In the Eyes of Young Children: A Study on Focused Attention to Digital Educational Games

Dinna N. Mohd Nizam
Department of Informatics
University of Leicester
University Road, Leicester,
LE1 7RH, UK
dmn9@le.ac.uk

Effie Lai-Chong Law
Department of Informatics
University of Leicester
University Road, Leicester,
LE1 7RH, UK
lcl9@le.ac.uk

Numerous research studies on digital educational games (DEGs) focused on *whether* they help improve children's learning performance. Nonetheless, only a few studies sought to address *how* children learn through DEGs. We were motivated to bridge this gap through an empirical study with the eye-tracking methodology. A total of 94 five-year-olds were involved in the study. They were asked to play with a DEG or cardboard game on numeracy. We analysed how fixation duration (a proxy for attention) was related to learning strategies based on the children's achievement level. The main findings include: the DEG did not yield any significant learning effect but its cardboard version did; the children's performance for recall-based tasks were significantly worse than that for recognition-based task; the achievement level played a significant role in varying attention given to the objects of the games. Reflections on applying the eye-tracking methods to young children were also discussed.

Digital Educational Games, Early Childhood Learning, Eye-tracking, Numeracy.

1. INTRODUCTION

The popularity of digital gadgets such as tablets has boosted the use of digital educational games (DEGs) by young children. 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 [9]. 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 [57]. Nonetheless, the number of studies on the learning effect of DEGs on young children aged three to five years remains low [1,8,37]. Methodologically, it is challenging to work with children of this age group, given their low verbal ability. Typically it is hard for these young children to describe on which objects they focus in a learning environment, let alone explain why they do so [35]; a viable alternative is to derive their attention from eye-tracking data.

The recent development of the eye-tracking technology, especially its applicability for portable devices such as tablets and smartphones, has enabled us to conduct a study with a DEG for young children. The study was aimed to understand via eye tracking data how learning activities were performed by 5-year-olds when playing a DEG on numeracy.

To understand the impact of the medium of an educational game on learning, a dedicated DEG and its cardboard version were designed and developed for this study. The game focused on numeracy from 1 to 20 and adapted the Early Years Foundation Stage (EYFS)[1] framework in the UK educational system. Both game versions and experimental setups were tested in a series of pilot studies in which 35 children were involved. To assess the learning effect, participants were asked to complete a pre- and post-test before and after playing the game, respectively. While questionnaires for evaluating gaming experience with young children exist (e.g. [44]), the issue of social desirability remains hard to address [28]. This and the other concern of prolonging an evaluation session with young children, who typically have a short attention

---

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

span, led us to decide not to use such a questionnaire but to rely on observations and videos.

Overall, the goal of our study was to evaluate the learning effect of the DEG and applicability of eye-tracking methodology for young children. The research questions (RQs) of our study are:

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

RQ2: How is the achievement level of learners related to their attention paid to relevant and irrelevant objects in the game as measured by eye-tacking data?

RQ3: Are learning tasks drawing on recall more difficult than those on recognition for young children for both media – digital and physical - of the educational game?

RQ4: What are the challenges of applying the eye-tracking method to young children?

The contribution of our work is twofold: First, improving the understanding of factors influencing how young children learn via DEGs (RQ1, RQ2, and RQ3). Second, reflecting on the applicability of the eye-tracking method for young children using DEGs to gain insights into the methodological challenges (RQ4).

A Disclosure and Barring Service check (DBS) and an ethics approval from the University Research Ethics Committee were obtained in advance. Such a check and review are critical for studies with young children to ensure that the researcher is eligible to work with this vulnerable group. Both parental and children's consents were acquired as prerequisites for the participation in the study. All participations were voluntary without any compensation.

2. RELATED WORK

2.1 Piaget's theories and Game-based Learning

Among different theories of cognitive development, Piaget's contributed significantly to today's pedagogical approaches to teaching young children [11]. Specifically, Piaget's work focuses on how children construct knowledge independently as they grow in five different stages [39]. Children of five years old – the target group of our study - are in the preoperational stage when they start to recognise symbols and use language to understand the world, despite the fact that their capability of logical thinking is still limited [22]. Furthermore, in Piaget's theory, play is seen as an essential 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). This developmental model on play can inform game-based learning (GBL) [15].

Games, given their strong motivational power in engaging players, have increasingly been used for supporting learning. Indeed, many GBL research studies have been conducted, addressing various aspects of learning such as performance [21,25,53], satisfaction [17], motivation [13,34,56], effectiveness [2,16] and confidence [33]. Prensky [41] discussed 6 basic features of a game: rules, goals, feedback, challenge, interaction, and story. With the advent of new technologies such as touchscreen, the ways games are played have changed. A useful design guideline on touchscreen applications for young children highlights several features such as navigation assistance, layout orientation, design placement and suitable multimedia [50]. The DEG in this study was primarily developed and designed based on Shi and Shih's [52] game-based learning design model, taking into account the features discussed previously.

Recognition and recall features were also essential in the game development with regards to information processing. The Stage Theory Model of information processing that was built upon a computer metaphor describes how information is stored and retrieved [18]. Accordingly, there are three stages of memory -sensory, short-term and long-term - and two access processes - recognition and recall. Recognition refers to the process of comparing information on current perception of an object or event with the memory, typically involving a single decision process based on perceived similarity Recall is related to remembering objects, facts or events that must be retrieved from the memory bank and normally involves a two-stage process - searching (retrieving) and decision-making (recognition) [4].

