10.86	6.32	-.052	-.078*	-.408
	FA duration *M4*	11.02	8.34	.151	.026	.176
Game S	Learning Effect Game *S*	.30	1.52		-1.362	
	FA duration *S1*	10.05	5.03	.360**	.077	.255
	FA duration *S2*	8.70	3.90	.141	-.012	-.030
	FA duration *S3*	9.47	4.32	.320**	.054	.153
	FA duration *S4*	10.47	4.72	.283*	.046	.143

*p<.05, ** p<.01

As for Game *S* of the same game learning media, a multiple linear regression was calculated. The predicted Learning Effect was calculated based on the Focused Attention given to the pages (*S1*, *S2*, *S3* and *S4*) of the DEG Game *S* (Table 6.5). A statistically non- significant regression was found (F (4, 45) = 2.426, p=.06) with an R^2 of 0.177 after controlling the other Focused Attention durations. There was no statistically significant predictive relationship between the Learning Effect and Focused Attention duration for any of the Game *S* pages. The scatterplot of standardised predicted values versus standardised residual met the assumptions of homogeneity of variance, and the residuals were normally distributed (p=.52). The results partially confirmed ***Hypothesis 3a*** predicting the impact of Focused Attention durations for the pages of the DEG on the Learning Effect.

As for the Cardboard game group (Table 6.6), the multiple linear regression analysis for all pages on Game M had a statistically non-significant result (F (4, 26) = 1.239, p=.32) with an R^2 =0.160 after controlling the other Focused Attention durations. No statistically significant relationship was found between the Learning Effect and Focused Attention durations for all pages of Game *M* (*M1* to *M4*).

Table 6.6: Statistics, correlations and regression analysis summary for the Focused Attention (FA) Duration on Learning Effect of Cardboard Game *M* and Game *S*.

Descriptive Statistics (Cardboard Game) M: (F (4, 26) = 1.239, p=.32, R^2 =0.160) – Not Normally Distributed S:(F (4, 29) = 0.228, p=0.92, R^2=0.031) – Not Normally Distributed					**Multiple Regression**	
	Variables	Mean	SD	Correlation with Learning Effect	b	β
Game M	Learning Effect Game *M*	.10	.70		-.063	
	FA duration *M1*	6.08	2.75	.200	.035	.137
	FA duration *M2*	9.35	5.87	-.242	-.036	-.305
	FA duration *M3*	10.15	4.74	.222	.038	.257
	FA duration *M4*	13.31	6.62	-.026	-.007	-.069
Game S	Learning Effect Game *S*	.35	1.01		.264	
	FA duration *S1*	17.86	8.02	.002	-.005	-.043
	FA duration *S2*	16.62	6.99	.044	.009	.059
	FA duration *S3*	16.74	7.22	-.057	-.022	-.160
	FA duration *S4*	19.62	8.13	.110	.021	.171

A similar result was found for Game *S* (Table 6.6) of the Cardboard game. A statistically non-significant effect (F (4, 29) = 0.228, p=0.92) with R^2=0.031 was found between the Learning Effect and Focused Attention duration for all pages of Game *S* (*S1* to *S4*). As a consequence, ***Hypothesis 3b*** predicting the impact of Focused Attention durations for the pages of the Cardboard game on the Learning Effect was rejected.

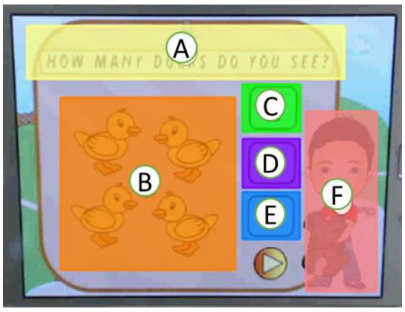
6.2.3 **Focused Attention on UI Objects.**

In this section the evaluation results of children's Focused Attention paid to relevant and irrelevant objects between high and low achievers of both DEG and Cardboard game for both Game *M* and Game *S* are reported. How to categorise an object as relevant or irrelevant is defined in Section 3.3.5. The relevant AOI objects are entities that are useful for playing the game task while irrelevant AOI objects are distractors. The Achievement Level was derived from the pre-test score (Prior Knowledge) over all participants of each group. The threshold for classifying low and high achievers was the average pre-test score of the children of both groups, DEG: Mean = 7.19, SD = 3.57 and Cardboard: Mean = 7.97, SD = 3.57. Those whose score was equal or above the average were categorised as ‘high’, otherwise ‘low’ achievers.

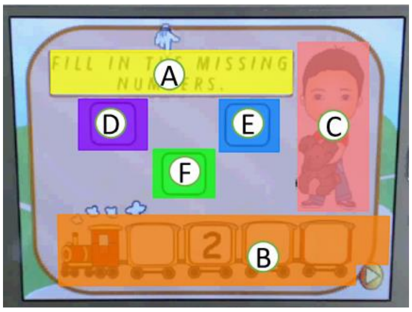
Main Analysis: The Mann-Whitney tests were used to evaluate the relationship between the Achievement Level (low or high) and the Focused Attention on relevant and irrelevant objects of both DEG and Cardboard game. For the DEG, in Game page *M1(*Table 6.7), the tests indicated that higher achievers paid statistically significantly more attention to relevant object *D* (answer) than low achievers (U=295, p=.03, r=0.27) did. Likewise, for Game *M* page *M4* results of the Mann-Whitney test showed that higher achievers paid more attention to relevant object *C* (answer) than low achievers did (U=312, p=.05, r=0.25). The results were consistent with the statistically significant correlations between the Pre-test and In-game scores discussed in Section 6.2.1. As it is logical that one gazes at an object when picking it, high achievers attended to (or fixated) the correct answers longer than low achievers. However, counterintuitively, no statistically significant differences were found for Game *M* page *M2* and *M3* of the DEG for all relevant and irrelevant objects. The incongruous pattern of the results across the four pages was unexplainable.

Furthermore, the zero Focused Attention indicated that low achievers did not look at the irrelevant object *F* (i.e. non-player character) at all on page *M2* or *M3*, and the negligible number of them looked at *F* on page *M4*. Similarly, high achievers did not look at *F* on page *M2* or *M4*. Only a small number of both groups of achievers, probably due to its novelty, looked at *F* on page *M1*, but with a relatively short duration. With *F* remaining unchanged, its attractiveness, as measured by Focused Attention, waned from the already quite low level for page *M1* to *M4*. However, a handful of high achievers did look at *F* on page *M3*. It might be triggered by the change of difficulty level from page *M2* to *M3* and some children might want to check if *F* was changed as well or if F would provide help. Since a statistically significant difference was found for pages *M1* and *M4* in Game *M* of the DEG, ***Hypothesis 4a(i)*** and ***Hypothesis 4a(iv)*** were accepted while ***Hypothesis 4a(ii)*** and ***Hypothesis 4a(iii)*** were rejected.

Table 6.7: Medians and ranges of Focused Attention on relevant and irrelevant objects of all pages in DEG Game *M* and Game *S*.



Game *M*



Game *S*

	Pages	Relevant Objects		Irrelevant Objects	
	Achiever	Low (n= 30)	High (n= 29)	Low (n= 30)	High (n= 29)
Game *M*	*M1*	Object *D**		Object *F*	
		0.42 (.23 –.65)	0.57 (.47 –.97)	0.00 (.00 -.34)	0.00 (.00 -.28)
	M2	Object *E*		Object *F*	
		0.41 (.00 –.61)	0.37 (.26 –.70)	0.00 (.00 -.00)	0.00 (.00 -.00)
	M3	Object *D*		Object *F*	
		0.45 (.22 –.57)	0.47 (.17 –.73)	0.00 (.00 -.00)	0.00 (.00 -.31)
	M4	Object *C**		Object *F*	
		0.40 (.00 –.69)	0.65 (.27 –.89)	0.00 (.00 -.04)	0.00 (.00 -.00)
Game *S*		Object *B*		Object *C*	
	S1	0.68 (.49 –.82)	0.57 (.43 –.68)	0.16 (.00 -.36)	0.00 (.00 -.37)
		Object *B* *		Object *C*	
	S2	0.66 (.53 -.76)	0.46 (.41 -.61)	0.00 (.00 -.48)	0.00 (.00 -.41)
		Object *B* *		Object *C*	
	S3	0.65 (.45 -.78)	0.51 (.42 -.61)	0.14 (.00 -.38)	0.00 (.00 -.10)
		Object *B* *		Object *C*	
	S4	0.66 (.53 -.86)	0.52 (.44 -.62)	0.00 (.00 -.28)	0.00 (.00 -.36)

*p<.05

As for the attention paid to the objects in Game *S* of the DEG (Table 6.7), given its different design, the Mann-Whitney tests for pages *S2*, *S3* and *S4* indicated that higher achievers paid less attention to relevant object *B* than for low achievers (U_{page2}=230, p_{page2}<.01, r=0.41; U_{page3}=299, p_{page3}<.05, r=0.27; U_{page4}=220, p_{page4}<.01, r=0.42). However, no statistically significant differences were found on page *S1* (i.e. the easiest level) of Game *S* for the relevant objects. The results revealed that low achievers tended to look at object *B* (i.e. the moving train) longer because they had no strategy in the gameplay even after going through the exploratory phase on page *S1*. Additionally, those children could also be attracted to the dynamic object *B* (incoming train) which was considered eye-catching for young children. Contrarily, high achievers looked less at object *B* because they had some gameplay strategy - distributing attention evenly among the other objects such as the answer choices *D*, *E* and *F*.

Furthermore, a pattern for irrelevant object *C* (non-player character) similar to Game *M* was observed for Game *S*. The median Focused Attention was zero on all four pages of the high achievers and low achievers on page *S2* and *S4*. A handful of low achievers looked at object *C* on page *S1* and *S3* with a short duration. Overall, the different patterns of Focused Attention for objects on the different game interface designs (Game *M* and Game *S*) suggested that low and high achievers applied different strategies to deal with a given task. Therefore, from the results for pages in Games *S* of the DEG ***Hypothesis 4b(i)*** was rejected while ***Hypothesis 4b(ii)***, ***Hypothesis 4b(iii)*** and ***Hypothesis 4b(iv)*** were accepted.

As for the Cardboard game (Table 6.8), the Mann-Whitney tests on Game *M* for pages *M1*, *M2* and *M4* indicated that lower achievers paid more attention to other relevant objects than for higher achievers ($U_{page1}=88$, $p_{page1}=.04$, r=0.41; $U_{page2}=84$, $p_{page2}=.03$, r=0.38; $U_{page4}=87$, $p_{page4}=.04$, r=0.36). The results revealed that lower achievers looked longer at other relevant objects that were not the correct answer (object D), except for page *M3* (i.e., the first level of hard-task) where there was no statistically significant difference in Focused Attention between relevant and irrelevant objects. Overall, for Cardboard Game *M*, low achievers paid more attention to its objects than high achievers. It might attribute to the fact that lower achieving children had to recount the number objects repeatedly or they looked at the objects without a clear goal or strategy. Therefore, from the results above, ***Hypothesis 4c(i)***, ***Hypothesis 4c(ii)*** and ***Hypothesis 4c(iv)*** were accepted while ***Hypothesis 4c(iii)*** was rejected.

Table 6.8: Medians and ranges of Focused Attention for relevant and irrelevant objects on all Game *M* and Game *S* pages of the Cardboard game.

	Pages	**Relevant Objects**		**Irrelevant Objects**	
	Achiever	Low (n= 15)	High (n= 20)	Low (n= 15)	High (n= 20)
Game *M*	*M1*	Object E*		Object F	
		0.21 (.00 -.36)	0.00 (.00 -.00)	0.44 (.33 -.53)	0.36 (.21 -.49)
	M2	Object C*		Object F	
		0.37 (.23 -.83)	0.24 (.00 -.35)	0.36 (.28 -.46)	0.35 (.03 -.41)
	M3	Object D		Object F	
		0.40 (.33 -.67)	0.32 (.00 -.58)	0.42 (.39 -.49)	0.25 (.05 -.40)
	M4	Object B*		Object F	
		0.50 (.42 -.63)	0.41 (.38 -.53)	0.43 (.31 -.50)	0.27 (.00 -.45)
Game *S*		Object B		Object C	
	S1	0.52 (.45 -.68)	0.56 (.47 -.67)	0.44 (.36 -.57)	0.47 (.34 -.61)
		Object B		Object C	
	S2	0.59 (.51 -.64)	0.54 (.49 -.65)	0.53 (.39 -.59)	0.49 (.28 -.62)
		Object B		Object C	
	S3	0.59 (.46 -.67)	0.55 (.43 -.64)	0.53 (.36 -.62)	0.35 (.00 -.47)
		Object E*		Object C	
	S4	0.13 (.00 -.47)	0.45 (.32 -.73)	0.42 (.28 -.53)	0.36 (.01 -.53)

*p<.05

As for the Focused Attention to the objects in Game *S* of the Cardboard game (Table 6.8), the Mann-Whitney tests for pages *S1*, *S2* and *S3* showed no statistically significant difference between high and low achievers. However, for page *S4*, higher achievers paid statistically significantly more attention to the relevant object *E* than low achievers (U=86, p=.03, r=0.36). Overall, given the different design of Game *S*, high and low achievers revealed that both groups gazed longer at the relevant objects as compared to irrelevant objects. This extended duration can be attributed to the observation that the children had more hands-on interactions (Velcro strips and Blu Tack) with the Cardboard game, making the irrelevant objects less distracting. Therefore, from the results, only Hypothesis ***4d (iv)*** was accepted while ***Hypothesis 4d (i)***, ***Hypothesis 4d (ii)*** and ***Hypothesis 4d(iii)*** were rejected.

In addition, the total Focused Attention for the actual gameplay was estimated by summing up the related measures over the four pages of Game *M* and Game *S*. Results are shown in Table 6.9; the total Focused Attention calculated for the DEG was 86.31 seconds (1.4 minutes). The duration was relatively short, which was attributed to the game design. If there were more game scenarios with more levels, it would be longer. Furthermore, there are four other pages in the game - introduction, selection, demo and reward -which were not analysed for total Focused Attention. Nonetheless, the total Focused Attention calculated for the Cardboard game was 116.34 seconds (1.9 minutes).

Table 6.9: Total Focused Attention (in seconds) for each of four pages of Game *M* and Game *S* of both DEG and Cardboard game.

		DEG		**Cardboard**	
	Pages/Level	Mean	SD	Mean	SD
Game *M*	*M1*	8.37	5.87	6.59	4.25
	M2	9.36	6.25	8.99	5.71
	M3	11.42	8.78	11.00	5.66
	M4	11.82	10.70	15.53	12.07
	Total *M*	40.97	31.6	42.11	27.69
Game *S*	*S1*	11.44	6.26	18.09	8.16
	S2	10.15	5.86	18.58	12.75
	S3	11.35	7.46	17.16	9.02
	S4	12.40	7.87	20.40	9.14
	Total *S*	45.34	27.45	74.23	39.07

6.2.4 **Recall and Recognition**

In this section, results on evaluating the relationships between the difficulty levels of both game learning media are reported (***Hypothesis 5a*** and ***5b***). The recognition-based tasks of the relatively low difficulty level of Game *M* and recall-based tasks of relatively high difficulty level of Game *S* of both DEG and Cardboard game were involved. For both Game M and Game S, the easier tasks presented in pages 1 and 2 involved numbers below and equal to 10 and the hard tasks presented in pages 3 and 4 were associated with numbers above 10 up to 20. The In-game scores were calculated by assigning 1 score for a correct answer and 0 for a wrong answer. The maximum in-game score for Game *M* was 4 and for Game *S* was 12.

Main Analysis: The Wilcoxon signed-ranks tests were performed between the recognition-based Game *M* and recall-based Game *S* with respect to the In-game scores of each individual page.

There was a statistically significant difference for both easy-tasks (Table 6.10) which were presented by page 1 (Z=8.66, p<.01, r=0.63) and page 2 (Z=8.22, p<.01, r=0.59) between Game *M* and Game *S.* For the hard-tasks (Table 6.10), there were also statistically significant differences for both page 3 (Z=7.67, p<.01, r=0.55) and page 4 (Z=6.70, p<.01, r=0.48) between Game *M* and Game *S.* The statistically significant difference (p<.01) between Game *M* and Game *S* showed that the children had more difficulty in dealing with the recall-based tasks in Game *S* as compared to the recognition-based tasks in Game *M*, thus accepting ***Hypothesis 5a*** and ***Hypothesis 5b***. This result corroborated the theoretical model that as recall entails a deeper level of information retrieval than recognition (Baddeley, 1997), it is more challenging.

