Central Lancashire Online Knowledge (CLoK)

Title	Using the memoline to capture changes in user experience over time with children
Type	Article
URL	https://clok.uclan.ac.uk/15268/
DOI	https://doi.org/10.1016/j.ijcci.2016.07.001
Date	2016
Citation	Sim, Gavin Robert, Nouwen, M, Vissers, J, Horton, Matthew Paul leslie, Slegers, K and Zaman, B (2016) Using the memoline to capture changes in user experience over time with children. International Journal of Child-Computer Interaction, 8. pp. 1-14. ISSN 2212-8689
Creators	Sim, Gavin Robert, Nouwen, M, Vissers, J, Horton, Matthew Paul leslie, Slegers, K and Zaman, B

Accepted Manuscript

Using the Memoline to capture changes in user experience over time with children

Gavin Sim, Marije Nouwen, Jorick Vissers, Matthew Horton, Karin Slegers, Bieke Zaman

Please cite this article as: G. Sim, M. Nouwen, J. Vissers, M. Horton, K. Slegers, B. Zaman, Using the Memoline to capture changes in user experience over time with children, *International Journal of Child-Computer Interaction* (2016), http://dx.doi.org/10.1016/j.ijcci.2016.07.001

Using the MemoLine to Capture Changes in User Experience Over Time with Children

Gavin Sim[1*], Marije Nouwen[2], Jorick Vissers[2], Matthew Horton[1], Karin Slegers[2], Bieke Zaman[2].

1 Affiliation

ChiCI Group

University of Central Lancashire

Preston, UK

PR1 2HE

2 Affiliation

CUO | Social Spaces

KU Leuven / iMinds

Parkstraat 45, bus 3605

3000 Leuven

Belgium

*Corresponding Author grsim@uclan.ac.uk

Using the MemoLine to Capture Changes in User Experience Over Time with Children

Abstract

In this paper, we focus on the MemoLine, a retrospective tool for capturing changes in long-term user experience of games with children, which has had little attention from the Child Computer Interaction community. To investigate the appropriateness of the MemoLine, two studies were performed. In the first study, 16 children aged 7 to 12 were instructed to use the MemoLine at home to reflect on their 4 month experience with a music game. The second study took place in a school context, with 32 children aged 10 to 11 who used MemoLine to report on their 3 month experience with an educational game. The results suggested that children along the age spectrum of 7 to 12 were able to complete the MemoLine Instruments. In the two different contexts children were able to recall experiences relating to the game and provide data that would be useful for developers to understand how and why their experiences changed over time. Finally, the results showed that the procedural choices for the data gathering could be adjusted to a home and school context. Based on the insights from the case studies, best practices are defined to facilitate the use and further development of the method.

Keyboards: Evaluation; User Experience; Children; Games; Longitudinal Studies;

1. Introduction

There has been considerable research published within both the Human Computer Interaction (HCI) and Child Computer Interaction (CCI) communities on user experience (Vermeeren et al., 2010; Xu, Read, Sim, McManus, & Qualter, 2009). According to the International Organization for Standardization (ISO) user experience is defined as 'a person's perceptions and responses that result from the use or anticipated use of a product or service' (ISO, 2010). Meaningful measurement criteria of user experience may focus, amongst others, on physical, sensual, emotional and/or aesthetic aspects of a users' interaction with systems. User experience is more subjective in nature than usability and therefore can present more of a challenge to measure. In the area of CCI, many evaluation methods have emerged over the years to capture user experiences with children, including evaluation methods that rely on survey (Barendregt, Bekker, & Baauw, 2008; Read, MacFarlane, & Casey, 2002), interview (Zaman & Abeele, 2010), and observational techniques (Sim, MacFarlane, & Horton, 2005). These evaluation methods have been devised to measure momentary user experiences; they have not been conceptualized to capture experiences of technology or software over a prolonged period of time. Long-term user experience studies are rather rare within CCI and HCI (Moschou & Zaharias, 2013). In a study analyzing user experience methods used with adults, it was found that 34 out of 95 methods were suitable for long-term usage (Vermeeren et al., 2010). This highlights the fact that the majority of adult methods are not suitable for measuring long-term user experience and it is conjectured within CCI there will be few applicable methods. Long term user experience measurements provide complementary insights to momentary measurements, for instance by yielding an understanding of changes in user experiences and informing the prediction of long term product evaluations (Karapanos, Zimmerman, Forlizzi, & Martens, 2009).