2.2 Eye-Tracking Technology

In the field of HCI, eye-tracking has become a standard research approach [27]. 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 [5]. Capturing participants' eye movements in natural conversations and behaviours is the primary benefit of mobile eye-tracking [6]. 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 [20]. At the same time, the quality of software packages for automatic analysis of eye movements has also been improved. These advanced features of the eye-tracking technology

have stimulated as well as supported research on cognition and learning [7,42]. A number of studies on eye-tracking and cognitive processes have been conducted in various areas such as reading [3,40], learning effects of multimedia [26,29,30] and questions answering strategies [55]. For instance, a study using different simulations and animations in classroom learning of middle-school students gained insights into multimedia effects through the analysis of eye fixation data [51].

Two standard measures used in the eye-tracking methodology are fixation and saccade [23]. Fixation refers to the static position of an eye on a specific area being viewed while saccades are quick movements of eyes from one fixation to another in a sequence [49]. Three derivative variables in eye-tracking are the position of fixation, fixation duration and saccade length [26]. Software applications for analysing such basic eye-tracking data have become increasingly sophisticated (e.g. dynamic images), enabling their users to address different research and practice goals. Nonetheless, the use of the eye-tracking methodology with young children is more challenging than with their older counterparts. This can partially explain the relatively lower number of studies in this regard. With our empirical study, we aimed to enrich the applied body of knowledge on this specific issue.

2.3 Attention

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 [47]. Elaboration of such types and issues, however, is beyond the scope of this paper.

Young children are known to have difficulty in maintaining attention on assigned tasks or objects and are easily distracted [45]. A number of studies have measured and studied attention in different ways. A study by Oakes [31] observed attention using the concept of attentional inertia and attentional state on infants between 6.5 to 9 months. Ruff [47] and Lander [24] used the idea of sustained attention in their studies among preschool children. Another study by Ruff [46] 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 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 [47,48].

In this study, the 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 [46]. According to Ruff et al [46], 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-year-olds' would be about 5 to 6 minutes[3]. However, we are fully aware that such an extrapolation can be speculative. As Ruff et al.'s [47] 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 [36,38]. Nonetheless, we have not been able to identify any more recent empirical evidence for the attention span of 5-year-olds, apart from some grey literature[4,5] suggesting that the span could be 10-25 minutes, which may include all three types of attention - casual, settled and focused - mentioned above (see Section 7: Discussion).

3. GAME DEVELOPMENT AND DESIGN

3.1 Game mechanics and content

The software used to develop the DEG was Adobe Animation CC, a platform for authoring animation and game design. ActionScript 3.0 programming language was used on top of Animation CC to allow interactivity in the game such as interactive feedback from a child's interaction. The game was then published using Adobe AIR to deliver the game onto an Android-based tablet that was deployed for the experiment.

For the game design, we took into account the issue of children's short attention span [37] (see also Section 2.3) and referred to existing learning materials in printed as well as digital formats. The game consists of two parts: Matching (Game M) and Sorting (Game S) with recognition- and recall-based

---

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

[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 [47] and 181 (~3 mins) seconds for 54-months-old [48].

[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

tasks, respectively, progressing from easy to hard in four levels. This is to build child's engagement with the game, which nevertheless has to be short to avoid the loss of interest. Both Game M and Game S were aimed to support children's ability to recognise and recall numeric symbols from 1 to 20, which is compatible with the recommendation of the UK national curriculum.

Figure 1: *Game M (Recognition) for Digital (left) and Cardboard (right) version.*

Figure 2: *Game S (Recall) for Digital (left) and Cardboard (right) version.*

In Game M, the child had to choose the correct answer from the three options presented in a column on the right of the screen to match the total number of objects (i.e., ducks, fish, bees, apples) on the left (Figure 1). In Game S, the child was required to sort the numbers correctly into the empty compartments of a train at the bottom of the screen (Figure 2). A non-player character (NPC) is present throughout the game; two gender options are offered to players to choose at the beginning of the game because children of this age tend to prefer gender-oriented activities [28]. The instructions (e.g. "Please choose" when presented the NPC girl/boy options), game activity questions (e.g. "How many apples do you see?" in Game M; "Fill in the missing numbers." in Game S) and feedback (e.g. "Well done!") were read out loud by a pre-recorded voice in the DEG. For the cardboard game the experimenter would read out loud in real-time the same lines of instructions.

3.2 Instruments

3.2.1. Eye-tracker
The eye tracker used for this study was Tobii X2-30 and was incorporated with a mobile device stand (MDS), supporting a tablet where the DEG was installed and a camera device (Figure 3). The mobile eye tracker that records eye-tracking data was placed below the participant, allowing her or him to play in a normal way with the tablet. The MDS has eight standard configurations for Tobii X2-30 which

can be applied dependent on the device type (tablet) used and the participant's height [54]. In this study, the configuration C2 was used since the participants in this study were young children who are typically short.

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 has to be in the range from 60 to 65cm allowing a maximum gaze angle of 31º (Figure 4, [54]).