Table 6.10: Medians and ranges of In-game score for Game *M* (recognition) and Game *S* (recall) for each level on both DEG and Cardboard Game

	Pages (level)	Z , p	Game	Median	Range
Easy tasks	Page 1	8.66, p<.01**	*M*	1.00	1.00 –1.00
			S	3.00	3.00 –3.00
	Page 2	8.22, p<.01**	*M*	1.00	.00 –1.00
			S	3.00	2.00 –3.00
Hard tasks	Page 3	7.67, p<.01**	*M*	1.00	.00 –1.00
			S	2.00	1.00 –3.00
	Page 4	6.70, p<.01**	*M*	1.00	.00 –1.00
			S	2.00	1.00 –3.00

**p<.01

6.3 Method 2

6.3.1 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, the variables (pre-test, post-test and in-game variables) were examined and observed for outliers. Cases with extreme outliers (values more than three IQRs 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.

6.3.2 Gaze Sub-Sequence Score on Learning Effect

In this section results of evaluating the relationship between the Gaze Sub-Sequence Scores (interaction strategy) and the Learning Effect for both game learning media, DEG and Cardboard are reported. First, the difference in the Gaze Sub-Sequence Scores between the Performance Level are discussed (***Hypothesis 6.1a***, ***Hypothesis 6.1b***). Then the correlations between the observed variables for both DEG (***Hypothesis 6.1c***) and Cardboard (***Hypothesis 6.1d***) game on both Game *M* and Game *S* are presented. The Learning Effect (Section 6.1.3) threshold for classifying the high and low performance was the average differences between the pre and post-test of both game learning media, DEG: Game *M* (Mean = 0.17, SD = 0.93), Game *S* (Mean = 0.24, SD = 1.006) and Cardboard: Game *M* (Mean = 0.09, SD = 0.74), Game *S* (Mean = 0.37, SD = 1.003).

Main Analysis: A Mann Whitney test was used to evaluate the difference in the Gaze Sub-Sequence Scores between the high and low performance on each Game *M* and Game *S* of both game learning media. For the DEG, no statistically significant difference in the Gaze Sub-Sequence Score 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 ($Mdn_{DEGGame M}$ =4.00, Range= 3.50-5.00; $Mdn_{DEGGame S}$ =4.00, Range= 2.00-5.00) performance. A similar result was seen for the Cardboard game, where no statistically significant difference in the Gaze Sub-Sequence Score 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 ($Mdn_{CardboardGame M}$ =4.00, Range= 3.50-5.00; $Mdn_{CardboardGame S}$ =4.00, Range= 3.00-6.00) performance.

The statistically non-significant difference between the high and low performance was probably due to the game design, which might not be a good differentiator in studying the Learning Effect of the game. The game design was developed based on young children's attention span and was tested in the pilot study. Another aspect was the ceiling event of the pre/post-test, although it was tested and improved in the pilot study. Therefore, ***Hypothesis 6.1a*** and ***Hypothesis 6.1b*** concerning Gaze Sub-Sequence Score and the Learning Effect were rejected.

In addition, a Spearman tests was performed to assess the correlation between the Learning Effect and Gaze Sub-Sequence Scores for both game tasks on the DEG and Cardboard game. For the DEG, both game tasks, Game *M* (N = 59, r=0.03, p=.82) and Game *S* (N = 59, r=0.20, p=.14) showed no correlation between the two variable. A similar result was observed for the Cardboard game with no statistically significant correlation was obtained between the Learning Effect and Gaze Sub-Sequence Score on both game tasks, *M* (N = 35, r=0.10, p=.57) and *S* (N = 35, r=0.20, p=.24). Therefore, ***Hypothesis 6.1c*** and ***Hypothesis 6.1d*** concerning the correlation between Learning Effect and Gaze Sub-Sequence Score were rejected.

6.3.3 Gaze Sub-Sequence Score on Gender

In this section, results of evaluating the relationship between the Gaze Sub-Sequence Scores (interaction strategy) and the Gender for both game learning media, DEG (***Hypothesis 6.2a***) and Cardboard game (***Hypothesis 6.2b***) are reported. Gaze Sub-sequence Score were calculated as discussed in Section 6.1.

Main Analysis: A Mann Whitney test was used to evaluate the difference in the Gaze Sub-Sequence Scores score between Genders on each Game *M* and Game *S* of both game learning media. For the DEG, no statistically significant difference in the Gaze Sub-Sequence Score was found for Game *M* (U =432, Z=0.05, p=.96) and Game *S* (U =332, Z=1.58, p=.12) between both boys and girls ($Mdn_{DEGGame\mathit{M}}$ =4.00, Range= 3.50-5.00; $Mdn_{DEGGame\mathit{S}}$ =4.00, Range= 2.00-5.00). A similar result was seen for the Cardboard game, where no statistically significant difference in the Gaze Sub-Sequence Score was found for Game *M* (U =112, Z=1.28, p=.20) and Game *S* (U =134, Z=0.54, p=.59) - between both genders ($Mdn_{CardboardGame\mathit{M}}$ =4.00, Range= 3.50-5.00; $Mdn_{CardboardGame\mathit{S}}$ =4.00, Range= 3.00-6.00). The statistically non-significant difference between genders showed that children at this age do not have any inherent difference in interaction strategy. Both girls and boys have the ability to apply an interaction strategy while playing a DEG related to numeracy at a young age. Therefore, based on the Mann Whitney results, ***Hypothesis 6.2a*** and ***Hypothesis 6.2b*** were rejected.

6.3.4 Gaze Sub-Sequence Score on Achievement Level

In this section results of evaluating the relationship between the Gaze Sub-Sequence Scores (interaction strategy) and the Achievement Level for both game learning media, DEG (***Hypothesis 6.3a***) and Cardboard (***Hypothesis 6.3b***) are reported. Gaze Sub-sequence Scores were calculated as discussed in Section 6.1.

Main Analysis: Mann Whitney tests were used to evaluate the difference in the Gaze Sub-Sequence Scores between high and low achievers on each Game *M* and Game *S* of both game learning media. For the DEG, there was a statistically significant difference in the Gaze Sub-Sequence Score on both Game *M* (U =288, Z=2.27, p=.02) and Game *S* (U =261, Z=2.67, p=.01) between high and low ($Mdn_{DEGGame M}$ =4.00, Range= 3.50-5.00; $Mdn_{DEGGame S}$ =4.00, Range= 2.00-5.00) achievers. However, for the Cardboard game, no statistically significant difference in the Gaze Sub-Sequence Score was found on Game *M* (U =137, Z=0.44, p=.66) and Game *S* (U =106, Z=1.48, p=.14) between high and low ($Mdn_{CardboardGame M}$ =4.00, Range= 3.50-5.00; $Mdn_{CardboardGame S}$ =4.00, Range= 3.00-6.00) achievers. 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 because of the nature of the physical game where interaction with the game was more sensory oriented (concrete interaction). In the DEG, children were less distracted since they did not interact with any concrete objects hence allowing the children to think abstractly and develop interaction strategies. While for the Cardboard game, the children's cognitive interest tends to be on the concrete objects they interact with, thereby limiting the children to think abstractly and develop interaction strategies due to their low cognitive flexibility. Therefore, interaction strategies were not observed for the Cardboard game thus rejecting ***Hypothesis 6.3a*** while ***Hypothesis 6.3b*** with regards to the DEG was accepted.

Sub-sequence Matrices Results: In Game *M* of the DEG (Figure 6.1), the game sub-sequence matrices showed that for 3 out of 15 female children with high achievement their first-fixated AOI of page *M1* was the "instruction" *A* (yellow) while 9 was the "counting objects" *B* (orange) and the other 3 was "answer boxes" *D* (purple) and *E* (blue).

Male And Low	M1(15)	M2(15)	M3(15)	M4(15)
	0	BABA	B	ABAB
	BABA	BABC	BABA	BABE
	BABA	BADA	BABC	BCBA
	BACB	BCAB	BCDA	BCBC
	BACB	BDBD	BCDB	BDBA
	BACB	BDBD	BDCD	BDBE
	BADB	BDBD	BEAB	BDBE
	BCAE	BDBD	BEBC	BDCE
	BCAE	BDCE	BEBC	BECA
	BDEB	BDEB	BEBD	BEFB
	BEBD	BDEB	BEBE	BFBA
	BECB	BDED	BEDB	DBAB
	CABD	BEBD	BEDC	DBDC
	CDBD	BEBD	DBDB	EBDC
	DFBD	FBCD	EBEB	FBEB

Male And High	M1(14)	M2(14)	M3(14)	M4(14)
	ACDA	BCAB	BABA	BABE
	BABA	BCBA	BABA	BABE
	BABA	BCBA	BABD	BACA
	BACA	BCDC	BCDC	BADB
	BAEB	BCDC	BCDC	BCBA
	BAFC	BDAC	BDAD	BCBC
	BCAB	BDBD	BDBD	BCBC
	BCAB	BDEB	BDCB	BCBD
	BCAB	BDEB	BDEB	BCDE
	BDBD	BEBA	BEBD	BDBD
	BDEA	BEBD	BEDE	BDBD
	BDEF	BEBD	EBAC	BDCB
	BEAB	DBEB	EBDB	BECD
	CABD	EBAB	EBDC	DBCB

Female And Low	M1(15)	M2(15)	M3(15)	M4(15)
	BABA	B	BABA	0
	BABC	BABA	BABA	BABD
	BABE	BABF	BABA	BABE
	BABE	BACB	BABE	BABF
	BADA	BADA	BACB	BADB
	BADB	BAEB	BADC	BCAB
	BCAB	BAEF	BCBA	BCBA
	BDAB	BCBA	BDBD	BDBE
	BDAC	BCBD	BDBD	BDEB
	BDAD	BDBA	BDCE	BDED
	DBDB	BDBD	BECB	BEAC
	DFDA	BEBD	BCBC	BEBC
	EBAB	BEBE	DEDE	BFCA
	EDBC	DBDB	EBCB	CDAB
	FHAD	FBCA	EBDB	FBAB

Female And High	M1(15)	M2(15)	M3(15)	M4(15)
	ABAD	BABD	BABA	BABA
	ABCA	BABF	BABA	BACA
	AEBE	BABD	BABC	BCAB
	BABC	BACB	BABC	BCBC
	BADE	BADB	BABD	BCBC
	BAEA	BCAC	BADB	BCBE
	BAEB	BCAC	BCBD	BCDB
	BAEB	BCBC	BDAC	BDCB
	BDAB	BCBD	BDEA	BDCB
	BDBA	BCDA	BEBA	BDED
	BDBC	BDBD	BECB	BECB
	BDEB	BDBD	BEDB	CDBE
	DBDB	BDCB	BEDC	DBCB
	DBDB	BDEB	BFDC	EBAB
	EABF	BEDE	EBCB	EBDE

Legend:	
Yellow	AOI A (Instructions)
Orange	AOI B (Counting objects)
Green	AOI C (Answer box)
Purple	AOI D (Answer box)
Blue	AOI E (Answer box)
Red	AOI F (NPC)
White	Invalid

Figure 6.1: Gaze sub- sequence matrices between genders and achievement levels (Male -Low; Male- High; Female-Low; Female-High) for Game *M* of the DEG.

The matrices also showed that across Game *M*, several low achievement children, both male and female, were distracted by the "NPC" *F* (red) at the beginning of the game pages. Note that white boxes in the matrices (Figure 6.1) indicate invalid data and they are not included in the sub-sequence percentage calculation.

Male And Low	S1 (15)	S2(15)	S3(15)	S4(15)
	ABFC	BDBE	BABE	BDFB
	BADA	BDBF	BACA	BEBA
	BAFA	BDBF	BEBD	BEBA
	BCEB	BDBF	BFBD	BEBA
	BCEF	BFBE	BFBF	BEBE
	BDBA	BFBF	BFBF	BFBF
	BEAE	BFBF	BFEB	BFBF
	BFBE	BFBF	BFEF	CABD
	BFAB	BFBF	CEAE	CFDF
	CBFB	BFBF	DBFA	CFEA
	CEAD	BFBF	DBFB	DBFB
	CEFB	CFBF	FBFD	FBAD
	ECAD	EBAF	FBFE	FBFA
	EFBA	EBDB	FCBF	FCED
	FECF	FEBE	FEBF	FDBD

Male And High	S1(14)	S2(14)	S3(14)	S4(14)
	ABFB	BDBD	BADF	BCBA
	BCEA	BDFE	BDBD	BCEC
	BEBF	BEBF	BEFB	BCFB
	BEBF	BEFB	BEFD	BCFE
	BEFD	BFBD	BFBF	BEDB
	BEFD	BFBE	BFBF	BEFB
	BFBE	BFBF	BFBF	BFBA
	BFBF	BFBF	BFDB	BFBC
	BFDA	BFBF	BFDE	BFEB
	BFEB	BFBF	BFDF	BFED
	CBCB	BFBF	CBEF	BCFD
	CEFB	BFBF	CEFD	EBFB
	EBFD	BFBF	EFBE	FBFB
	ECEB	FBFB	FBFB	FEDB

Female And Low	S1 (15)	S2(15)	S3(15)	S4(15)
	ABDB	ABCF	BADB	AFED
	BADB	BCBF	BDAD	BDBD
	BCAE	BCEF	BDBE	BEAF
	BDBD	BCFE	BDBF	BEBF
	BEBE	BDCB	BDBF	BEBF
	BEBF	BFBD	BEAE	BFBE
	BFBE	BFBE	BEDF	BFBE
	BFBE	BFBE	BFBE	BFBE
	BFBE	BFBE	BFBF	BFBF
	BFCB	BFBF	BFCE	BFDB
	BFDF	BFBF	BFDB	BFDF
	CBCF	DBDF	BFEB	EBFD
	EBAB	FBFB	EBEB	EFBE
	EFDB	FBFD	FBCD	FBFD
	FBFB	FDED	FBEB	FECB

Female And High	S1 (15)	S2(15)	S3(15)	S4(15)
	BCFA	BDEB	BEAC	AEBE
	BDBF	BEBF	BEDB	BEAB
	BEBE	BFAC	BFBA	BFAF
	BECA	BFBE	BFBE	BFBD
	BEFB	BFBF	BFDA	BFDB
	BEFB	BFBF	BFDE	BFEB
	BFBD	BFBF	CAEF	BFEB
	BFDA	BFDA	DABF	DABE
	BADB	BFDF	DBFB	DBFB
	BFBA	EBFB	EBFA	DBFE
	EBEB	EFBD	FBDE	DFBE
	EFDB	FBAB	FBED	EDFD
	FBAB	FBEF	FBFB	EFDF
	FBFA	FBFB	FBFB	FBFB
	FDBD	FBFD	FBFD	FBFB

Legend:	
Yellow	AOI A (Instructions)
Orange	AOI B (Counting objects)
Red	AOI C (NPC)
Purple	AOI D (Answer box)
Blue	AOI E (Answer box)
Green	AOI F (Answer box)
White	Invalid

Figure 6.2: Game sub- sequence matrices between genders and achievement levels (Male -Low; Male- High; Female-Low; Female-High) for Game *S* of the DEG.

As for Game *S* of the DEG (Figure 6.2), for male children with low achievement, red boxes can be seen across the levels from *S1* to *S4*, indicating the number of children distracted by the “NPC” *C* (red) in the beginning of each page in Game *S*. The matrices also showed that male children were prone to be distracted to the “NPC” *C* (red) compared to female children. 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 maybe because some children tended to stare in the middle of the tablet screen while waiting for the next level to appear.