Previous literature has reported on three dominant approaches to understanding changes in user experiences over time: 1) Cross-sectional research designs; 2) longitudinal research designs using pre-

post test, or more data gathering moments through repeated measures; and 3) retrospective recall (Kujala, Roto, V, et al., 2011; von Wilamowitz Moellendorff et al. 2006). Changes in user experience over time have been classified by von Wilamowitz Moellendorff et al. (2006) based on different resolutions, those being, a micro perspective (e.g. a few hours), a meso perspective (e.g. weeks) and a macro perspective (e.g. years)

Firstly, in a cross-sectional research design, changes in user experience are being evaluated based on their level of expertise, for example novices versus experts. In a study by Lozonder et al. (2000) they used a cross-sectional study to understand the difference between novices and experienced users in searching the internet and found that the experts were superior in locating websites. Clearly, the long-term use of the system impacts positively on the performance. Cross-sectional studies are rare within CCI. Expertise within the context of CCI research has mainly focused on differences in children based on age. For example, Couse and Chen (2010) showed that engagement with technology can vary across age spectrums.

Secondly, in longitudinal research, pre-post studies use the same participants at two points in time to capture momentary experiences. Karaponos et al. (2008) studied the changes in user experience of 10 individuals using a novel pointing device over a 4 week period, capturing data through a questionnaire in week 1 when the device was introduced and again after 4 weeks of use. Pre-post studies have been criticised, as they only take two momentary measurements, and therefore interpolation is used to understand the changes in experiences between the two measures. Contrary to pre-post studies, repeated measures take more than two measurements, using a variety of techniques to capture data on the same variable and with the same participants on a number of different occasions. These techniques include questionnaires, self-report and psycho-physiological measures. However, longitudinal studies with repeated measures have been criticised for being laborious in execution (Karapanos, Martens, & Hassenzahl, 2012).

Finally, retrospective measures capture (changes in) user experience data in a single contact, and hence they are more likely to yield insight in overall memories that are more than the simple sum of momentary experiences. Even though memory biases may occur with the use of retrospective methods, the relevance of memories cannot be underestimated when studying user experience over a prolonged period of time. As people reflect on memories of past experiences, and report them to others, these memories have the power to guide their future behaviour (Norman, 2009).

Many retrospective studies in HCI have reported on the use of self-reporting techniques, and to facilitate this, various visual and /or internet-mediated tools have been developed including DrawUX (Varsaluoma & Kentta, 2012) and iScale (Karapanos, Zimmerman, Forlizzi, & Martens, 2010). In addition to these tools, interview methods have been devised including UX Curve (Kujala, Roto, Väänänen-Vainio-Mattila, & Sinnelä, 2011) and Corpus (Von Wilamowitz-Moellendorf, Hassenzahl, & Platz, 2006), which was used to evaluate adults' pragmatic and hedonic experiences with interactive products over time. Within CCI, the MemoLine (Vissers, De Bot, & Zaman, 2013) has been synthesized based on the UX Curve (Kujala, Roto, Väänänen-Vainio-Mattila, Karapanos, & Sinnelä, 2011) to enable children to report their experiences over time in a retrospective way. The MemoLine has had limited application and therefore would benefit from insights in best practices on how to use this tool, and a critical understanding to assess how appropriate and meaningful it is. In line with the observation that research benefits from replication studies, which happens to be very scare in HCI research (Hornbæk et al. (2014),

we argue that as far as the development of methods concerns, a number of iterations is necessary to fully understand their potential and limitations, and refine them where necessary.

With this paper, we aim to contribute to the knowledge about the MemoLine for evaluating changes in user experience over time with children. von Wilamwitz Moellendorff et al. (2006) suggest there are three types of longitudinal studies. The focus of the research presented in this paper follows the Meso Perspective (weeks rather than hours or years). In order to assess the appropriateness and meaningfulness of the MemoLine in a wide variety of contexts, we report on two case studies that span the involvement of a broad age spectrum (children between the ages of 7 to 12), varying in terms of setting of data gathering (home versus school), the interactive system under investigation (entertainment app versus educational game), the time span for the retrospective evaluation (9 versus 13 weeks), along with procedural differences.