Parameters in X config.tool.	C1 / C2
Display area: Width (1)	14 cm
Display area: Height (2)	14 cm
Display area: Angle (3)	0 º
Eye tracker: Angle (4)	-5,2º
Eye tracker: Horizontal difference (5)	-8 cm
Eye tracker: Vertical difference (6)	5,7 cm

Figure 3: *Configuration for the Mobile Device Stand in this study. Source from Tobii Pro [54].*

The distance is indicated via the track status window on the Tobii Studio software. The calibration process then continues by going through the calibration points which are indicated with numbers from 1 to 5 on the calibration plate.
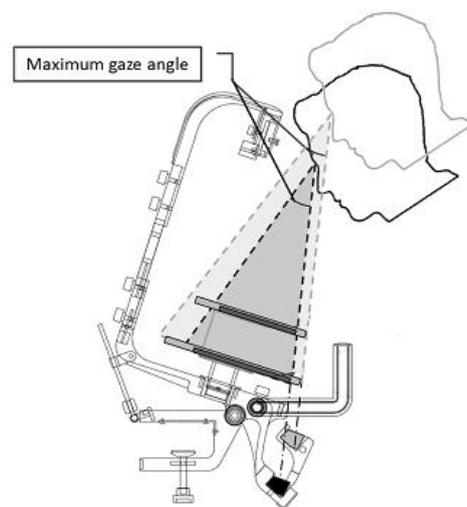

Figure 4: *The gaze angle range in this study is represented with the dark grey area. Image source from Tobii Pro [54].*

3.2.2. Paper-based test
To evaluate the learning effect, a paper-based test with five matching (recognition) tasks (maximum score = 5) and five sorting (recall) tasks (maximum score = 7) similar to those in the game was developed. The same paper test was used for the pre- and post-test which were completed by the participants before and after the gameplay session.

3.3 Pilot Studies

A series of pilot studies were conducted before the main study to improve the instruments, including the eye-tracking setup, the paper-based test and both educational game versions. The first pilot study involved five children in a non-school environment; it focused on the eye-tracking set-up for young children and the game design. Issues that were identified in the pilot were improved for the main study. The main issues identified were:

Lighting: Reflections from ceiling lights prevented the eye tracker from reading or recording a participant's pupil (eye). To allow smooth recording of the gameplay, the eye tracker device had to be located away from direct celling lights to avoid reflections beaming into the eye tracker. In addition, dark areas were also avoided so as to allow the camera to capture participants' facial expressions.

Body features. In some cases, a five-year-old may have small body features. One of the participants was unable to reach the tablet (DEG) around the mobile device stand (MDS) due to the participant's petite body features. A solution was to remove the handles of the MDS to allow the participant to be within the reach of the tablet. Unfortunately, this particular participant still had another problem, the child's body was now covering the eye tracker surface after removing the handles. Due to this, the participant had to be eliminated from the study. However, in the second pilot study, 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. The implication is to be discussed later.

Calibration plate. The original calibration plate that came with the MDS (Figure 5) and eye tracker was not feasible for the young participants. The small number indicators on the plate were not attractive enough for these young participants to focus on. The lack of constant attention on the calibration plate made the calibration process more difficult because not enough calibration data was gathered by the software to measure and evaluate individual participant's pupil (eye) characteristics.

Modification of the calibration plate was done for the second pilot study by replacing the small number indicators with coloured stars (i.e., red, blue, yellow, green, black) glued on to the calibration plate. Colour indicators were preferable since 5-year-olds were more familiar with colours. The children were asked to identify the colours before the calibration process to make sure they recognised each star colour.
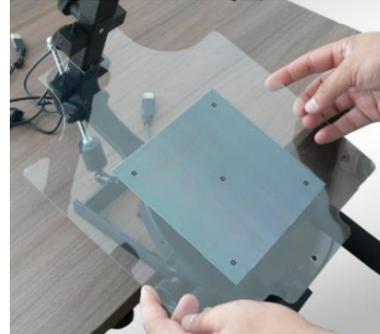

Figure 5: *Calibration Plate used in this study. Source from Tobii Pro [54].*

Calibration duration. The calibration process was a time-consuming process dependent on individual participants' characteristics and cooperation. The process started by measuring the distance between the eye-tracker device and the participant. Next the participant was asked to follow the calibration points; this step was error-prone when the points could not draw the participant's attention. If calibration data collected was not enough, another round of calibration had to be done again. This repeated process caused the young participant to lose interest in the task, given their short attention span.

Game design flaws. According to Peirce [37], young children's attention span is very short, however, the exact duration is 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 (~ 2 – 7 minutes (cf. the focused attention span in Section 2.3), depending on the child's characteristics) to avoid fatigue in the participants, but it was still grounded in Shi and Shih's GBL design model [52]. The first phase of the DEG design had a few flaws which were identified and improved. The flaws included the navigation assistance, interaction and feedback.

The second pilot study involved 31 children from a local school in England. It was of a larger scale than the first pilot study and thus provided more comprehensive input to the main study (Section 4). The study was conducted in a dedicated meeting room separate from the classroom to minimise distractions such as concurrent classroom activities. Issues identified and subsequently improved are described in the following.