Male And Low	M1(7)	M2(7)	M3(7)	M4(7)
	ABCF	ABCF	BDBD	ADBA
	AFDB	ABDF	BEBA	BABA
	BDBC	AFAB	CBDF	BCBF
	FBDF	BDBD	FABE	BCDB
	FDAB	BFBD	FAFB	BEBE
	FDBE	FBDB	FDFC	FBFB
	FDEB	FBFE	FEFE	FCBE

Male And High	M1(8)	M2(8)	M3(8)	M4(8)
	A	A	BABA	ABCB
	BABA	BABD	BCFA	ABCD
	BAFB	BADB	BEDC	ABCD
	BFBA	BFAB	CACA	BABC
	DBAC	BFCF	FABA	BABE
	FABF	CBCB	FBCB	BEDC
	FBFB	DBFE	FBDB	DFBD
	FCAB	FABA	FBDE	FDBC

Female And Low	M1(8)	M2(8)	M3(8)	M4(8)
	ACBA	ABAB	ACBA	BCBC
	BABC	ACBC	BCBD	BDBA
	BAFB	ACBC	DBAC	BDBD
	BCBA	BDEB	FACB	BEBF
	CFBD	BFDB	FBCD	CBEB
	CFCA	DACB	FBDE	DBAC
	FBFB	DBFB	FDFE	FCAB
	FDCE	EFBA	FEBF	FDFB

Female And High	M1(12)	M2(12)	M3(12)	M4(12)
	ABDB	ABCA	ACFA	ABDB
	ABFD	ABCB	BABC	B
	ACFC	AFBE	BCBD	BAFA
	AFBA	BCBC	BDBC	BCBC
	BABD	BDBA	BDBF	BCBC
	BAFC	BDCA	BFCD	BCBD
	BFBF	BDED	CBAC	BDCB
	CABA	BFCA	CBDB	BFBA
	CBAF	CBCF	CBFD	BFBD
	CFAB	DCFB	FCFB	CBDB
	DFDA	ECBA	FDCB	DEDC
	FBFC	FDCB	FDCD	FCFB

Legend:	
Yellow	AOI A (Instructions)
Orange	AOI B (Counting objects)
Green	AOI C (Answer box)
Purple	AOI D (Answer box)
Blue	AOI E (Answer box)
Red	AOI F (NPC)
White	Invalid

Figure 6.3: Gaze sub- sequence matrices between genders and achievement levels (Male -Low; Male- High; Female-Low; Female-High) for Game *M* of the Cardboard game.

As for Game *M* of the Cardboard game (Figure 6.3), the gaze sub-sequence matrices showed that female children with high achievement were likely to fixated 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*, all four groups of children were distracted by "NPC" *F* (red) but male children with low achievement tended to be so distracted more frequently.

For Game *S* of the Cardboard game (Figure 6.4), similar to Game *M* of the Cardboard, the gaze sub-sequence matrices showed that female children with high achievement tended to look at "instruction" *A* (yellow) first in the beginning of each page in Game *S*. Both female children with high and low achievement were constantly looking at the "answer box" *F* (green) across the levels from *S1* to *S4*, indicating a number of children from this group scanned from the middle to the other areas of the game. Overall, similar to Game *M* of the Cardboard game, all four groups of Game *S* were distracted by the "NPC" *C* (red), but low achievement children of both genders tended to be so distracted more frequently.

Male And Low	S1(7)	S2(7)	S3(7)	S4(7)
	BDBA	BADA	ABCB	BCBC
	BEBE	BEBE	BABC	CBED
	BEDC	BFBF	BEFB	CBFB
	BFAB	CBEB	BFBF	CFEB
	CEBF	CFBD	BFBF	EFCB
	CFBC	EBCE	CEDB	FBDC
	DFBC	EFEC	FBCF	FBEB

Male And High	S1(8)	S2(8)	S3(8)	S4(8)
	ADBF	BDBD	BDBD	BDBC
	AFBD	BDBD	BDBF	BDBE
	BCAB	BECF	BEBE	BEFE
	BCEB	CBDB	BFBE	BFBA
	BDAB	CBEC	BFDB	BFBD
	BFBF	CBEF	CBCB	BFBE
	EBCB	CBFB	CBCB	BFEB
	FBEB	EDBA	FDEB	CBCB

Female And Low	S1(8)	S2(8)	S3(8)	S4(8)
	BCBE	ACAB	BCBD	BEBF
	BEBE	AEAC	BEBF	BFEB
	BFBC	BCDB	BFBD	CAFC
	CECE	BCEB	CBDB	CBEF
	CEFA	BFEC	DAEA	EBFB
	DEBF	CEBF	ECDA	EFAF
	FBAB	DBDA	EDBE	EFBC
	FDBA	FABD	FBDE	FECF

Female And High	S1(12)	S2(12)	S3(12)	S4(12)
	ABEB	BADA	ABDA	ABDB
	AEAB	BADB	AEDF	BCBE
	AEFC	BCEC	BDBE	BEBF
	AFAB	BDBE	BDED	BFBE
	BADB	CEAB	BFEA	BFCE
	BCBD	CEFB	BFEB	BFEF
	BEDB	DEDB	DFBF	CAEB
	CAFE	EAFB	EAFE	CBEF
	EBDB	FBAE	FBEB	CEBC
	EBEB	FBDB	FBFE	EDAD
	ECBE	FBDB	FCBE	FABF
	FABE	FCFB	FEBD	FBEF

Legend:	
Yellow	AOI A (Instructions)
Orange	AOI B (Counting objects)
Red	AOI C (NPC)
Purple	AOI D (Answer box)
Blue	AOI E (Answer box)
Green	AOI F (Answer box)
White	Invalid

Figure 6.4: Gaze sub- sequence matrices between genders and achievement levels (Male -Low; Male- High; Female-Low; Female-High) for Game *S* of the Cardboard game.

In conclusion, children using the Cardboard game were more distracted by the "NPC" (red) as compared to those using the DEG game. This may be because children tended to think the "NPC" on the Cardboard game had a function to play while the DEG was only a decorative character.

Sub-sequence Percentage (on UI objects) Results: In this section the results of the sub-sequence percentage (%) calculated earlier in Figure 5.8 are discussed. The percentage results are then used to compare with those of the heat map visualisations analysed by the Tobii Studio software in the next section (Section 6.3.5). A Chi-square tests was performed to check the significance differences of the observed variables used in the sub-sequence percentage results and was reported earlier in Section 6.1.4.

For Game *M* of the DEG (Table 6.11), both male and female children with high achievement on average gazed 100% on Relevant UI objects in the beginning of the game page levels as compared to 87% and 80% for the children with low achievement. This showed that higher achieving children did not get distracted with the NPC when entering every level page.

Table 6.11: Sub-sequence percentage (on relevant (R) UI object) between genders and achievement level on both Games *M* and *S* of the DEG.

Groups (DEG) ***R= relevant***	**Game *M* (%)**				**Mean *M* (%)**	**Game *S* (%)**				**Mean *S* (%)**
	M1	*M2*	*M3*	*M4*		*S1*	*S2*	*S3*	*S4*	
Male x Low (R)	100	74	100	74	87	44	74	74	44	59
Male x High (R)	100	100	100	100	100	55	100	55	100	78
Female x Low (R)	74	72	100	72	80	74	100	100	100	94
Female x High (R)	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. Female children, both high and low achievers on average gazed 94% on Relevant UI object in the beginning of the game levels compared to male children with 59% and 78% for the high and low achievement group, respectively. The result implied that in Game *S* of the DEG, male children were easily distracted with the NPC and male children with low achievement were constantly distracted across the four levels of Game *S*. Overall for the DEG, female children with high achievement had the highest average percentage of gazing at Relevant UI objects when entering every level page with 100% in Game *M* and 94% for Game *S*.

As for Game *M* of the Cardboard game (Table 6.12), on average female children, low and high, had a higher percentage in gazing at Relevant UI objects, 61% and 47%, compared to male children with 38% and 28%, respectively. None of the groups maintained a higher percentage gazing at Relevant UI object across the four levels of Game *M*. All groups of children playing Game *M* of the Cardboard game were most distracted with the NPC on page *M3*. However, an obvious drastic low percentage on relevant objects from page *M2* to *M3* can be observed for female children with low achievement. This may be because of the difficulty in progressing from the easy to the hard level. The hard level task began on page *M3* which might prompt the children to wander around the Cardboard game and female children with low achievement tended to wander a lot more than the other groups.

Table 6.12: Sub-sequence percentage (on relevant (R) UI object) between genders and achievement level on both Games M and S of the Cardboard game.

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

As for Game *S* of the Cardboard game (Table 6.12), children with high achievement of both genders had a higher percentage in gazing at Relevant UI objects, 64% (female) and 53% (male), compared to low achievement children with 48% (female) and 36% (male). Male and female children with high achievement gazed 100% at Relevant UI object once across the four levels however none of the groups constantly maintain a higher percentage. A drastic fall of percentage on the relevant object from page *S1* to *S2* can be observed for male children 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, female children with high achievement had the highest average percentage of gazing at the relevant UI objects with 61% for Game *M* and 64% for Game *S*.

Based on the results, no consistent gaze sequence can be identified among male and female children. However, for both DEG and Cardboard game, female with high achievement showed a higher percentage of gazing at relevant 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.

6.3.5 Sub-sequence Percentage (Gaze Sub-Sequence Analysis) and Heat Maps (Visualisation Data)

In this section the similarity and dissimilarity between gaze sub-sequence analysis (sub-sequence percentage) results (Section 0) and the visualisation data (heat map) results analysed by the Tobii Studio software are analysed. Learning the difference between both eye-tracking data analysis methods can gain some insight into the human cognitive process (Le Meur & Baccino, 2013).

Heat Maps Preparation: Heat maps were analysed from the Tobii Studio software. Using the software, recordings were categorised according to the clusters described in Section 0, which were Male -Low; Male- High; Female-Low; Female-High. The number of heat maps for all four clusters for both game learning media with all pages of Game M (*M1, M2, M3, M4)* and *(S1, S2, S3, S4)* was so large that it would be too resource-demanding to analyse all of them. Therefore, to streamline the analysis, focuses were set on page 1 and page 4 of Game *M* and Game *S*, representing the easy-task and hard-task of the games, respectively. The use of this subset of pages was reasonable because the design layouts for all pages of each game (Game *M* and Game *S*) were highly similar.

Comparing Sub-Sequence Percentage and Heat Maps Results: For Game *M* of the easy-task (page *M1*) in the DEG, comparison was made between the sub-sequence percentages results (Table 6.11) with the heat maps (Figure 6.5) extracted from the Tobii Studio. The heat map colour scheme begins from red down to green. Red indicated the highest fixated or deepest concentration degree while green indicated the least, and yellow indicated an average concentration degree. In Game *M1* of the DEG the sub-sequence percentage results and visualisation data (heat maps) were partially consistent. From the heat maps there were no visual fixations on the Irrelevant “NPC” object for both male-high and female-high cluster group, a similar result presented by the sub-sequence percentage results discussed earlier (Table 6.11), where there was 0% (i.e., Relevant (R) =100%) gazed at Irrelevant objects.

(a) R=100 (b) R=100 (c) R=74 (d) R=100

Figure 6.5: Gender vs. Achievement heatmaps for DEG Page *M1* (matching): (a) Male-Low, (b) Male-High, (c) Female-Low, and (d) Female-High.

Results for the female-low cluster were also consistent, a detectable but not so intense visual fixations (i.e., green indicator) for the Irrelevant "NPC" object in the heat-map could be seen and a 26% (R=74%) gaze on Irrelevant objects was presented by the sub-sequence percentage results. However, for the male-low cluster group the heat map result was not consistent with the sub-sequence percentage result, a green indicator on the Irrelevant "NPC" object was seen in the heatmap but had 0% (R=100%) gaze on Irrelevant objects.

(a) R=74 (b) R=100 (c) R=72 (d) R=100

Figure 6.6: Gender vs. Achievement heatmaps for DEG Page *M4* (matching): (a) Male-Low, (b) Male-High, (c) Female-Low, and (d) Female-High.

As for Game *M* of the hard-task (page *M4*) in the DEG (Figure 6.6), the sub-sequence percentages results and heat map were also partially consistent. Accumulated fixations on the Irrelevant "NPC" object for both male-low and female-low visualised greenish yellow areas on the heat map which was consistent with the sub-sequence percentage results, 26% (i.e., Relevant (R) =74%) and 28% (R=72%) gaze on Irrelevant object. The male-high cluster was also consistent with the sub-sequence percentage results, having no visual fixations for the Irrelevant "NPC" object in the heat map and a 0% (R=100%) gaze for the Irrelevant objects. However, for the female-high cluster, the heat map results were not consistent, very light visual fixations (i.e., green indicator) for the Irrelevant "NPC" object in the heat-map was seen but had 0% (R=100%) gaze for the Irrelevant objects presented by the sub-sequence percentage results. Slightly different results showed between both analysis methods. Overall for both page *M1* and page *M4*, it can be observed that as the level of difficulty increased the more visual fixations were accumulated for the relevant objects.

As for Game *S*, the easy-task (page *S1*) in the DEG, the heat maps for page *S1* (Figure 6.7) were substantially consistent with the sub-sequence percentage results in Table 6.11. Visual fixations for the Irrelevant “NPC” object for female-high cluster were not observed in the heat maps (i.e. no green indicator) a similar result presented by the sub-sequence percentage results where there was 0% (i.e. Relevant (R) = 100%) gaze for the Irrelevant “NPC” object. Consistency was also observed for the other three clusters: male-low, male-high and female-low. Green visual fixations on the Irrelevant “NPC” object in the heat-map was observed for all 3 cluster groups; each cluster had 44% (R=56%), 55% (R=45%) and 74% (R=26%) gaze for Irrelevant objects, respectively presented by the sub-sequence percentage results.

R=44 (b) R=55 (c) R=74 (d) R=100

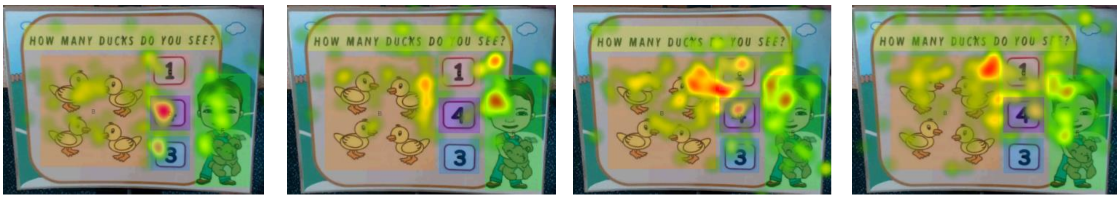
Figure 6.7: Gender vs. Achievement heatmaps for DEG Page *S1* (sorting): (a) Male-Low, (b) Male-High, (c) Female-Low, and (d) Female-High.

As for Game *S*, the hard-task (page *S4*) in the DEG, the heat maps for page *S4* (Figure 6.8) were partially consistent with the sub-sequence percentage results (Table 6.11). There were no visual fixations for the Irrelevant “NPC” object for both male-high and female-high children similar to the sub-sequence percentage results discussed earlier (Table 6.11), with having 0% (i.e., Relevant (R)=100%) gaze for the Irrelevant “NPC” objects. However male-low and female-low cluster heat map results were not consistent with the sub-sequence percentage results. The male-low cluster had no visual fixations for the Irrelevant “NPC” object on the heat map but had 56% (R=44%) gaze for the Irrelevant “NPC” object presented from the sub-sequence percentage results.

(a) R=44 (b) R=100 (c) R=100 (d) R=100

Figure 6.8: Gender vs. Achievement heatmaps for DEG Page *S4* (sorting): (a) Male-Low, (b) Male-High, (c) Female-Low, and (d) Female-High.

Contrariwise, the female-low cluster visualised greenish yellow areas on the heat map but had 0% (R=100%) gaze for the Irrelevant “NPC” object. This showed a slight inconsistency between both methods. However, this inconsistency may be because the sub-sequence percentage only revealed the very early phase of interaction on every page whereas the heat map interaction is the accumulative fixation from the entire session of every page/level. The sub-sequence percentage delivers information such as the percentage of objects that attracted the child when approaching a page. While the heat map only showed visible information (colour scheme) without any percentage of the objects fixated.



(a) R=13 (b) R= 21 (c) R=38 (d) R=69

Figure 6.9: Gender vs. Achievement heatmaps for Cardboard Game Page *M1* (matching): (a) Male-Low, (b) Male-High, (c) Female-Low, and (d) Female-High.