Synchronizing video capture. Initially, the video capture was performed through an external camera

situated next to the MDS. Unfortunately, the eye-tracker software did not allow external video files to be imported into the software. This created complication when synchronising the video capture data with the eye tracking data. The video capture was essential for providing supplementary data such as children's facial expressions. To address this issue, the external camera was replaced with an embedded laptop camera which allowed video files and eye-tracker data simultaneously saved into the Tobii Studio software to avoid data inconsistency problems. Nonetheless, this benefit of synchronicity is gained at the expense of the flexibility of the external camera angles.

Improved DEG and cardboard design flaws. The improved DEG version was tested in the second pilot study. The navigation assistance remained to be ignored by the children. This unawareness caused misunderstanding, particularly for Game S. Hence, a demo page was created and included in Game S. In addition, other navigation objects were animated on every page to draw the child's attention to them. Apart from the DEG design, the cardboard was also tested. For the cardboard game, instructions were manually given by the experimenter. Since the cardboard game was played on the MDS device, some modification had to be made. For instance, the number boxes had to be glued with blu tack to allow the boxes to stay and be pulled out while playing.

Paper-based test confusion. Another issue investigated in the second pilot study was the design of the pre and post-test. Initially, the test questions were arranged and squeezed into a single page. However, this structure confused the children who could not focus on one question at one time.
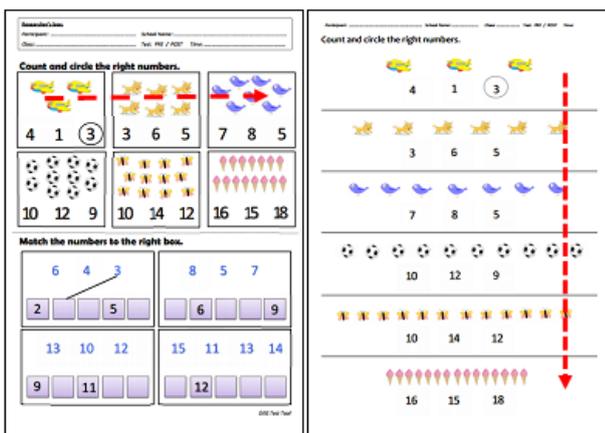

Figure 6: *The paper-based test before and after the re-arrangement.*

Hence, the presentation of the questions was rearranged, displaying each question horizontally and distributing the questions over two pages with appropriate spacing (Figure 6). The way of revealing the test questions to the participants was also adapted to allow them to focus on one question at one time.

4. MAIN STUDY

4.1 Participant and Procedure

This study was conducted in a local school in England for seven weeks. The experimental sessions were carried 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 5 and were in the foundation (playgroup) stage. Each session took place in an uninterrupted room in the school where 59 children played the DEG, and 35 played the cardboard game.

For every experimental session, a child played either a DEG or cardboard game in front of the MDS while the eye-tracker and the built-in camera of the laptop captured data into the Tobii Studio software. The session began by performing the calibration process which proved challenging for some cases as it was not easy for a child to listen to instructions. Repeated calibrations had to be done when not enough calibration data was collected. Throughout the session, the experimenter was present in the room to observe the child's interaction behaviour and provided help when necessary. For the duration of the cardboard session, the experimenter's presence was essential as she played a role in the gameplay by delivering feedback to the child.

The average duration of the experimental session for the DEG and that for the cardboard version were 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 could be accounted for by the other activities – greeting and seating the participant, giving the instructions, filling out the questionnaire, and debriefing. Overall, the range of the duration of gameplay was within the attention span of 5-year-olds [47]. Details are presented in Results and Discussion.

5. DATA ANALYSIS

5.1 Scene Segmentation

Each participant's recordings were divided into short scenes based on the pages of the games. In total there were 13 scene pages, including the welcome, selection, demonstration and thank-you pages, per participant for both games, but only the game scene pages M1 to M4 (Game M) and S1 to S4 (Game S) were analysed for this study. M1 to M4 were pages associated with the matching activities (i.e.

recognition) whereas S1 to S4 were pages related to the sorting activities (i.e. recall).

Figure 7 shows the scenes taken from a participant's eye-tracking recording of Game S. At the bottom left side of the screen, one can view the video capture of the session (it is made blurry for the sake of anonymity), which helped identify unrecorded or missing data.
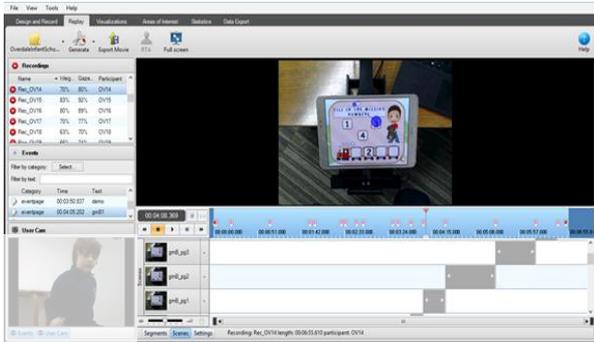

Figure 7: *Divided scenes taken from a participant's recording using Tobii Studio*

5.2 Area of Interest Positions

In the eye-tracking methodology, Areas of Interest (AOIs) are used to link eye movement measures (e.g. fixation duration) to parts of the stimulus displayed. The AOI statistics facilitates the interpretation of eye-tracking data [19]. According to Orquin [32], 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 (e.g., a usability test of a website design [12]) ; (ii) based on quality criteria where the stimuli design allows maximising the distance between objects and high accuracy is required (e.g., a research study on visual cognition [43]). Due to the experimental stimuli design in this study where overlapping of AOIs may happen, a smaller AOI margin ($\approx 0^{\circ}$ margin) was used to balance the proportion of fixations [32]. AOIs were defined based on relevant (e.g. counting objects) and irrelevant images (e.g. non-player character).

Once the scene segmentation process was completed on each participant's record, the AOIs were carefully defined and positioned for each page of Game M and Game S. 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 8. 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.
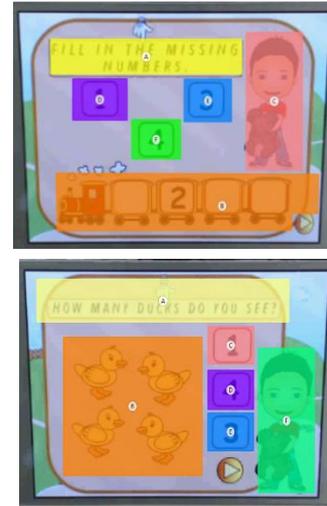

Figure 8: *Defined Areas of Interest (AOIs) for Game M (lower) and Game S (upper)*

5.3 In-Game Performance

Apart from analysing the pre- and post-test difference in studying the learning effect of both the DEG and cardboard game, the in-game performance of the participants was also examined. The in-game score was calculated as 1 for every correct answer and 0 for every wrong answer. The maximum score for Game M was 4 and for Game S was 12. Pages were divided into easy and hard tasks. The easier tasks presented in page 1 and page 2 involved numbers below and equal to 10 and the harder tasks presented in pages 3 and 4 were associated with numbers above 10 up to 20. As the game design was based on the EYFS standard for foundation classes the numbers used in the tasks did not go beyond number 20.

6. RESULTS

6.1 Measures of learning effects

As the data are not normally distributed (Shapiro-Wilk test, $p<.05$), the non-parametric Spearman correlation tests were performed to assess the relationships between the pre- and post-based tests and in-game performance for both the DEG and cardboard game. For the DEG there were highly 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 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: $r = 0.63$). These results suggest that the participating children applied their knowledge of numeracy consistently across the evaluative situations.

6.2 Role of the learning media

In studying the learning effect of both learning media, the DEG and cardboard game, the non-parameter Wilcoxon signed-ranks tests were used. Results of the DEG (Mdn_{pre}=7.00, range = 4.00 – 11.00; Mdn_{post}=8.00, range = 4.00 –11.00) indicated that no significant improvement in learning was achieved (Z=1.10, p=0.27). However, for the cardboard game (Mdn_{pre}=9.00, range = 5.00 –11.00; Mdn_{post}=10.00, range = 6.00 –12.00), the Wilcoxon signed-ranks tests indicated that there was a significant improvement in knowledge (Z=1.90, p=0.05). These results could be attributed to the typical learning environment of 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 [14]. 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.

6.3 Recall and Recognition

The Wilcoxon signed-ranks tests were performed between the recognition-based Game M and recall-based Game S with respect to the in-game performance of each individual level. There was a significant difference for both easy-level tasks on page 1 and page 2 between Game M and Game S which involves numbers 1 to 10 (Table 1). For the hard-level tasks involving numbers 11 to 20, there were also significant difference for both page 3 and page 4 between Game M and Game S (Table 1)

Table 1: Medians and ranges of in-game score for the recognition (Game M) and recall game (Game S) for each game level.

Pages	*Z , p*	Game	Median	Range
Easy-level 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-level 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

The findings showed that the children had more difficulty in answering the recall-based Game M as compared to the recognition-based Game S. They

corroborated the theoretical model that as recall entails a deeper level of information retrieval than recognition [4], it is more challenging.

6.4 Attention to relevant and irrelevant objects

In studying the attention paid to relevant and irrelevant object on the DEG for Game M and Game S, the Mann-Whitney tests were used to evaluate the relation between the levels of achievement (low or high) and the fixation durations for relevant and irrelevant objects. The threshold for classifying low (n=30) and high achievers (n=29) was the average pre-test performance of the children (M=7.19, SD=3.57).

In page 1 of Game M (Table 2), the test indicated that higher achievers paid significantly more attention to relevant object *D* (answer) than low achievers (U=295, p=.03). Likewise for page 4 of Game M results of the Mann-Whitney test showed that higher achievers paid more attention to relevant object *C* (answer) than for low achievers did (U=312, p=.05). The results are consistent with the significant correlations between the pre-test and in-game scores (Section 6.1). As it is logical that one gazes at an object when picking it, high achievers attended to (or fixated at) the correct answers longer than low achievers. However, counterintuitively, no significant differences were found on page 2 and page 3 of Game M for all relevant and irrelevant objects. We are unable to explain the incongruous pattern of the results across the four pages.