As for Game *M* of the easy-task (page *M1*) in the Cardboard game, comparison was made between the sub-sequence percentages results in Table 6.12 with the heat maps in Figure 6.9. The sub-sequence percentages and heat maps were considerably consistent. Overall, across all four cluster groups, the heat map showed every group had roughly looked at the Irrelevant “NPC” object, a similar result presented by the sub-sequence percentage results. However, accumulated fixations on the Irrelevant “NPC” object for female-high visualised small red areas on the heat map (indicating concentration), given that this cluster group showed only 31% (i.e., Relevant (R)=69%) gaze at the Irrelevant “NPC” objects as presented by the sub-sequence percentages results. In comparison, the male-low cluster showed a higher 87% (R=13%) gaze at the Irrelevant “NPC” objects but visualised no red (concentration) areas and only the greenish yellow indicator on the heat map. This showed different results between both analysis methods. Similar to the previous discussion for the DEG, the difference could be attributed to the fact that the sub-sequence percentage was derived from the early phase interaction while the heat map was derived from accumulative fixation that took place throughout a page session. However, the result differences from both methods may complement each other to understand better between relevant and irrelevant UI object in game design for 5-year-old children.

(a) R=33 (b) R=58 (c) R=38 (d) R=67

Figure 6.10: Gender vs. Achievement heatmaps for Cardboard Game Page *M4* (matching): (a) Male-Low, (b) Male-High, (c) Female-Low, and (d) Female-High.

As for Game *M* of the hard-task (page *M4*) in the Cardboard game, the heat maps for page *M4* (Figure 6.10) were consistent with the sub-sequence percentage results (Table 6.12). No noticeable difference was seen between the two methods. For instance, male-high achievement did not visualise any red area but only a greenish yellow area (no concentration) on the Irrelevant “NPC” in the heat map, a similar result showed by the sub-sequence percentage where 42% (i.e., R= 58%) gaze at Irrelevant object was presented.

(a) R=33 (b) R=100 (c) R=38 (d) R=69

Figure 6.11: Gender vs. Achievement heatmaps for Cardboard Game Page *S1* (sorting): (a) Male-Low, (b) Male-High, (c) Female-Low, and (d) Female-High.

As for Game *S* of the easy-task (page *S1*) in the Cardboard game, the results between the two analysis methods, sub-sequence percentage (Table 6.12) and heat map (Figure 6.11) were not consistent. For the male-high cluster, the sub-sequence percentages result presented a 0% (i.e., R= 100%) gaze at the Irrelevant "NPC" object. However, from the heat map, accumulated visual fixations (i.e., the indicator is green yellowish) can be seen on the irrelevant "NPC" object, indicating that children did gaze at the "NPC" but possibly later in the interaction. A similar situation happened to other clusters showing no consistency.

As for Game *S* of the hard-task (page *S4*) in the Cardboard game, the heat maps for page *S4* (Figure 6.12) were considerably consistent with the sub-sequence percentages results (Table 6.12). No noticeable difference was seen between the two methods. This may be because of the low sub-sequence percentage rate results making it difficult to compare in detail with the visualisation data (heat map). However, the differences in the difficulty of the game may also be a factor to the low percentage of relevant objects.

(a) R=21 (b) R=58 (c) R=38 (d) R=38

Figure 6.12: Gender vs. Achievement heatmaps for Cardboard Game Page *S4* (sorting): (a) Male-Low, (b) Male-High, (c) Female-Low, and (d) Female-High.

Overall, the findings show that the two analysis methods that were employed - sub-sequence percentages analysis and heat map analysis - were partially consistent. Nonetheless, the analyses can indeed complement each other for understanding the interaction strategy used by children. Heat maps (accumulated visual fixations) enable us to infer where the children focused their attention over a specific duration whereas sub-sequence percentage provides information on children's attention to specific objects (i.e. relevant and irrelevant objects). In addition, the sub-sequence matrices provide information on possible strategies used by children while interacting with the game.

6.4 Chapter Summary

In analysing the Learning Effect on both game learning media, a statistically significant improvement in learning was found through the Cardboard game but no statistically significant improvement was found for the DEG. There was a partial statistical significant difference in Learning Effect based on Gender and School Attended between the two game learning media. The statistically significant differences on Gender and School Attended were both observed for the Cardboard game participants. Table 6.13 presents the hypotheses outcomes of the Fundamental Analyses (Non-eye-tracking data).

Table 6.13: Results for hypotheses of the Fundamental Analysis (Non-eye-tracking data).

No	Hypotheses: Fundamental	Result
Evaluation Material and Learning Effect		
1a	*There is a statistically significant correlation between (i) the DEG in-game scores and pre-test; (ii) the DEG in-game and post-test scores; (iii) the Cardboard game in-game and pre-test scores; (iv) the Cardboard game in-game and post-test scores.*	Accepted
1b	*There is a statistically significant learning effect of (i) the DEG and (ii) the Cardboard game.*	partially accepted
1c	*There is a statistically significant difference in the learning effect between the DEG and Cardboard game.*	rejected
Gender, Attended School and Game Experience on Learning Effect		
2.1	*There is a statistically significant correlation between Gender and Achievement Level for (i) the DEG participants; (ii) the Cardboard game participants.*	Rejected
2.2	*There is a statistically significant difference in the learning effect between Genders of (i) the DEG participants; (ii) the Cardboard game participants*	partially accepted
2.3	*There is a statistically significant difference in the learning effect between both schools of (i) the DEG group; (ii) the Cardboard group*	partially accepted
2.4a	*For the DEG, there is a statistically significant correlation between the learning effect and Game Experience for (i) Game M; (ii) Game S.*	partially accepted
2.4b	*For the Cardboard game, there is a statistically significant correlation between the learning effect and Game Experience for (i) Game M; (ii) Game S.*	partially accepted

The data analysis Method 1 and Method 2 involved synchronic (focused attention) and diachronic (gaze sequence) indicators in eye-tracking methodology. In Method 1, the Focused Attention on relevant and irrelevant UI objects between both game learning media had statistically moderate significant differences between high and low achievers. For the DEG, high achievers paid attention to the relevant objects in Game *M* and demonstrated gameplay strategies (e.g., paying less attention as indicated by a shorter fixation duration to the object *B* “train”) in Game *S*. This contrasted with their counterparts, who were distracted by the eye-catching moving train. As for the Cardboard game, lower achieving children spent more time looking at the cardboard Game *M* due to recounting the number of objects repeatedly or were entirely lost (no strategy). A similar result was observed for Game *S*. The relation between the recall-based task and the recognition-based task proved that the recall task was more difficult than recognition for both game learning media, thus confirming that recall entails a deeper level of information retrieval. As an overview, Table 6.14 presents the outcomes of all hypotheses in relation to Method 1.

In analysing the Interaction Strategy, there was no difference of the Gaze Sub-Sequence Scores between high and low Performance Level for both game learning media. No correlation was also found between the Gaze Sub-Sequence Scores and Learning Effect. As for Gender and Gaze Sub-Sequence Scores, both genders in this age applied the similar interaction style when playing the game. Nonetheless, there was a statistically significant difference in the Gaze Sub-Sequence Scores based on the Achievement Level of both Game *M* and Game *S* in the DEG but not in the Cardboard game participants. In summary, Table 6.15 presents all hypotheses and results in relation to Method 2.

Exploring further, the gaze sub-sequence analysis helped to understand the strategy used among young children when interacting with the game in the beginning of every level/page. The sub-sequence matrices, which were used to analyse the strategies for both the DEG (Figure 6.1 and Figure 6.2) and Cardboard (Figure 6.3 and Figure 6.4), showed that male and female high achievers applied different learning strategies (cognitive style) as compared to their low achieving counterparts. From the sub-sequence percentage result, female participants showed a higher percentage of gazing at relevant UI objects as compared to their male counterparts for both Game *M* and Game *S*. In addition to the gaze sub-sequence analysis, the results from two eye-tracking analysis methods - sub-sequence percentages (gaze sub-sequence analysis) and heat maps (visualisation data) – showed that both methods complement each other by contributing different information for understanding children’s gazing behaviour.

Table 6.14: Results for hypotheses in study Method 1

No	Hypotheses: Method 1	Results
Focused Attention		
3a	*For the DEG, focused attention is a statistically significant predictor of the learning effect for each of the four difficulty levels of (i) Game M; (ii) Game S*	partially accepted
3b	*For the Cardboard game, focused attention is a statistically significant predictor of the learning effect for each of the four levels of (i) Game M and (ii) Game S.*	rejected
Focused Attention towards UI Objects		
4a	*There is a statistically significant difference between high and low achievers in focused attention given to the relevant objects of the DEG Game M (i) level 1; (ii) level 2; (iii) level 3; (iv) level 4.*	partially accepted
4b	*There is a statistically significant difference between high and low achievers in focused attention given to the relevant objects of the DEG Game S (i) level 1; (ii) level 2; (iii) level 3; (iv) level 4.*	partially accepted
4c	*There is a statistically significant difference between high and low achievers in focused attention given to the relevant objects of the Cardboard Game M (i) level 1; (ii) level 2; (iii) level 3; (iv) level 4.*	partially accepted
4d	*There is a statistically significant difference between high and low achievers in focused attention given to the relevant objects of the Cardboard Game 2 (i) level 1; (ii) level 2; (iii) level 3; (iv) level 4.*	partially accepted
Recall and Recognition		
5a	*There is a statistically significant difference in the in-game scores between the easy level tasks of Recognition and Recall for both the DEG and Cardboard game.*	accepted
5b	*There is a statistically significant difference in the in-game scores between the hard level tasks of Recognition and Recall for both the DEG and Cardboard game.*	accepted

Table 6.15: Results for hypotheses in study Method 2

No	Hypotheses: Method 2	Result
Gaze Sub-sequence Score based on Learning Effect		
6.1a	*For the DEG, there is a statistically significant difference in the Gaze Sub-Sequence Score between the high and low performers in terms of Learning Effect for (i) Game M; (ii) Game S.*	rejected
6.1b	*For the Cardboard game, there is a statistically significant difference in the Gaze Sub-Sequence Score between the high and low performers in terms of Learning Effect for (i) Game M; (ii) Game S.*	rejected
6.1c	*For the DEG, there is a statistically significant correlation between the learning effect and Gaze Sub-Sequence Score for (i) Game M; (ii) Game S.*	rejected
6.1d	*For the Cardboard game, there is a statistically significant correlation between the learning effect and Gaze Sub-Sequence Score for (i) Game M; (ii) Game S.*	rejected
Gaze Sub-sequence Score on Gender		
6.2a	*For the DEG, there is a statistically significant difference in the Gaze Sub-Sequence Score between Genders for (i) Game M; (ii) Game S.*	rejected
6.2b	*For the Cardboard game, there is a statistically significant difference in the Gaze Sub-Sequence Score between Genders for (i) Game M; (ii) Game S.*	rejected
Gaze Sub-sequence Score based on Achievement Level		
6.3a	*For the DEG, there is a statistically significant difference in the Gaze Sub-Sequence Score between the two groups with high and low Achievement Level (Prior Knowledge) for (i) Game M; (ii) Game S.*	accepted
6.3b	*For the Cardboard game, there is a statistically significant difference in the Gaze Sub-sequence Score between the two groups with high and low Achievement Level (Prior Knowledge) for (i) Game M; (ii) Game S.*	rejected

Chapter 7: Lessons Learned

This chapter presents and discusses the lessons learned throughout the entire research project. The practical, methodological and theoretical challenges encountered are pointed out with improvement suggestions for future studies. The implications of game design for 5-year-old children are also presented in this chapter. Finally, the limitations that had influenced the results or restricted the process of the empirical research are explained.

7.1 Challenges

Challenges are the difficulties encountered during the implementation of the study. Challenges that occurred in each step of the empirical study process are divided into practical, methodological and theoretical. This section presents the challenges encountered throughout this study and their implications.

7.1.1 Practical

Recruiting participants. Recruiting schools and young children for the Main Study was very challenging. Administratively, it was difficult to get schools involved since taking individual children out of regular classes for 25 to 30 minutes are usually not appreciated as it may affect their learning. Additionally, the time constraint imposed by the school term calendar impedes schools from accommodating research studies that may take 2 to 3 weeks to complete, especially when participating schools had to prepare and provide resources (e.g., a separate room) to support the empirical study. Hence, for future research work these issues on recruitment should be taken into account and other strategies could be planned in advance, for instance, preparing some budget and time for travelling beyond a local city for data collection.

Mobile Device Stand (MDS) Design. From the practical experience of the Main Study, the MDS design could be considered a good design to study mobile applications on tablets for adults. However, the MDS design is not suitable for studies involving young children that have petite body features. Modification has been made to allow the participating young children to reach the tablet in the Main Study. For the design of eye-tracking devices, especially for studying mobile applications and games, a more children-friendly

MDS and similar hardware should be carefully evaluated. This is especially relevant as the number of promising game applications for young children is growing. Furthermore, Participatory Design (PD) methods are not only applicable to software but also, if not more important, hardware technology.

Eye-tracking Device on Young Children. Eye-tracking technology is known to support studies related to focused attention (Lai et al., 2013) and cognitive processing (Alemdag & Cagiltay, 2018; Eraslan et al., 2016) among adult participants. However, implementing the eye-tracking methodology on young children can be very challenging and time-consuming. Preparations concerning the calibration process, sitting behaviour, configuration (eye-tracker distance) and experiment room need to be thoroughly tested in advance. Preparation such as lightings in the experiment room had to be identified before setting up the experiment, making sure that the reflection does not interfere with the eye-tracking recording or the distance the child sits within the eye-tracker's ideal range needs to be identified before the experiment. The most challenging part with young children was the calibration process. 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. Details of other aspects of preparing eye-tracking studies among young children can be referred to in Section 4.1.1.

7.1.2 Methodology

Game engagement versus loss of data. The use of the MDS (Section 3.3.2) with young children could be limited by a trade-off between loss of data and gameplay engagement. Despite instructions of keeping a certain distance from the tablet, children tended to lean forward towards it; a posture enabling them to engage in the gameplay. However, by allowing the child to lean forward may cause the eye-tracker recording to be out of range (60 to 65cm to enable maximum gaze angle, Figure 3.11) and can cause loss of data. Contrarily, giving repeated reminders to sit appropriately to a child during the gameplay might irritate them for being nagged and controlled, causing them to lose interest in the game. Hence, for future research this issue has to be considered in advance, such as applying another type of eye-tracking device on young children as well as specifically designed furniture for studying the game base learning effect with focused attention.

Facial expression approach. Although questionnaires for evaluating game experience with children exist (Read & MacFarlane, 2006), using the questionnaires among 5-year-old children who have limited reading abilities (McIntyre et al., 2006) may invalidate the questionnaire approach. On the other hand, applying the interview approach was not applicable either because it may prolong the experiment duration, which was not possible due to the school's policy for restricting the time that children can be taken out from their classroom individually. This inflexible timing was because children might miss out certain lessons or exercise if they were taken out of the classroom beyond 30 minutes. Besides children at this age whose verbal ability in expressing themselves, especially describing emotions, are quite limited (Parker et al., 2013). Therefore, analysing facial expressions was the feasible approach to study the game experience in this study.

Gaze sequence analysis challenge. According to Eraslan and his colleagues (2016), most scanpath analysis software applications on the Internet are inaccessible and need to be purchased. In this study the use of *eyePatterns* was only limited to extracting data from the extended sequence form (fixation sequence) to collapsed sequence (gaze sequence) form and did not use the software for pattern identification. This is because *eyePatterns* sometimes detects very short sequences that may not be meaningful for this study. Another challenge for using *eyePatterns* is its inadequate documentation. It is not easy for a user to understand the software program by itself. The documentation provided is too generic without enough details to enable users to apply it for operating the software tool. In this study the gaze analysis was limited to the use of collapsed sequence. This allowed us to perform a gaze sequence analysis in the order of AOIs that participants looked at, thereby inferring their learning strategies. In future work, the application of fixation sequences in inferring learning behaviours will be looked into.

Heat map and scanpath challenge. In general, heat maps are less time-consuming to achieve since the eye-tracker software can automatically analyse the visualisation data. However, a heat map only reveals areas being viewed, visualising the relative fixation durations by colours with red indicating the most intensely gazed at spots and light green the least (or no colour where it is not looked at). As scanpath 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. However, scanpath data can reveal how a participant views the objects in order, making the analysis more valuable. Given the goal of gaining insights into how children interact with both game learning media thereby deriving their learning strategies, the efforts are worthwhile.