Table 2: Medians and ranges of fixation duration on the objects of all pages in Game M. (p<.05).*

Game M page	Achiever n_{low} = 30 n_{high} = 29	Relevant Objects (second)	Irrelevant Objects (second)
		Object D*	*Object F*
M1	*Low*	*0.42 (.23 –.65)*	*0.00 (.00 -.34)*
	High	*0.57 (.47 –.97)*	*0.00 (.00 -.28)*
		Object E	*Object F*
M2	*Low*	*0.41 (.00 –.61)*	*0.00 (.00 -.00)*
	High	*0.37 (.26 –.70)*	*0.00 (.00 -.00)*
		Object D	*Object F*
M3	*Low*	*0.45 (.22 –.57)*	*0.00 (.00 -.00)*
	High	*0.47 (.17 –.73)*	*0.00 (.00 -.31)*
		Object C *	*Object F*
M4	*Low*	*0.40 (.00 –.69)*	*0.00 (.00 -.04)*
	High	*0.65 (.27 –.89)*	*0.00 (.00 -.00)*

Furthermore, the zero fixation durations indicate that low achievers did not look at the irrelevant object *F*

(i.e. non-player character) at all on page 2 or page 3, and the negligible number of them looked at *F* on page 4. Similarly, high achievers did not look at *F* on page 2 or page 4. Only a small number of both groups of achievers, probably due to its novelty, looked at *F* on page 1, but with a relatively short duration. With *F* remaining unchanged, its attractiveness, already quite low, waned from page 1 to page 4. However, a handful of high achievers did look at *F* on page 3. It might be triggered by the change of difficulty level from page 2 to page 3 and some children might want to check if *F* was changed as well.

As for the attention paid to the objects in Game S (Table 3), given its different design, the Mann-Whitney tests for pages 2, 3 and 4 indicated that higher achievers paid less attention to relevant object *B* than for low achievers (Table 3), U_{page2}=230, p_{page2}<.05; U_{page3}=299, p_{page3}=.04; U_{page4}=220, p_{page4}<.05. However, no significant differences were found on page 1 (i.e. the easiest level) of Game S for all relevant and irrelevant 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 1. Additionally, those children could also be attracted to the dynamic object *B* 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*.

Table 3: Medians and ranges of fixation duration on the objects of all pages in Game S (p<.05)*

Game S Page	Achiever n_{low} = 30 n_{high} = 29	Relevant Objects (second)	Irrelevant Objects (second)
		Object B	Object C
S1	Low	0.68 (.49 –.82)	0.16 (.00 -.36)
	High	0.57 (.43 –.68)	0.00 (.00 -.37)
		Object B*	Object C
S2	Low	0.66 (.53 -.76)	0.00 (.00 -.48)
	High	0.46 (.41 -.61)	0.00 (.00 -.41)
		Object B*	Object C
S3	Low	0.65 (.45 -.78)	0.14 (.00 -.38)
	High	0.51 (.42 -.61)	0.00 (.00 -.10)
		Object B *	Object C
S4	Low	0.66 (.53 -.86)	0.00 (.00 -.28)
	High	0.52 (.44 -.62)	0.00 (.00 -.36)

Furthermore, a pattern for irrelevant object *C* (non-player character) similar to Game M was observed for Game S. A zero median fixation duration was for all four pages for high achievers and for page 2 and page 4 for low achievers. A handful of low achievers looked at *C* on page 1 and page 3 with a short duration. Overall, the patterns of fixation durations for the different game interface objects suggest that low and high achievers applied different strategies to deal with the tasks given.

In addition, fixation duration is a proxy for focused attention (Section 2.3). To estimate it for the actual gameplay, we measured the total fixation durations over the four pages of Game M and Game S. Results are shown in Table 4; the total duration was found to be 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. Note also that there are five other pages – introduction and demo, which we have not analysed for fixation duration.

Table 4: Total fixation durations (in seconds) over four pages of Game M and Game S.

Game M					Total
	M1	M2	M3	M4	All
Mean	8.37	9.36	11.42	11.82	40.97
SD	5.87	6.25	8.78	10.70	31.6
Game S					Total
	S1	S2	S3	S4	All
Mean	11.44	10.15	11.35	12.40	45.34
SD	6.26	5.86	7.46	7.87	27.45

7. GENERAL DISCUSSION

In this section we reflect on the issues encountered during the empirical process and their implications.

Better education support. Results of the DEG learning effect may be due to children having different tutoring support and/or better home education which allows them to know numeracy better than others. Nonetheless, the design of this study followed the EYFS curriculum standard; accordingly children should be able to count from 1 to 20 by the end of 5 years of age [10]. The extraneous influences of the social factors are beyond control of this study. Ideally, if we could recruit a much bigger sample of children from foundation years, the impact of the confounding variables would be mitigated. .

Participant Recruitment. Recruiting schools and young children for this study was very challenging. From the organisational perspective, it was difficult to get schools involved since taking individual

children out of regular classes usually is not appreciated as it may interfere their learning programme. Furthermore, the time constraint imposed by the calendar of school terms impedes schools from accommodating research studies that may take 2 to 3 weeks to complete, especially when the school involved needs to provide resources (e.g., a separate room) to support the 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.

Lack of established publications on attention span. As mentioned in Section 2.2, some grey literature has discussed the decrease of attention span among young children due to the use of technology [36,38]. 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 4) could be considered as a means to estimate focused attention of 5-years-old, our study, unlike [47], 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 relevant data and with the systematic manipulation of the influencing factors (e.g. the interaction with the experimenter).

Overall, we 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 engagement versus loss of data. The use of the mobile device stand (MDS) (Section 3.2.1) 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 4) and can cause loss of data. Contrarily, giving repeated reminders to sit appropriately to a child during the gameplay might agitate them for being nagged and controlled, causing them to lose interest in the game.