7.1.3 Theoretical

Lack of established publications on attention span. As mentioned in Section 2.5, some grey literature has discussed the decrease of attention span among young children due to the use of digital technology (Patel, 2017; Perles, 2013). However, there have been hardly any empirical studies in the recent years on children's focused attention span. While the results on the total fixation durations (Table 6.9) could be considered as a proxy of estimated focused attention of 5-years-old, the Main Study, unlike (Ruff et al., 1998) was not specifically designed to measure focused attention and the related metrics. The precision of the rough estimate could have been improved with different game designs (e.g., a broader range of game scenarios with more levels), a more sophisticated experimental setup (e.g., high resolution cameras focusing on the child and the room) to allow collection of more relevant data and with the systematic manipulation of the influencing factors (e.g. the interaction with the experimenter). Overall, this is a call for more future research work to be conducted to study attention span in young children. Empirical findings thus obtained will serve as a valuable reference for the research community.

Game-based Learning framework. There are a number of game-based learning models that can be used to develop educational games (All, Nuñez Castellar, & Van Looy, 2016; Zin, Jaafar, & Yue, 2009). The Shi & Shih (2015) GBL model applied in this study provided a useful and clear framework in defining factors that could be considered in designing an entertaining game. However, the challenge was defining the learning component in the DEG for the Main Study. The main purpose of the applied framework was to assist designers to build fun games without taking into account the learning factor; it assumed that learning would occur when a player's engagement in the game increased. In a DEG, the learning factor is an important component and one of the main focuses in this study. Therefore, additional references that included learning content (e.g. Simões, Redondo, & Vilas, 2013) as a factor for educational game design and existing educational games were reviewed to gain insights for developing the dedicated DEG for this study.

7.2 Implications for Game Design involving 5-year-old children

Implications derived from the results and observations of the Main Study are presented in this section.

Game Duration. Although the DEG was tested and improved from the pilot studies, the game design was not a good differentiator to examine the learning effect in the Main Study due to the short duration of the game. The game design was developed based on the literature of young children's attention span. However, from results and observations of the Main Study, the short attention span of young children derived from previous studies taken place in non-technological contexts appears no longer valid; interactive technologies presumably can engage users, young as well as older ones, for a longer 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 the probably out-dated literature on children attention span. However, it should not be too long either because of the potential fatigue effect.

Game Difficulty. According to the Early Years Foundation Stage (EYFS) outcome evaluation (Department for Education, 2013) a child by the end of 5 years old should be capable of counting and sorting numbers from 1 to 20. However, from the Main Study observation, the DEG for 5-year-old children game content and difficulty level should not be limited by the EYFS evaluation outcome. The difficulty level should at least go beyond the number 20 to avoid the ceiling effect of these young children which have different learning paces. Nevertheless, the number of levels in the game should also be added to enable the learning effect to be differentiated.

Instruction. The gaze sub-sequence in the Main Study helped identify which objects attracted children as they entered each page of the game. Typically, when a person plays a game, he or she reads the instruction before doing anything else. However, from the gaze sub-sequence result, a number of children tended to not look at the instruction text (AOI A) as they entered the game pages (Figure 6.1 and Figure 6.2). This behaviour could possibly be because the instruction text did not play a role in the game due to the children's low reading ability. Instead, the children listened to the audio instruction. Therefore, game designers could drop the instruction text that could occupy much of the real estate of the game, replacing it with an object that could be more engaging to young children. In addition, it is important to have audio instruction as it plays an important role in a game design for young children who have limited reading ability.

Non-Player Character (NPC). From the gaze-sub-sequence results, a number of male (Figure 6.2) and low achieving (Figure 6.1) children were distracted by the NPC as they entered each page. This distraction could probably indicate that the children were expecting something from the NPC, which could have been made as a suitable means to deliver the instruction. Therefore, instead of static NPCs, game designers for young children could create interactive NPCs that can speak and tell children what to do next, creating a more enjoyable and attractive environment.

7.3 Limitations

Besides the challenges discussed earlier, there were limitations that affected the results of the study as follow:

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 the lack of literature and empirical studies of digital educational games on numeracy for 5-years-old. Therefore, the findings of this study may only be generalizable for DEGs on arithmetic/numeracy for 5-years-old children in the UK.

Lack of Prior Knowledge data of the young children. The Prior Knowledge in the Main Study was limited to the Achievement Level (i.e. the pre-test result) and had no additional data that can represent the children's knowledge prior to participating in the experiment. At such a young age no formal assessment was available and applied in the reception or 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 with the researcher. Therefore, the Achievement Level was used as the measurement for children's Prior Knowledge in the Main Study.

Device for data collection may have limitation. The screen mounted eye-tracker device used in this study may have limitations, especially on young children who typically can only sit still for a short period of time. However, the decision to use this kind of eye-tracker was to minimise the intrusiveness of the measuring device to be used with young children such as a wearable eye-tracker, which can cause discomfort or distress to this vulnerable group due to the device appearance. In addition, parents or guardians of young children may not give their consent for their child to participate in experiments involving intrusive devices due to worries caused by the lack of understanding of the devices.

Reliance on quantitative method. The eye-tracking data such as Focused Attention and Gaze Sub-Sequences Scores presented in numeric forms led to the reliance on quantitative data in this study. Nonetheless, as the interview approach was not applied due to the low verbal ability of young children to express themselves, qualitative data analyses were limited in the Main Study.

Statistically non-significant Learning Effect. Learning Effect in the Main Study was defined as the performance gain after the children played the game in one of the learning media. However, the statistically non-significant Learning Effect of the DEG on both game learning media could be attributed to a few factors. The celling effect of the test material (i.e. pre and post-test) makes it improbable to demonstrate any change in the dependent variable. 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. Nonetheless, the design of the Main Study was based on the EYFS curriculum standard; accordingly children should be able to count from 1 to 20 by the end of 5 years of age (Department for Education, 2013). 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.

Statistically non-significant results related to gender. Studying the difference in the learning effect between two genders and between two game learning media is intriguing. It could give insights into customised games or gender-based games design. However, the non-significant statistical results on gender from the Main Study can also contribute to other insights. The results eradicate the stereotype (Arroyo et al., 2013; Sullivan &

Bers, 2016) that one gender is better than the other in terms of learning via digital educational games. The results showed that young children did not have any inherent differences in applying interaction strategy in learning arithmetic via games. It can be inferred that 5-year-olds, irrespective of gender, may have the comparable capacity for learning through a game. Therefore, boys and girls of this age can be trained in the same way without any bias.

7.4 Chapter Summary

The challenges as well as implications of game design for 5-year-old children and the limitations of the study were discussed in this chapter. The practical challenges were related to device appropriateness and participant recruitment. Specifically, the eye-tracker used was not child-friendly and requesting schools to allocate their space and time for getting involved in the research study was difficult. The methodological challenges were related to data availability and refining the existing gaze sequence analysis methods. Theoretical challenges were due to the lack of relevant literature such as the up-to-date references on attention spans of young children and a limited number of validated game-based learning models on DEGs for young children. Implications for game design for 5-year-old children include game duration, game difficulty and voice-based instructions and interactivity of non-player character. Finally, the limitations with regard to restriction to a particular age group, unavailable prior knowledge, non-child-friendly eye-tracker, reliance on quantitative data and gender stereotypes were discussed.

Chapter 8: Conclusion

This chapter summarises how each research question in this thesis was addressed, discusses the contributions of this research project and finally concludes with possible future work that can be explored. Figure 8.1 illustrates a brief overview of each research question (RQ1 to RQ6) and how they were approached for each Analysis Method (Section 6.1, Section 6.2 and Section 6.3) in this thesis.

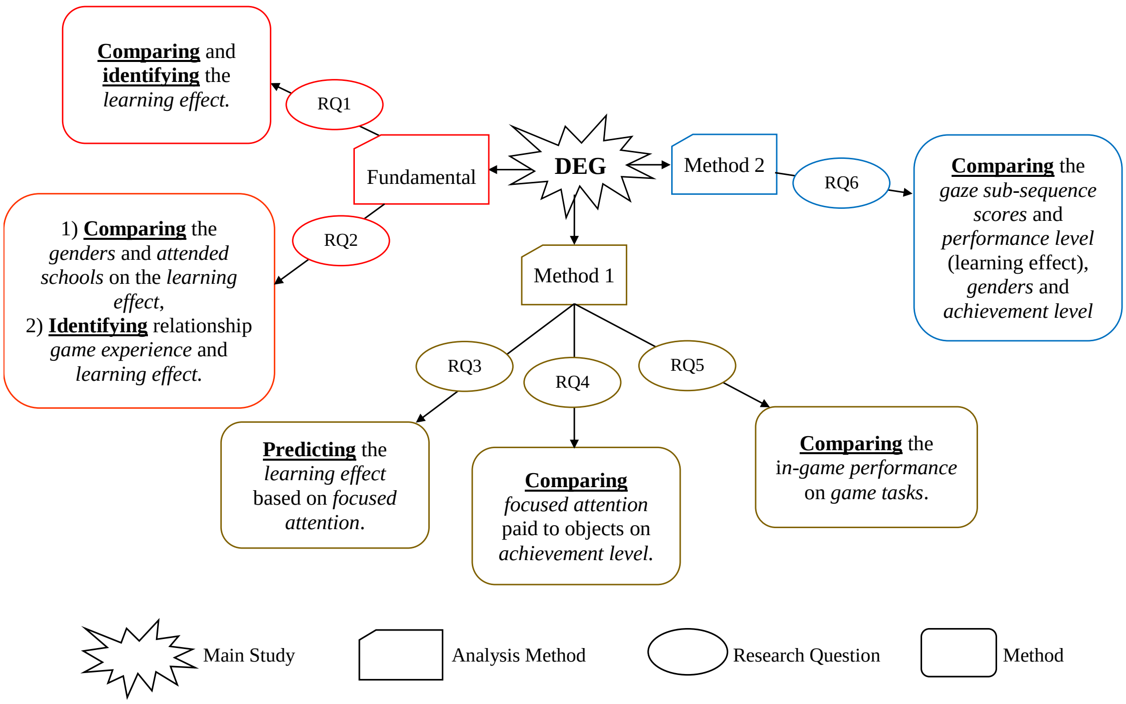


Figure 8.1: A brief overview of the RQs and their methods in this thesis. Red labels are related to the Fundamental Analysis, Brown label related to Method 1 Analysis and Blue labels are related to Method 2 Analysis.

8.1 Research Questions Concluded

Each research question is revisited and summarized as below:

RQ1: To what extent does the learning effect *of a dedicated DEG on numeracy differ from its cardboard version?*

RQ1 was to assess the Learning Effect between both game learning media, DEG and Cardboard game. RQ1 was addressed by comparing and identifying the Learning Effect obtained by calculating the differences between the pre and post-test. The measurement of the Learning Effect in this thesis was implemented as: 1) Difference between the pre and post-test and 2) In-game scores. The relationship between the learning measures was first evaluated to make sure that the knowledge applied was consistent across the evaluation conditions. The results showed that there was no difference in the Learning Effect between both game learning media which could be contributed by the sample size between both game learning media. However, the Cardboard game participants gained improvement in learning while the DEG ones did not demonstrate any statistically significant gain by playing the games. The main possibility could be because of the typical playgroup learning environment in schools and exposure to DEGs that the children had experienced before taking part in the study. Related results and hypotheses were discussed in Section 6.1.2 and Section 6.1.3.

RQ 2: How can children's attributes and game experience help understand the learning effect induced by both game learning media?

The objective of RQ2 was to explore further the influence of Gender, Attended School and Game Experience on the Learning Effect. RQ2 was addressed by 1) Comparing the difference between the Genders and Attended Schools on the Learning Effect, 2) Identifying the relationship between the Game Experience on the Learning Effect (Section 6.1.4). For the DEG no statistically significant difference in Learning Effect was found between boys and girls (Buffum et al., 2015) while for the Cardboard game there was a statistically significant difference. Girls perform better in the hands-on learning environment (Fails et al., 2005) probably because that girls among 5-year-olds had better fine motor skills compared to boys (T. Moser & Reikerås, 2016). For Attended School, no difference between School 1 and School 2 on the Learning Effect for the DEG was found. However, a difference was found for the Cardboard game which was probably because most children in School 2 were from low-medium income family backgrounds with limited access to DEGs, but they were probably familiar with hands-on experience with cardboard games in general.

The game experience on learning effect for both DEG and Cardboard game were partially correlated by the type of game tasks (Game M and Game S). Game M tasks (recognition-based task) were preferably played through a physical environment while Game S tasks (recall-based task) were preferably played through a digital game-based learning environment. This could be caused by the attractiveness and flexibility of the game tasks. Game *M* of the Cardboard involved more hands-on interaction than its digital counterpart and thus more attractive. On the other hand, Game *S* of the DEG was attractive because of the animated train and flexible boxes.

RQ3: How can children's focused attention help understand the learning effect induced by the two game learning media?

The rationale underpinning RQ3 was to identify how Focused Attention could help define the Learning Effect on both the DEG and Cardboard game. RQ3 was approached by predicting the Learning Effect based on the Focused Attention given on each page of the games on both learning media. Focused Attention refers to the total fixation duration of each page retrieved from the eye-tracker software. Only page *M2* and *M3* of the DEG had a statistically significant predictive relationship between the Focused Attention and Learning Effect. All pages of the Cardboard game for both Game *M* and Game *S* had no statistically significant predictive relationship with the Learning Effect. These findings suggested that Focused Attention might not be a good predictor for understanding the Learning Effect for both game learning media. Nonetheless, the relatively simple game designs could account for the statistically non-significant relationships observed. The related results and discussions were discussed in Section 6.2.2.

RQ4: How is the achievement level of children related to their focused attention (i.e. fixation duration) to relevant and irrelevant objects in the games?

The purpose of RQ4 was to explore the relationship between Focused Attention on UI objects and the Achievement Level of the participants. RQ4 was addressed by comparing the Focused Attention paid to relevant and irrelevant objects between high and low achievers on the game tasks of both the DEG and Cardboard game. For the two game learning media, the results revealed that Focused Attention paid to the relevant object between high and low achievers were partially different across the four pages of Game *M* and Game *S*. Both high and low achievers had their individual preferences for exploring relevant objects based on the type of game task. The related results and hypotheses of this RQ were discussed in Section 6.2.3.

RQ5: Are learning tasks drawing on recall more difficult than those on recognition for young children for both game learning media?

The rationale of RQ5 was to examine the effect of the difficulty level between the recognition-based tasks and recall-based tasks. RQ5 was addressed by comparing the in-game performance of both recognition-based (easy) task and the recall-based (hard) task. The results revealed (Section 6.2.4) that the recall based- tasks were harder for the children to answer as compared to the recognition-based task. This finding affirms Baddeley's (1997) theoretical model in human memory.

RQ6: How are children's interaction strategies derived from gaze sequence analyses related to the learning effect induced by both game learning media?

The purpose of RQ6 was to evaluate the gaze sub-sequence scores (Interaction Strategy) based on the Learning Effect, Gender and Achievement Level for both game learning media. RQ6 was addressed by comparing the difference in the Gaze Sub-Sequence Scores between the performance level (Learning Effect), Genders and Achievement Level. In comparing the Interaction Strategy based on the performance level (Learning Effect), no statistically significant difference was found between the high and low performance on both game tasks of the DEG and Cardboard game. Additionally, no relationship was found between the Interaction Strategy and the Learning Effect. This no statistically significant difference and no relationship was probably because the game design and ceiling effect (Section 6.3.2).

In comparing the Interaction Strategy based on Gender, no statistically significant difference was found between boys and girls on the game tasks of both the DEG and Cardboard game. The possible reason was that both boys and girls do not have any inherent difference in interaction strategy from the mathematical perspective. Therefore both genders are potentially capable of applying an interaction strategy while playing a DEG on numeracy at a young age (Section 6.3.3).
As for the interaction strategy based on the achievement level, no statistically significant difference was found between high and low achievers on the game tasks of both the DEG and Cardboard game. However, a statistically significant difference was observed between the achievement levels for the DEG. A possible reason was that children's

achievement level could influence their interaction strategy. The statistically non-significant difference results for the Cardboard game were because children's concentration could be distracted by the concrete objects and thus compromised their attention to develop interaction strategies (Section 6.3.4).

8.2 Contribution

This research project contributed to three main general body of knowledge; Human-Computer Interaction, Game-Based Learning and Educational Technology. Under the HCI field, applying the eye-tracing methodology among young children bridged the gap of insufficient work of eye-tracking with young children. A modified calibration board that was recognizable among young children during the empirical experiment with eye-tracking (practical) was created and available for used in future research work. The strengths and limitations of eye-tracking methodology on 5-year-olds were learned and improvement suggestion have been reflected in Section 7.1.1. In analysing eye-tracking data, a modified Gaze-Subsequence marking scheme to learn children's interaction strategy in playing the game was contributed.

As contribution to the Game-based Learning, studying the learning effect induced by playing a digital game helped to learn the different influences of children's attributes and consequently gave a deeper understanding of the role and design of such games elements (e.g.; NPC and Instructions). Moreover, the DEG design based on game-based learning not only validated the GBL framework but also refined children's focused attention on digital games.

As to Educational Technology, studying the interaction strategy in digital game contributed to the usage of digital games among young children in education. Analysis of focused attention among 5-year-olds on DEGs informed the present attention span of children with technology.

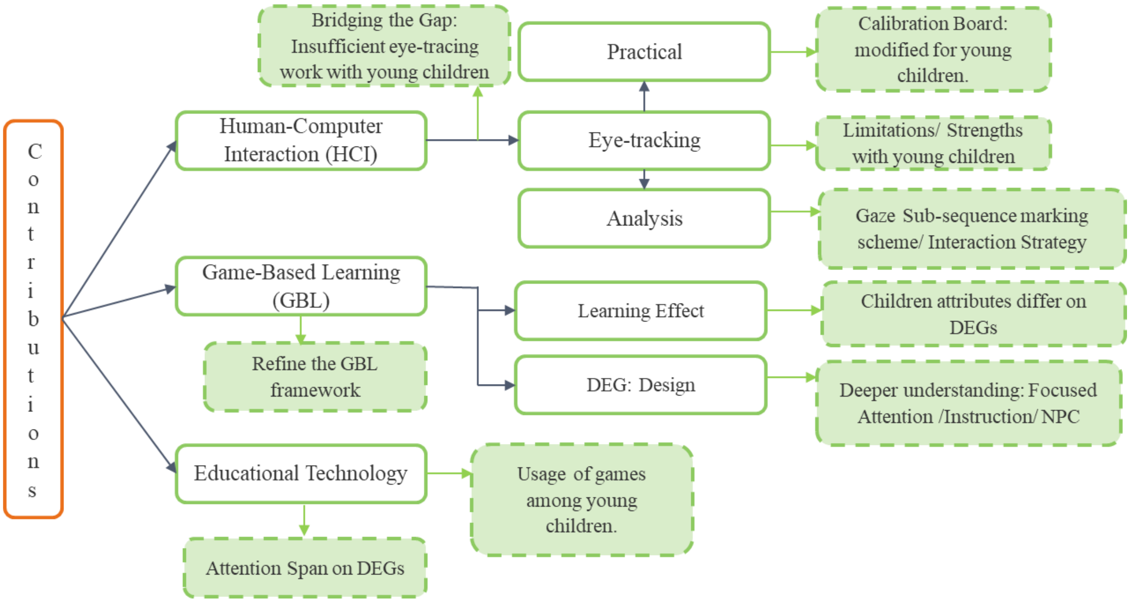


Figure 8.2: Contributions of this Research Project.

8.3 Future work

Based from the results and the lesson learned, the outlook for future work on this specific area is discussed in the following.

Sample with different backgrounds. The Main Study only involved 5-year-olds from two different schools in Leicestershire. Recruitment of participants could be done beyond the local city and with schools of different academic performance levels and backgrounds. Samples with different background and levels may generalise the findings and gain better understanding of learning on DEGs.

Apply wearable eye-trackers. As technology improves rapidly and parents become more aware of eye-tracking and wearable devices, they are thus more likely to give consent for their children to take part in a study where they will be asked to wear child-friendly mobile eye-trackers. Screen mounted eye-tracker device can still be used. However, the MDS frame that supports the DEG or Cardboard game together with the eye-tracker needs to be customised for young children. Other than that, the eye-tracking device manufacturers should search for better ways to optimise the calibration process for young participants as there are growing numbers of mobile device users among young children (Mayer, 2017).

Improved Game Design. The game designed for this study was relatively short as to accommodate the short attention span of young children as a requirement. However, the notion needs to be adapted in order to study the learning effect of the two game learning media. Specifically, a reasonable time-based game with more game levels, which suit children with different attention span and numeracy, should be developed to study the learning effect of game learning media for future work.

Focus study on Game Experience. The main scope of this study was to study the learning effect of both game learning media via the eye-tracking data and the covariates that may influence the learning. The next phase of this study could focus Game Experience on DEGs via eye-tracking data. Some existing research use the pupil size derived from the eye-tracking data to recognize a person's emotion (Oliva & Anikin, 2018). However, it is an emerging research and no consistent findings or directions are concluded at this moment, implying a call for future work. Furthermore, other scanpath measures such as 'fixation sequence' or longer sub-sequences (Eraslan et al., 2016) can be applied to study the 'gaze sequence'.

8.4 Final Remarks

The goal of this research was to evaluate the learning effects of DEGs and the applicability of eye-tracking methodology of 5-year-old children. Outcome of this study contributed to the general body knowledge of Human-Computer Interaction, Game-Based Learning and Educational Technology. The six research questions addressed and the corresponding empirical findings reported have laid some valuable groundwork for future research work, given the increasing use of DEG among young children. The lessons learned could inform future researchers of eye-tracking methodology on young children with useful recommendations when implementing a similar research design. Given the challenges in recruiting participants in this study, it is recommended more support from schools and parents through systematic outreach activities to convince them the potential impact of research activities in the long run, should be sought. A child-friendly eye-tracker and mobile device stand (MDS) may encourage further insight into DEG through focused attention and gaze sequences. Nonetheless, a practical game design for 5-year-old children needs to take into account factors such as game duration, game difficulty, audio /text instructions and interactive NPC in the development.

Appendix

A.1 DBS and Ethics Approval

UNIVERSITY OF LEICESTER

University Ethics Sub-Committee for Science and Engineering and Arts Humanities

25/01/2016

Ethics Reference: 2907-dmn9-computerscience

TO:
Name of Researcher Applicant: Dinna Mohd Nizam
Department: Computer Science
Research Project Title: Digital Educational Games on Numeracy for Preschool Children

Dear Dinna Mohd Nizam,

RE: Ethics review of Research Study application

The University Ethics Sub-Committee for Science and Engineering and Arts Humanities has reviewed and discussed the above application.

1. Ethical opinion

The Sub-Committee grants ethical approval to the above research project on the basis described in the application form and supporting documentation, subject to the conditions specified below.

2. Summary of ethics review discussion

The Committee noted the following issues:
Approved

3. General conditions of the ethical approval

The ethics approval is subject to the following general conditions being met prior to the start of the project:

As the Principal Investigator, you are expected to deliver the research project in accordance with the University's policies and procedures, which includes the University's Research Code of Conduct and the University's Research Ethics Policy.

If relevant, management permission or approval (gate keeper role) must be obtained from host organisation prior to the start of the study at the site concerned.

4. Reporting requirements after ethical approval

You are expected to notify the Sub-Committee about:

- Significant amendments to the project
- Serious breaches of the protocol
- Annual progress reports
- Notifying the end of the study

5. Use of application information

Details from your ethics application will be stored on the University Ethics Online System. With your permission, the Sub-Committee may wish to use parts of the application in an anonymised format for training or sharing best practice. Please let me know if you do not want the application details to be used in this manner.

Best wishes for the success of this research project.

Yours sincerely,

Prof. Paul Cullis
Chair

A.2 Study Information Leaflet (Front Page)

The Game...

The game developed for this project is divided into TWO parts:

Game A: Matching
The child needs to count the total number of objects and match it with the correct number out of the three options given.

Game B: Sorting
The child needs to drag the numbers to the correct boxes on the train below. The numbers have to be in the right sequence.

Each part consists of 4 levels of activity.
Below are screen examples of the game.

Data Protection...

All information gathered during the research will be used anonymously. Children's data will only be released by the researcher if any safeguarding issue arises. All information recorded and provided is protected by the Data Protection Act. The research project has been approved by the University's Research Ethics Committee.

Your Contribution...

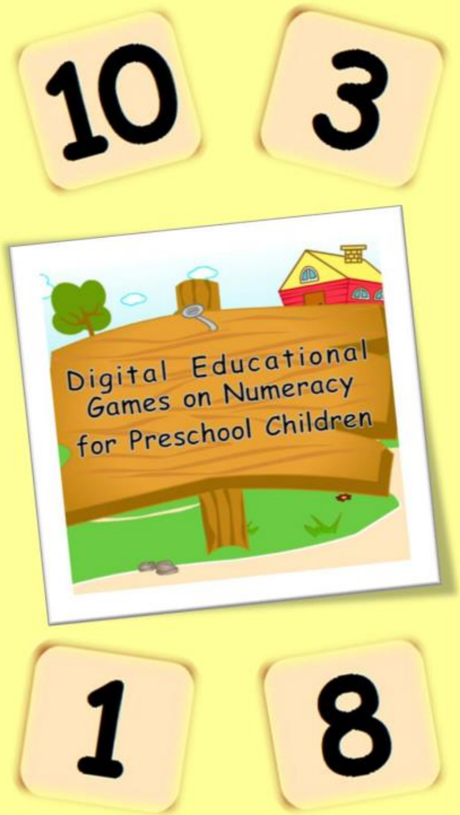
- Gain insights into research on technology with young children.
- Contribute to the better development of educational games for schools and at home.
- Take part in helping the society to understand the usage of technology with young children.
- Share your views on the use of technology in supporting young children to learn.

Find out More...

Please do not hesitate to contact us at:

Dinna Mohd Nizam (Researcher)
Email: dmn9@le.ac.uk
Phone: 07873914082

Dr Effie Law (Research Ethics Officer and Supervisor)
Email: lcl9@le.ac.uk



STUDY INFORMATION LEAFLET

UNIVERSITY OF LEICESTER

A.3 Study Information Leaflet (Back Page)

Introduction...

Educational game-based mobile applications are becoming more commonly used by children in learning. Insights into young children's learning process during a game session can contribute to improving digital education games (DEGs) in school and at home. This assumption has driven a research project being run at the Department Of Computer Science, University of Leicester.

About the Project...

What is the study about?

This study is about learning on how young children aged 5 learn numeracy with digital objects presented on a tablet and physical objects.

Who can take part?

Young children who just entered 5 years old are invited to participate in this project. Parents are encouraged to join their young child.

Where will it be?

It should take place in a room with least distraction in your child's school.

How long will it take?

A single session should take about 20-30 minutes to complete.

How it will be done...

In a single session, a child will be asked to play an educational game (tablet- or cardboard-based) while being observed by the researcher.

At the same time, a video focusing on the gameplay and the child's eye movements will be recorded by an attached camera on the stand and an eye tracker in front of the child. Another video camera will take the child's body reaction situated on the right-hand side of the child.

Images below show the setup of a session.

Equipment used...

An eye tracker will be used in this project to gather the child's eye movements. The eye tracker incorporates with a mobile stand to hold the devices used in the project together. Please note that **no** device is attached to any part of a child's head or body at all. An image of the eye tracker and stand can be seen below.

This is an example of how the child's eye movement data looks like and is essential for this project.

To participate...

An acknowledge letter will be given prior to the study through your child's nursery or school. Parents whom disagrees may complete the enclosed form and return it to the school.

A.4 Permission Request Letter

UNIVERSITY OF LEICESTER

Department of Computer Science
University Road
Leicester LE1 7RH
UK

[redacted]

14th September 2016

Dear [redacted],

I am Dinna Mohd Nizam, a PhD student under the supervision of Dr Effie Law from the Department of Computer Science, University of Leicester. For my PhD research, I aim to develop an effective and enjoyable digital educational game for young children aged five years (Reception/Nursery) to learn numeracy with a tablet. I have developed the game and plan to evaluate it to demonstrate its efficiency. Here I am writing to enquire for permission to recruit participants for my evaluation study in your school, [redacted].

Educational game-based mobile application is becoming prevalent for children of all ages. Your generous support will contribute to advancing this increasingly important effort. Future children in your school will benefit from new insights to be gained through research of this kind.

Here I provide detailed information about the study, which consists of individual sessions, and the help that I'd like to request. In a single session, a participating child and a researcher (myself) will be present. The child's parent (or carer) will be strongly encouraged to be present in this session. The child will be asked to play an educational game (tablet- or cardboard-based) and be observed. In addition, videos focusing on the gameplay will be recorded with an eye-tracker (see leaflet for more details). After the child finishes playing the game, I will conduct a short interview with the child and with the parent/carer, if available, to understand the child's gameplay experience.

The session should take place in a room with least distraction. Teachers and parents/carers will fully be informed about the process through a leaflet. An acknowledge letter will be sent to parents/carers prior to the study. An example of the letter has been attached.

Furthermore, I am much willing to work as a volunteer in classrooms involving 5 years old children (Reception/Nursery) at your school for a few days before the study to build rapport with the children.

I have passed the Disclosure and Barring Service check, certificate no: 001500710472. This research has been approved by the University of Leicester Ethics Committee, reference: 2907-dmn9-computerscience. A full version of the Code of Practice for research ethics can be found on this link http://www2.le.ac.uk/institution/ethics/code.

If you would kindly grant me the permission to conduct the aforementioned study in your school, [redacted], please email the enclosed form to dmn9@le.ac.uk by the 22nd **September 2016.** An email version of this request letter has also been sent to the school's email written above.

If you have any questions about this study or would like to have additional information, please do not hesitate to contact me or my supervisor, Dr Effie Law using the contacts provided on the leaflet.

Thank you in advance for your interest and support of this study.
With kind regards,

[redacted]

Dinna M. Nizam
Student ID: [redacted]
E-mail: dmn9@le.ac.uk

A.5 Head Teacher Consent Form

SCH	CF	

UNIVERSITY OF LEICESTER

Digital Educational Games on Numeracy for Preschool Children

School Consent Form

School Name: ...[redacted]...

I am aware and,

	Yes	No
Have read the Data Protection Act information in the leaflet and have been fully informed of the relevant data protection procedures.	☑	☐
Give permission for my school to participate in this study.	☑	☐
Give permission for my school to help identify groups of **reception aged** children for this study.	☑	☐
Give permission for my school to help handout parental consent letters to parents/carers of **reception aged children** enrolled in this school.	☑	☐
Give permission to the researcher to use a space with least distraction to carry out this study in the school.	☑	☐
Give permission to allow the researcher to work as a volunteer in the school for a few days (at least 2 days) before running the actual study with children.	☑	☐

I also understand,

	Yes	No
That the children who participate in this study are free to withdraw themselves anytime without any reason.	☑	☐
That I must keep the identity of all children who participate confidential.	☑	☐
That I am free to contact the researcher and University of Leicester Research Ethic Officer to discuss any questions that I might have.	☑	☐
That at the end of the study I will be provided with additional information and feedback about the study.	☑	☐

I, ...[redacted]... give permission to the researcher, Dinna Mohd Nizam, to proceed with this study under the supervision of Dr Effie Law.

Signature: ...[redacted]...

Date: 23/02/17

School Representative's contact person: ...[redacted]...

A.6 Parent Consent Form

P	Acknowledge	

UNIVERSITY OF LEICESTER

Digital Educational Games on Numeracy for Preschool Children Research Study

Parents' Acknowledge Letter

School: [redacted]

Dear Parents,

A research study on Digital Educational Games in Numeracy among reception aged children will be done in your child's school for a week. Every child in Rainbows will take part in this 20 to 30 minutes research study. Please sign this letter below and return it to the school by Friday the 14th October 2016.

Please be informed in this research:

- Every child will be observed as part of the child's research data.
- Every child will be videoed as part of the child's research data.
- Every child will be photographed as part of the child's research data.
- Every child's observation records will be reproduced and published anonymously as findings of the research.
- Every child's video records will be reproduced and published anonymously as findings of the research.
- Every child's photographs will be reproduced and published anonymously as findings of the research.
- We would like to reassure you that any data, video or photographs mention above **will not** be published on the internet.

Should you have any questions or would like additional information please do not hesitate to contact me at 07873914082 or dmn9@le.ac.uk. You may also contact the Research Ethics Officer Dr. Effie Law at lcl9@le.ac.uk about child's rights as a research participant. We would be very happy to answer any queries that you may have.