From the practical experience of the study reported above, 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

this 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. .

8. CONCLUSION

To summarise our work, here we revisit the four RQs raised earlier and report the findings as follows:

RQ1: No significant difference improvement was achieved via the DEG but a significant improvement in learning was found through the cardboard game.
RQ2: High achievers paid attention to relevant object (Game M) and set strategies in the gameplay (Game S).
RQ3: Recall game task proved to be more difficult than recognition for both the digital and cardboard game versions.
RQ4: The calibration process and the trade-off between data loss and gameplay engagement were among the challenges of the eye-tracking methodology for young children.

Notwithstanding the limitations, our study contributes to the applied body of knowledge on young children's learning via DEGs and the methodological challenges as well as resolutions for applying the eye-tracking methodology for young children.

ACKNOWLEDGEMENTS

REFERENCES

1. Fashina Aladé, Alexis R Lauricella, Leanne Beaudoin-Ryan, and Ellen Wartella. 2016. Measuring with Murray: Touchscreen technology and preschoolers' STEM learning. *Computers in Human Behavior* 62: 433–441.

2. Anissa All, Elena Patricia Nuñez Castellar, and Jan Van Looy. 2016. Assessing the effectiveness of digital game-based

learning: Best practices. *Computers & Education* 92–93: 90–103.

3. Nicola Ariasi and Lucia Mason. 2011. Uncovering the effect of text structure in learning from a science text: An eye-tracking study. *Instructional Science* 39, 5: 581–601.

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

5. Andreas Bulling, Andrew T Duchowski, and Paivi Majaranta. 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*, 627–628.

6. Andreas Bulling and Hans Gellersen. 2010. Toward Mobile Eye-Based Human-Computer Interaction. *Pervasive Computing, IEEE* 9, 4: 8–12.

7. Andreas Bulling and Daniel Roggen. 2011. Recognition of visual memory recall processes using eye movement analysis. *Proceedings of the 13th international conference on Ubiquitous computing*, 455–464.

8. Cathy Burnett. 2010. Technology and literacy in early childhood educational settings: A review of research. *Journal of Early Childhood Literacy* 10, 3: 247–270.

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

10. Department for Education (DfE). 2013. Early Years Outcomes: A non-statutory guide for practitioners and inspectors to help inform understanding of child development through the early years. 1–34. Retrieved from https://www.foundationyears.org.uk/files/2012/03/Early_Years_Outcomes.pdf.

11. Jonathan Doherty and Malcolm Hughes. 2009. *Child development: theory and practice 0-11*. Longman, Harlow.

12. Sukru Eraslan, Yeliz Yesilada, and Simon Harper. 2015. Eye tracking scanpath analysis techniques on web pages: A survey, evaluation and comparison. *Journal of Eye Movement Research* 9(1), 2: 1–19.

13. Séverine Erhel and Eric Jamet. 2013. Digital game-based learning: Impact of instructions and feedback on motivation and learning effectiveness. *Computers & Education* 67: 156–167.

14. Eva L Essa. 2012. *Introduction to early childhood education*. Cengage Learning.

15. Joe L Frost, Sue Clark Wortham, and Robert Stuart Reifel. 2008. *Play and child development*. Pearson/Merrill Prentice Hall Upper Saddle River, NJ.

16. Tsai Fu-Hsing, Yu Kuang-Chao, and Hsiao Hsien-Sheng. 2012. Exploring the Factors Influencing Learning Effectiveness in Digital Gamebased Learning. *Journal of Educational Technology & Society* 15, 3: 240–250.

17. David Furió, M-C Juan, Ignacio Seguí, and Roberto Vivó. 2015. Mobile learning vs. traditional classroom lessons: a comparative study. *Journal of Computer Assisted Learning* 31, 3: 189–201.

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

19. Roy S Hessels, Chantal Kemner, Carlijn van den Boomen, and Ignace T C Hooge. 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.

20. Kenneth Holmqvist, Marcus Nyström, Richard Andersson, Richard Dewhurst, Halszka Jarodzka, and Joost Van de Weijer. 2011. *Eye tracking: A comprehensive guide to methods and measures*. OUP Oxford.

21. Cheng-Yu Hung, Fang- O Kuo, Jerry Chih-Yuan Sun, and Pao-Ta Yu. 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.

22. Jane Johnston. 2005. *Early Explorations in Science 2nd Edition*. McGraw-Hill, Inc.

23. Meng-Lung Lai, Meng-Jung Tsai, Fang-Ying Yang, et al. 2013. A review of using eye-tracking technology in exploring learning from 2000 to 2012. *Educational Research Review* 10: 90–115.

24. Lorraine Lander. 2002. Sustained Attention, Self-regulation, and Motivation During a Maze Task for Preschool Children: A Motivational, Cognitive Analysis of the Process of Sustained Interest. Retrieved

from
https://commons.esc.edu/lorrainelander/wp-content/uploads/sites/182/2011/07/Dissertation-Lorraine-Lander.pdf.

25. Eric Zhi Feng Liu and Po-Kuang Chen. 2013. The Effect of Game-Based Learning on Students' Learning Performance in Science Learning – A Case of "Conveyance Go." *Procedia - Social and Behavioral Sciences* 103: 1044–1051.

26. Han-Chin Liu and Hsueh-Hua Chuang. 2011. An examination of cognitive processing of multimedia information based on viewers' eye movements. *Interactive Learning Environments* 19, 5: 503–517.