Regards,

Dinna M.N.
Department of Computer Science,
University of Leicester.

I, ... **agree** / **disagree** to let my child participate in this
(Parents or carers Name)
research study.

Child's Name:... Class:...

Parents or Carers Signature: ....

A.7 Pre Test and Post Test Material (Front Page)

Participant: .... School Name: .... Class: .... Test: PRE / POST Time:

Count and circle the right numbers.

4 1 3

3 6 5

7 8 5

10 12 9

10 14 12

16 15 18

A.8 Pre Test and Post Test Material (Back Page)

Find the missing numbers.

7 4 6

1 2 3 ___ 5

---

1 7 3

2 ___ 4 5 6

---

9 11 8

6 7 ___ ___ 10

---

12 14 10

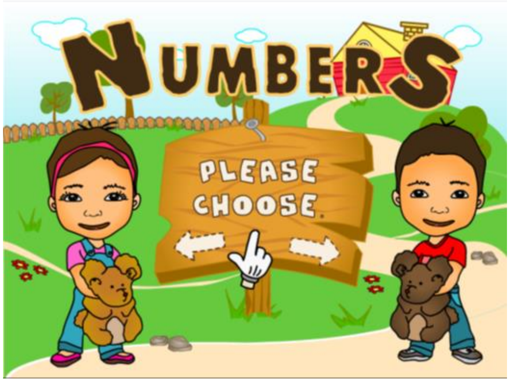
9 ___ 11 ___ 13

---

15 17 14

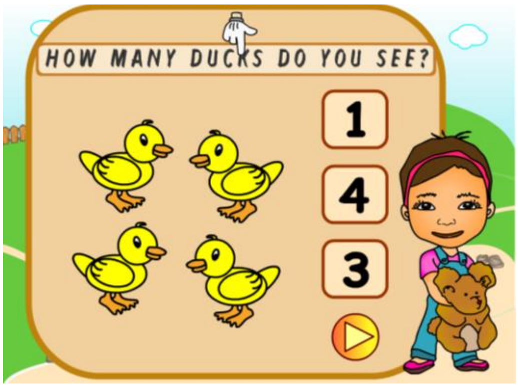
12 13 ___ ___ 16

A.9 Game Design Interfaces/ DEG (Screenshots)

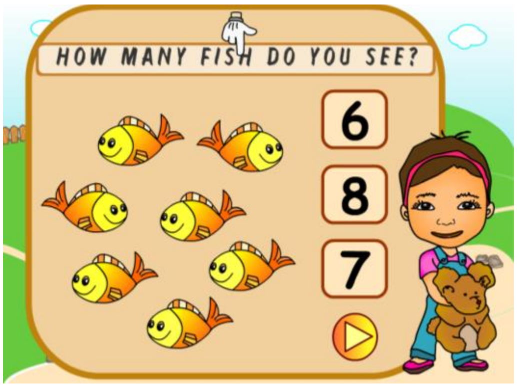


Main Page

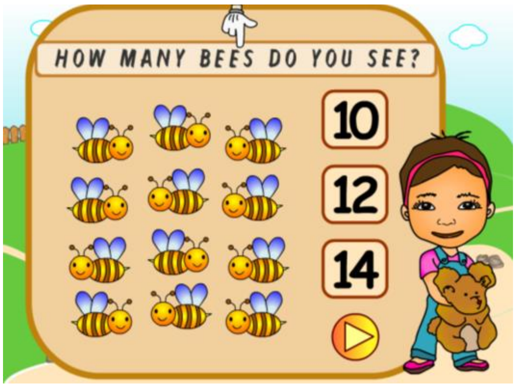
Selection Page for Game M



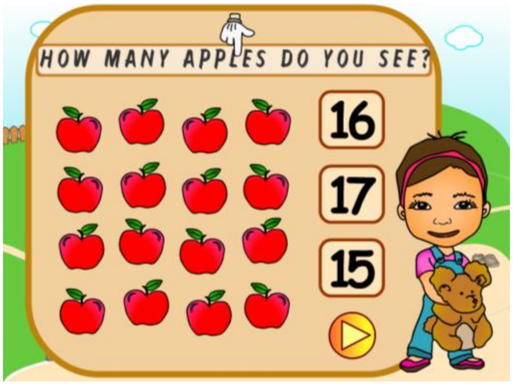
Game Page M1 (Level 1)



Game Page M2 (Level 2)



Game Page M3 (Level 3)



Game Page M4 (Level 4)

Selection Page for Game S

Game S Demonstration Page

Game Page S1 (Level 1)

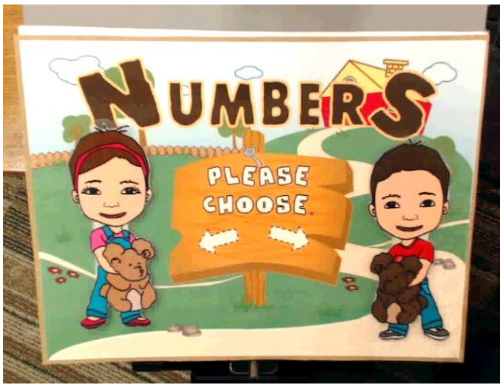
Game Page S2 (Level 2)

Game Page S3 (Level 3)

Game Page S4 (Level 4)

Reward Page

A.10 Cardboard Game Pages (Tobii Screenshots)



Main Page

Selection Page for Game M

Game Page M1 (Level 1)

Game Page M2 (Level 2)

Game Page M3 (Level 3)

Game Page M4 (Level 4)

Selection Page for Game S

Game Page S1 (Level 1)

Game Page S2 (Level 2)

Game Page S3 (Level 3)

Game Page S4 (Level 4)

Reward Page

A.11 Game Experience scores coded by the raters.

		Coding 1						Coding 2						Agreed Coding				
partid	**partcode**	**Selection**	**Game M**	**Game S**	**End**	**Total Score**		**Selection**	**Game M**	**Game S**	**End**	**Total Score**		**Selection**	**Game M**	**Game S**	**End**	**Total Score**
		2	2	2	2	8		2	2	2	2	8		2	2	2	2	8
1	OV01	2	2	2	2	8		1	1	1	1	4		2	2	2	2	8
2	OV02	1	1	1	1	4		1	1	1	1	4		1	1	1	1	4
3	OV03	1	0	1	1	3		1	0	1	1	3		1	0	1	1	3
4	OV04	2	2	2	2	8		2	2	2	2	8		2	2	2	2	8
5	OV05	1	1	1	1	4		1	1	1	1	4		1	1	1	1	4
6	OV06	1	2	2	2	7		1	1	1	2	5		1	2	2	2	7
7	OV07	2	2	2	2	8		1	1	1	2	5		2	2	2	2	8
8	OV08	2	2	2	1	7		2	2	2	0	6		2	2	2	1	7
9	OV09	1	2	1	2	6		1	2	1	2	6		1	2	1	2	6
10	OV10	1	2	2	2	7		1	2	2	2	7		1	2	2	2	7
11	OV11	2	2	1	2	7		2	2	1	2	7		2	2	1	2	7
12	OV12	1	2	1	2	6		1	2	1	2	6		1	2	1	2	6
13	OV13	2	1	1	1	5		2	1	1	1	5		2	1	1	1	5
14	OV14	1	1	1	1	4		1	1	1	1	4		1	1	1	1	4
15	OV15	0	1	1	0	2		1	0	1	0	2		0	1	1	0	2
16	OV16	1	2	1	1	5		1	1	1	2	5		1	2	1	1	5
17	OV17	2	2	2	2	8		2	2	2	2	8		2	2	2	2	8
18	OV18	2	2	2	1	7		2	2	2	1	7		2	2	2	1	7
19	OV19	1	2	2	1	6		1	2	1	2	6		1	2	2	1	6
20	OV20	1	1	1	1	4		1	1	1	1	4		1	1	1	1	4
21	OV21	1	2	1	2	6		1	2	1	2	6		1	2	1	2	6
22	OV22	1	2	1	1	5		1	2	1	1	5		1	2	1	1	5
23	OV23	1	1	1	1	4		1	1	1	1	4		1	1	1	1	4
24	OV24	1	2	2	2	7		1	2	2	2	7		1	2	2	2	7
25	OV25	1	1	1	1	4		1	1	1	1	4		1	1	1	1	4
26	OV26	1	0	0	1	2		0	0	0	1	1		1	0	0	1	2
27	OV27	1	2	1	1	5		1	2	1	2	6		1	2	1	1	5
28	OV28	1	2	1	2	6		1	2	1	2	6		1	2	1	2	6
29	OV29	2	1	1	1	5		2	1	1	1	5		2	1	1	1	5
30	OV30	1	2	1	2	6		2	2	1	2	7		1	2	1	2	6
31	OV31	1	1	1	1	4		2	2	2	2	8		1	1	1	1	4
32	OV32	1	1	1	1	4		1	1	1	1	4		1	1	1	1	4
33	OV33	1	2	2	2	7		1	2	2	2	7		1	2	2	2	7
34	OV34	1	1	1	1	4		1	1	1	1	4		1	1	1	1	4
35	OV35	1	1	1	1	4		1	1	1	1	4		1	1	1	1	4
36	OV36	2	2	2	2	8		2	2	2	2	8		2	2	2	2	8
37	OV37	1	1	1	1	4		1	1	1	1	4		1	1	1	1	4
38	OV38	2	2	2	2	8		2	2	2	2	8		2	2	2	2	8
39	OV39	1	1	1	1	4		1	1	1	1	4		1	1	1	1	4
40	OV40	1	2	2	1	6		1	2	2	1	6		1	2	2	1	6

41	OV41	1	1	1	1	4		1	1	1	1	4		1	1	1	1	4
42	OV42	1	1	1	1	4		1	1	1	1	4		1	1	1	1	4
43	OV43	1	2	1	1	5		1	1	2	1	5		1	2	1	1	5
44	OV44	1	1	1	1	4		1	2	2	2	7		1	1	1	1	4
45	OV45	1	2	1	2	6		1	2	1	2	6		1	2	1	2	6
46	OV46	1	1	1	1	4		1	1	1	1	4		1	1	1	1	4
47	OV47	1	2	1	1	5		1	2	2	2	7		1	2	1	1	5
48	OVBd01	1	1	1	0	3		1	1	1	1	4		1	1	1	0	3
49	OVBd02	1	2	1	1	5		1	2	1	1	5		1	2	1	1	5
50	OVBd03	1	2	2	1	6		1	2	2	1	6		1	2	2	1	6
51	OVBd04	1	1	1	1	4		1	1	1	1	4		1	1	1	1	4
52	OVBd05	1	2	2	1	6		1	2	2	1	6		1	2	2	1	6
53	OVBd06	1	2	2	1	6		1	2	2	1	6		1	2	2	1	6
54	OVBd07	1	2	1	1	5		1	2	1	1	5		1	2	1	1	5
55	OVBd08	1	2	1	1	5		1	1	2	2	6		1	2	1	1	5
56	OVBd09	1	1	1	1	4		1	1	1	1	4		1	1	1	1	4
57	OVBd10	1	1	1	1	4		1	1	1	1	4		1	1	1	1	4
58	OVBd11	1	1	1	1	4		1	2	2	2	7		1	1	1	1	4
59	OVBd12	1	2	2	1	6		2	2	2	2	8		1	2	2	1	6
60	OVBd13	1	1	1	1	4		1	1	1	1	4		1	1	1	1	4
61	OVBd14	1	2	2	2	7		1	2	2	2	7		1	2	2	2	7
62	OVBd15	1	1	1	1	4		1	2	2	2	7		1	1	1	1	4
63	UP48	1	1	1	2	5		1	1	1	2	5		1	1	1	2	5
64	UP49	1	1	1	1	4		2	2	1	2	7		2	2	1	2	7
65	UP50	1	2	1	2	6		1	2	1	2	6		1	2	1	2	6
66	UP51	0	1	1	1	3		1	2	2	2	7		1	2	2	2	7
67	UP52	1	0	1	0	2		1	1	1	1	4		1	1	1	1	4
68	UP53	1	2	1	2	6		1	2	1	2	6		1	2	1	2	6
69	UP54	0	1	1	1	3		1	2	1	1	5		1	2	1	1	5
70	UP55	1	2	1	2	6		1	2	1	2	6		1	2	1	2	6
71	UP56	1	1	1	1	4		1	1	1	1	4		1	1	1	1	4
72	UP57	1	2	2	2	7		1	2	2	2	7		1	2	2	2	7
73	UP58	1	2	1	2	6		1	2	1	2	6		1	2	1	2	6
74	UP59	1	2	2	1	6		1	2	2	1	6		1	2	2	1	6
75	UPBd_16	1	1	1	1	4		1	1	1	1	4		1	1	1	1	4
76	UPBd_17	1	2	1	1	5		1	2	1	1	5		1	1	1	1	4
77	UPBd_18	1	1	1	1	4		1	1	1	1	4		1	1	1	1	4
78	UPBd_19	2	2	2	2	8		2	2	2	2	8		2	2	2	2	8
79	UPBd_20	1	1	1	1	4		1	1	2	1	5		1	1	2	1	5
80	UPBd_21	2	2	2	2	8		2	2	2	2	8		2	2	2	2	8
81	UPBd_22	1	2	2	1	6		1	2	2	1	6		1	2	2	1	6
82	UPBd_23	1	2	2	2	7		1	2	2	2	7		1	2	2	2	7
83	UPBd_24	2	2	2	2	8		2	2	2	2	8		2	2	2	2	8
84	UPBd_25	1	2	2	2	7		1	2	2	2	7		1	2	2	2	7
85	UPBd_26	1	1	1	1	4		1	2	2	2	7		1	2	2	2	7
86	UPBd_27	1	2	1	1	5		1	2	1	1	5		1	2	1	1	5
87	UPBd_28	1	1	1	1	4		2	2	2	2	8		2	2	2	2	8
88	UPBd_29	1	2	1	1	5		1	2	1	1	5		1	2	1	1	5
89	UPBd_30	1	1	1	1	4		1	1	1	1	4		1	1	1	1	4
90	UPBd_31	1	1	1	1	4		1	0	2	1	4		1	0	2	1	4
91	UPBd_32	1	2	2	2	7		1	2	2	2	7		1	2	2	2	7
92	UPBd_33	2	2	2	2	8		2	2	2	2	8		2	2	2	2	8
93	UPBd_34	1	2	2	2	7		1	2	2	2	7		1	2	2	2	7
94	UPBd_35	2	2	2	2	8		2	2	2	2	8		2	2	2	2	8

Bibliography

Adobe System Incorporated. (2018). Adobe AIR. Retrieved February 5, 2018, from https://www.adobe.com/uk

Aladé, F., Lauricella, A. R., Beaudoin-Ryan, L., & Wartella, E. (2016). Measuring with Murray: Touchscreen technology and preschoolers' STEM learning. *Computers in Human Behavior*, *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. *Computers and Education*, *125*(July), 413–428. https://doi.org/10.1016/j.compedu.2018.06.023

All, A., Nuñez Castellar, E. P., & Van Looy, J. (2016). Assessing the effectiveness of digital game-based learning: Best practices. *Computers & Education*, *92–93*, 90–103. https://doi.org/http://dx.doi.org/10.1016/j.compedu.2015.10.007

Animate CC. (2017). ActionScript with Animate CC. Retrieved February 5, 2018, from https://www.adobe.com/uk

Ariasi, N., & Mason, L. (2011). Uncovering the effect of text structure in learning from a science text: An eye-tracking study. *Instructional Science, 39*(5), 581–601.

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. *Journal of Educational Psychology*, *105*(4), 957–969. Retrieved from http://search.ebscohost.com.ezproxy4.lib.le.ac.uk/login.aspx?direct=true&db=pdh&AN=2013-32700-001&site=ehost-live

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*(1–2), 135–152. https://doi.org/10.3233/JAI-2011-020

Atkinson, R. C., & Shiffrin, R. M. (1968). Human memory: A proposed system and its control processes1. In *Psychology of learning and motivation* (Vol. 2, pp. 89–195). Elsevier.

Baddeley, A. D. (1997). Retrieval. In *Human memory : Theory and Practice (Revised Edition)* (Revised Ed, p. Chapter 11, 191-210). England: Psychology Press.