27. Päivi Majaranta and Andreas Bulling. 2014. Eye tracking and eye-based human–computer interaction. In *Advances in Physiological Computing*. Springer London, 39–65.

28. Panos Markopoulos, Janet C Read, Stuart MacFarlane, and Johanna Hoysniemi. 2008. *Evaluating children's interactive products: principles and practices for interaction designers*. Morgan Kaufmann.

29. Lucia Mason, Patrik Pluchino, Maria Caterina Tornatora, and Nicola Ariasi. 2013. An Eye-Tracking Study of Learning From Science Text With Concrete and Abstract Illustrations. *The Journal of Experimental Education* 81, 3: 356–384.

30. Lucia Mason, Maria Caterina Tornatora, and Patrik Pluchino. 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.

31. Lisa M. Oakes, Shannon Ross-Sheehy, and Kathleen N. Kannass. 2004. Attentional engagement in infancy: The interactive influence of attentional inertia and attentional state. *Infancy* 5, 2: 239–252.

32. Jacob L. Orquin, Nathaniel J.S. Ashby, and Alasdair D.F. Clarke. 2016. Areas of Interest as a Signal Detection Problem in Behavioral Eye-Tracking Research. *Journal of Behavioral Decision Making* 29, 2–3: 103–115.

33. Ku Oskar, Y Chen Sherry, H Wu Denise, C C Lao Andrew, and Chan Tak-Wai. 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.

34. Marina Papastergiou. 2009. Digital Game-Based Learning in high school Computer Science education: Impact on educational effectiveness and student motivation. *Computers & Education* 52, 1: 1–12.

35. Alison E Parker, Erin T Mathis, and Janis B Kupersmidt. 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.

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

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

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

39. Jean Piaget. 1971. The theory of stages in cognitive development. In *Measurement and Piaget*. McGraw-Hill, New York, NY, US, ix, 283-ix, 283.

40. Hector R Ponce and Richard E Mayer. 2014. An eye movement analysis of highlighting and graphic organizer study aids for learning from expository text. *Computers in Human Behavior* 41: 21–32.

41. Marc Prensky. 2001. *Digital game-based learning*. McGraw-Hill, New York;London;

42. Keith Rayner. 2009. Eye movements and attention in reading, scene perception, and visual search. *The Quarterly Journal of Experimental Psychology* 62, 8: 1457–1506.

43. Keith Rayner and Eyal M Reingold. 2015. Evidence for direct cognitive control of fixation durations during reading. *Current Opinion in Behavioral Sciences* 1: 107–112.

44. Janet C Read and Stuart MacFarlane. 2006. Using the fun toolkit and other survey methods to gather opinions in child computer interaction. *Proceedings of the 2006 conference on Interaction design and children*, ACM, 81–88.

45. Anna Richter and Mary L. Courage. 2017. Comparing electronic and paper storybooks for preschoolers: Attention, engagement, and recall. *Journal of Applied Developmental Psychology* 48: 92–102.

46. Holly A. Ruff and Mary C. Capozzoli. 2003. Development of attention and distractibility in the first 4 years of life. *Developmental Psychology* 39, 5: 877–890.

47. Holly A Ruff, Mary Capozzoli, and Renata Weissberg. 1998. Age, individuality, and context as factors in sustained visual attention during the preschool years. *Developmental Psychology* 34, 3: 454–464.

48. Holly A Ruff and Katharine R Lawson. 1990. Development of sustained, focused attention in young children during free play. *Developmental Psychology* 26, 1: 85–93.

49. Andrew Schall and Jennifer Romano Bergstrom. 2014. Introduction to Eye Tracking. In J.R.B.J. Schall, ed., *Eye Tracking in User Experience Design*. Morgan Kaufmann, Boston, 3–26.

50. 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.

51. Hsiao-Ching She and Yi-Zen Chen. 2009. The impact of multimedia effect on science learning: Evidence from eye movements. *Computers & Education* 53, 4: 1297–1307.

52. Yen-Ru Shi and Ju-Ling Shih. 2015. Game factors and game-based learning design model. *International Journal of Computer Games Technology* 2015: 11.

53. Han-Yu Sung and Gwo-Jen Hwang. 2013. A collaborative game-based learning approach to improving students' learning performance in science courses. *Computers & Education* 63: 43–51.

54. Tobii AB. 2016. *Tobii Pro Mobile Device Stand User ' s Manual*. .

55. Meng-Jung Tsai, Huei-Tse Hou, Meng-Lung Lai, Wan-Yi Liu, and Fang-Ying Yang. 2012. Visual attention for solving multiple-choice science problem: An eye-tracking analysis. *Computers & Education* 58, 1: 375–385.

56. Jui-Mei Yien, Chun-Ming Hung, Gwo-Jen Hwang, and Yueh-Chiao Lin. 2011. A Game-Based Learning Approach to Improving Students' Learning Achievements in a Nutrition Course. *Turkish Online Journal of Educational Technology-TOJET* 10, 2: 1–10.

57. Nicholas Zaranis, Michail Kalogiannakis, and Stamatios Papadakis. 2013. Using Mobile Devices for Teaching Realistic Mathematics in Kindergarten Education. *Creative Education* 4, 7: 1–10.