Barr, R. (2013). Memory Constraints on Infant Learning From Picture Books, Television, and Touchscreens. *Child Development Perspectives*, *7*(4), 205–210. https://doi.org/10.1111/cdep.12041

Buffum, P. S., Frankosky, M., Boyer, K. E., Wiebe, E., Mott, B., & Lester, J. (2015). Leveraging collaboration to improve gender equity in a game-based learning environment for middle school computer science. In *Research in Equity and Sustained Participation in Engineering, Computing, and Technology (RESPECT), 2015* (pp. 1–8). https://doi.org/10.1109/RESPECT.2015.7296496

Bulling, A., Duchowski, A. T., & Majaranta, P. (2011). PETMEI 2011: the 1st international workshop on pervasive eye tracking and mobile eye-based interaction. *Proceedings of the 13th International Conference on Ubiquitous Computing*. Beijing, China: ACM. https://doi.org/10.1145/2030112.2030248

Bulling, A., & Gellersen, H. (2010). Toward Mobile Eye-Based Human-Computer Interaction. *Pervasive Computing, IEEE*, *9*(4), 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. *Proceedings of the 13th International Conference on Ubiquitous Computing*. Beijing, China: ACM. https://doi.org/10.1145/2030112.2030172

Burnett, C. (2010). Technology and literacy in early childhood educational settings: A review of research. *Journal of Early Childhood Literacy*, *10*(3), 247–270. https://doi.org/10.1177/1468798410372154

Catsambis, S. (2004). The Gender Gap in Mathematics: Merely a Step Function ? In A. M. Gallagher & J. C. Kaufman (Eds.), *Gender differences in mathematics : an integrative psychological approach* (pp. 220–245). Retrieved from https://ebookcentral-proquest-com.ezproxy4.lib.le.ac.uk

Chiong, C., & Shuler, C. (2010). Learning: Is there an app for that. In *Investigations of young children's usage and learning with mobile devices and apps.* New York: The Joan Ganz Cooney Center at Sesame Workshop. 2010.

Department for Education. (2013). Early Years Outcomes: A non-statutory guide for practitioners and inspectors to help inform understanding of child development through the early years. Retrieved from https://www.foundationyears.org.uk/files/2012/03/Early_Years_Outcomes.pdf

Department for Education. (2014). *Statutory framework for the early years foundation stage*. (U. K. Department For Education, Ed.), *DfE publications*. Department for Education. https://doi.org/00266-2008BKT-EN

Department Of Education, T. and E. (2012). *Mobile Devices in Special Education: advice for schools*. The State of Queensland, Australia. Retrieved from http://education.qld.gov.au/smartclassrooms/documents/enterprise-platform/pdf/mobile-devices-special-education.pdf

DeVries, R. (2000). Vygotsky, Piaget, and Education: a reciprocal assimilation of theories and educational practices. *New Ideas in Psychology*, *18*(2), 187–213. https://doi.org/https://doi.org/10.1016/S0732-118X(00)00008-8

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

Drachen, A., Göbel, S., Nacke, L. E., Drachen, A., & Göbel, S. (2010). Methods for evaluating gameplay experience in a serious gaming context. *International Journal of Computer Science in Sport*, *9*(2), 1–12. https://doi.org/10.1109/TLT.2010.21

Duh, E. S., Koceska, N., & Koceski, S. (2017). Game-based learning: educational game Azbuka to help young children learn writing Cyrillic letters. *Multimedia Tools and Applications*, *76*(12), 14091–14105. https://doi.org/10.1007/s11042-016-3829-9

Early Education. (2012). *Development Matters in the Early Years Foundation Stage*. Retrieved from https://www.early-education.org.uk/development-matters-early-years-foundation-stage-eyfs

Eraslan, S., Yesilada, Y., & Harper, S. (2016). Eye tracking scanpath analysis techniques on web pages: A survey, evaluation and comparison. *Journal of Eye Movement Research*, *9(1)*(2), 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. *Computers & Education*, *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. Retrieved from https://books.google.co.uk/books?hl=en&lr=&id=ispqH-v-8PAC&oi=fnd&pg=PP1&dq=Essa,+A.+L.+(2012).+Introduction+to+early+childhood+education.+Cengage+Learning.&ots=yd1pWeo99F&sig=Y8vnDLwquAxnF2Ariwd5-6wbtGQ

Fails, J. A., Druin, A., Guha, M. L., Chipman, G., Simms, S., & Churaman, W. (2005).

Child’s play: A Comparison of Desktop and Physical Interactive Environments. *Proceeding of the 2005 Conference on Interaction Design and Children - IDC ’05, 08*, 48–55. https://doi.org/10.1145/1109540.1109547

Feng, J., Spence, I., & Pratt, J. (2007). Playing an Action Video Game Reduces Gender Differences in Spatial Cognition. *Psychological Science*, *18*(10), 850–855. https://doi.org/10.1111/j.1467-9280.2007.01990.x

Frost, J. L., Wortham, S. C., & Reifel, R. S. (2008). *Play and child development*. Pearson/Merrill Prentice Hall Upper Saddle River, NJ.

Fu-Hsing, T., Kuang-Chao, Y., & Hsien-Sheng, H. (2012). Exploring the Factors Influencing Learning Effectiveness in Digital Gamebased Learning. *Journal of Educational Technology & Society*, *15*(3), 240–250.

Furió, D., Juan, M.-C., Seguí, I., & Vivó, R. (2015). Mobile learning vs. traditional classroom lessons: a comparative study. *Journal of Computer Assisted Learning*, *31*(3), 189–201. https://doi.org/10.1111/jcal.12071

Gosvāmī, U., & Goswami, U. (2008). *Cognitive development: the learning brain*. Hove: Psychology Press.

Goswami, U. (2010). *Blackwell Handbooks of Developmental Psychology : Wiley-Blackwell Handbook of Childhood Cognitive Development (2nd Edition)*. Hoboken, NJ, USA: Wiley-Blackwell.

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. *Behavior Research Methods*, *48*(4), 1694–1712.

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.

Huitt, W. (2003). The information processing approach to cognition. *Educational Psychology Interactive*, *3*(2), 53–67.

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 Transactions on Learning Technologies*, *7*(1), 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*, *321*(5888), 494 LP-

495. Retrieved from http://science.sciencemag.org/content/321/5888/494.abstract

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 & Behavior*, *11*(4), 437–442.

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

KementerianPelajaranMalaysia. (2010). Kurikulum Standard Prasekolah Kebangsaan. Kuala Lumpur: Bahagian Pembangunan Kurikulum.

Lai, M.-L., Tsai, M.-J., Yang, F.-Y., Hsu, C.-Y., Liu, T.-C., Lee, S. W.-Y., … Tsai, C.-C. (2013). A review of using eye-tracking technology in exploring learning from 2000 to 2012. *Educational Research Review*, *10*, 90–115. https://doi.org/http://dx.doi.org/10.1016/j.edurev.2013.10.001

Lander, L. (2002). *Sustained Attention, Self-regulation, and Motivation During a Maze Task for Preschool Children: A Motivational, Cognitive Analysis of the Process of Sustained Interest*. University of Rochester. Retrieved from https://commons.esc.edu/lorrainelander/wp-content/uploads/sites/182/2011/07/Dissertation-Lorraine-Lander.pdf

Le Meur, O., & Baccino, T. (2013). Methods for comparing scanpaths and saliency maps: Strengths and weaknesses. *Behavior Research Methods*, *45*(1), 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*(5), 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. *Interactive Learning Environments*, *19*(5), 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. *Cognition and Emotion*, *32*(5), 1122–1130. https://doi.org/10.1080/02699931.2017.1365046

Lourenço, O. (2012). Piaget and Vygotsky: Many resemblances, and a crucial difference. *New Ideas in Psychology*, *30*(3), 281–295. https://doi.org/http://dx.doi.org/10.1016/j.newideapsych.2011.12.006

Lutz, S. T., & Huitt, W. G. (2003). Information processing and memory: Theory and applications. Retrieved from http://www.edpsycinteractive.org/papers/infoproc.pdf

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

Markopoulos, P., Read, J. C., MacFarlane, S., & Hoysniemi, J. (2008). *Evaluating children's interactive products: principles and practices for interaction designers*. Morgan Kaufmann. Retrieved from http://www.myilibrary.com?ID=179543

Mason, L., Pluchino, P., Tornatora, M. C., & Ariasi, N. (2013). An Eye-Tracking Study of Learning From Science Text With Concrete and Abstract Illustrations. *The Journal of Experimental Education*, *81*(3), 356–384. https://doi.org/10.1080/00220973.2012.727885

Mason, L., Tornatora, M. C., & Pluchino, P. (2013). Do fourth graders integrate text and picture in processing and learning from an illustrated science text? Evidence from eye-movement patterns. *Computers & Education*, *60*(1), 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. *Journal of Computer Assisted Learning*, *33*(5), 403–423. https://doi.org/10.1111/jcal.12197

McIntyre, E., Rightmyer, E., Powell, R., Powers, S., & Petrosko, J. (2006). How much should young children read? A study of the relationship between development and instruction. *Literacy Teaching and Learning*, *11*(1), 51–72.

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. *Journal of Experimental Child Psychology*, *137*, 137–155. https://doi.org/10.1016/j.jecp.2015.04.002

Moser, T., & Reikerås, E. (2016). Motor-life-skills of toddlers – a comparative study of Norwegian and British boys and girls applying the Early Years Movement Skills Checklist. *European Early Childhood Education Research Journal*, *24*(1), 115–135. https://doi.org/10.1080/1350293X.2014.895560

Neumann, M. M. (2018). Using tablets and apps to enhance emergent literacy skills in young children. *Early Childhood Research Quarterly*, *42*(August 2017), 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 Childhood Education Journal*, *42*(4), 231–239.

https://doi.org/10.1007/s10643-013-0608-3

Oakes, L. M., Ross-Sheehy, S., & Kannass, K. N. (2004). Attentional engagement in infancy: The interactive influence of attentional inertia and attentional state. *Infancy*, *5*(2), 239–252. https://doi.org/10.1207/s15327078in0502_8

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. *Scientific Reports*, *8*(1), 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. *Journal of Behavioral Decision Making*, *29*(2–3), 103–115. https://doi.org/10.1002/bdm.1867

Oskar, K., Sherry, Y. C., Denise, H. W., Andrew, C. C. L., & Tak-Wai, C. (2014). The Effects of Game-Based Learning on Mathematical Confidence and Performance: High Ability vs. Low Ability. *Journal of Educational Technology & Society*, *17*(3), 65–78.

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

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 Education and Development*, *24*(2), 188.

Patel, D. (2017). Will Technology Ruin Your Children's Development? Retrieved March 23, 2018, from https://journal.thriveglobal.com/will-technology-ruin-your-childrens-development-663351c76974

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

Perles, K. (2013). Is Technology Killing Your Child's Attention Span? Retrieved March 23, 2018, from https://gb.education.com/magazine/article/attention-problems-blame-technology/

Piaget, J. (2000). Piaget's theory of cognitive development. *Childhood Cognitive Development: The Essential Readings*, 33–47.

Ponce, H. R., & Mayer, R. E. (2014). Qualitatively different cognitive processing during online reading primed by different study activities. *Computers in Human Behavior*, *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. *Encyclopedia of Human Computer Interaction*, *1*, 211–219.

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

Rayner, K. (2009). Eye movements and attention in reading, scene perception, and visual search. *The Quarterly Journal of Experimental Psychology*, *62*(8), 1457–1506. https://doi.org/10.1080/17470210902816461

Rayner, K., & Reingold, E. M. (2015). Evidence for direct cognitive control of fixation durations during reading. *Current Opinion in Behavioral Sciences*, *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* (p. 16:1--16:10). New York, NY, USA: ACM. 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* (pp. 81–88). ACM.

Richter, A., & Courage, M. L. (2017). Comparing electronic and paper storybooks for preschoolers: Attention, engagement, and recall. *Journal of Applied Developmental Psychology*, *48*, 92–102. https://doi.org/10.1016/j.appdev.2017.01.002

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

Ruff, H. A., & Capozzoli, M. C. (2003). Development of attention and distractibility in the first 4 years of life. *Developmental Psychology*, *39*(5), 877–890. https://doi.org/10.1037/0012-1649.39.5.877

Ruff, H. A., Capozzoli, M., & Weissberg, R. (1998). Age, individuality, and context as factors in sustained visual attention during the preschool years. *Developmental Psychology*, *34*(3), 454–464. https://doi.org/10.1037/0012-1649.34.3.454

Ruff, H. A., & Lawson, K. R. (1990). Development of sustained, focused attention in young children during free play. *Developmental Psychology*, *26*(1), 85–93. https://doi.org/10.1037/0012-1649.26.1.85

Sabol, T. J., Bohlmann, N. L., & Downer, J. T. (2018). Low-Income Ethnically Diverse Children's Engagement as a Predictor of School Readiness Above Preschool Classroom Quality. *Child Development*, *89*(2), 556–576. https://doi.org/10.1111/cdev.12832

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

SesameWorkshop. (2012). Best Practices : Designing Touch Tablet Experiences for Preschoolers. Retrieved from http://www.sesameworkshop.org/wp_install/wp-content/uploads/2013/04/Best-Practices-Document-11-26-12.pdf

Shabani, K., Khatib, M., & Ebadi, S. (2010). Vygotsky's zone of proximal development: Instructional implications and teachers' professional development. *English Language Teaching*, *3*(4), 237.

She, H.-C., & Chen, Y.-Z. (2009). The impact of multimedia effect on science learning: Evidence from eye movements. *Computers & Education*, *53*(4), 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. *International Journal of Computer Games Technology*, *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*.

Simões, J., Redondo, R. D., & Vilas, A. F. (2013). A social gamification framework for a K-6 learning platform. *Computers in Human Behavior*, *29*(2), 345–353. https://doi.org/10.1016/j.chb.2012.06.007

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 V. Dimitrova, T. Kuflik, D. Chin, F. Ricci, P. Dolog, & G.-J. Houben (Eds.), *User Modeling, Adaptation, and Personalization: 22nd*

International Conference, UMAP 2014, Aalborg, Denmark, July 7-11, 2014. Proceedings (pp. 183–194). Cham: Springer International Publishing. 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. *Journal of Information Technology Education: Innovations in Practice*, *15*(1), 145–165.

Sung, H.-Y., & Hwang, G.-J. (2013). A collaborative game-based learning approach to improving students' learning performance in science courses. *Computers & Education*, *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. *Behaviour & Information Technology*, *31*(7), 679–695. https://doi.org/10.1080/0144929X.2012.692101

Tobii AB. (2014). *User Manual - Tobii Studio*. *Tobii* (Vol. 3.3.0). Retrieved from www.tobii.com

Tobii AB. (2016). *Tobii Pro Mobile Device Stand User ' s Manual*. Retrieved from https://www.tobiipro.com

Tobii Dynavox. (2017). How eye tracking works - Tobii Dynavox. Retrieved May 6, 2018, from https://www.tobiidynavox.com/about/about-us/how-eye-tracking-works/

Tobii Eye Tracker. (2015). How do Tobii Eye Trackers work? Retrieved from http://www.tobiipro.com/learn-and-support/learn/

Tsai, M.-J., Hou, H.-T., Lai, M.-L., Liu, W.-Y., & Yang, F.-Y. (2012). Visual attention for solving multiple-choice science problem: An eye-tracking analysis. *Computers & Education*, *58*(1), 375–385. https://doi.org/http://dx.doi.org/10.1016/j.compedu.2011.07.012

Voyer, D., & Voyer, S. D. (2014). Gender differences in scholastic achievement: A meta-analysis. *Psychological Bulletin*, *140*(4), 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* (pp. 2893–2902). New York, NY, USA: ACM.

https://doi.org/10.1145/2470654.2481400

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

Widen, S., & A Russell, J. (2010). Differentiation in Preschooler's Categories of Emotion. *Emotion*, *10*, 651–661.

Zaranis, N., Kalogiannakis, M., & Papadakis, S. (2013). Using Mobile Devices for Teaching Realistic Mathematics in Kindergarten Education. *Creative Education*, *4*(7), 1–10. https://doi.org/10.4236/ce.2013.47A1001

Zin, N. A. M., Jaafar, A., & Yue, W. S. (2009). Digital game-based learning (DGBL) model and development methodology for teaching history. *WSEAS Transactions on Computers*, *8*(2), 322–333. https://doi.org/10.1016/j.apsusc.2010.11.070