2. User Experience Evaluation Methods and Children

Previous research has reported on a variety of empirical methods to evaluate user experience and/or fun in children, of which (a combination of) the following techniques are most common: survey tools, diaries, verbalization methods, and/or observation methods (Zaman et al., 2013).

Firstly, survey tools are primarily used within CCI to capture quantifiable data relating to user experience. Contrarily to research with adults, the administration of survey tools for product evaluation preferably takes a face-to-face approach in the case of child participants (Read, 2008). Within the category of survey tools, we can distinguish between techniques that measure the degree of an overall acceptance judgment based on the evaluation of a single system, versus techniques that aim for preference evaluations based on the comparison of two or more experiences (Zaman et al., 2013). For instance, the Smileyometer, one of the child-friendly survey tools of the Fun Toolkit (Read & MacFarlane, 2006) is one that yields a quantifiable overall acceptance evaluation. Contrarily, the Fun Toolkit's Again Again Table (Read & MacFarlane, 2006) and the This or That method (Zaman et al., 2013) yield a relative evaluative evaluation based on the comparison of two or more experiences. Although the majority of survey-based evaluation methods in CCI have been used to capture momentary experiences, they have the potential to be used within a repeated measures study to better account for changes in user experiences over time. Barendregt, Bekker, Bouwhuis, & Baauw (2006) have used the Smileyometer instrument in a pre-post research design that spanned over a 3 week period. These findings suggest that the Smileyometer may be a viable option for use in long-term user experience studies when employed in a pre-post study design, but further work would still be required to validate the tool within this context. Survey tools are not without their problems and there is concern over the reliability of the tools over time. In a study looking at children's technology ownership, a survey was designed and administered to the children on two occasions, a week apart, and it was shown that the reliability of responses was only about 50% (Horton, Read, & Sim, 2011).

Secondly, diary studies are common in user experiences evaluation studies in the field of CCI, both for understanding momentary experiences as changes in user experiences over time (Markopoulos, Read, MacFarlane, & Hoysniemi, 2008). There are various approaches to implement diaries in evaluation studies, for instance diaries can be in paper or digital form, and capture textual, verbal and/or visual

information (e.g., by encouraging the use of digital videos or cameras). Oftentimes, diaries include survey-elements or come with an interview afterwards in order to exploit the power of triangulation in measuring children's experiences with technology. Illustrating its use for understanding changes in experiences over time, Colombo and Landoni (2014) used diaries in the form of a questionnaire that needed to be completed at random intervals whilst children interacted with ebooks. The rationale for using the questionnaire was that they wanted to minimize the interference of the evaluation method on the user experience. The diaries were completed online and were in the form of a series of Likert style questions that were completed over a two-week period. The latter form of data capture would be more aligned to a repeated measure study. Diaries have the potential advantage of increasing the likelihood of retrospection, achieved by reducing the time elapsed between an experience and the account of this experience (Reis, 1994).

Finally, verbalisation methods have been introduced in various types of research designs and adapted in various ways to account for the particularities of evaluating children's user experiences. Some of the research designs focus on momentary experiences, whereas others are aimed at revealing memories and changes over times. Sometimes, interview-based methods are implemented in a research design that includes triangulation with other methods, like visual and observation-based methods. For instance, Barendregt et al. (2008)'s Problem Identification Picture Card (PIPC) method, stipulates that the researcher hands out picture cards to the participating child as visual cues to facilitate the child thinking out loud and reporting on emotions and momentary experiences. The child can choose to verbalize how he or she feels and/or select one or more picture cards portraying an emotional expression corresponding to how the child feels at that moment. This information is being complemented with the data from observing the child interacting with the system, hereby providing a richer data set to interpret the child's interactions. The PIPC protocol does not foresee in subsequent follow-up probing to understand the deeper reasons of certain experiences and emotions. In this respect, PIPC differs from for instance the User eXperience Laddering method (Zaman & Abeele, 2010). In the latter interview protocol, the interviewer is expected to ask a series questions to understand why children like or dislike a particular technology, aiming to reveal reasons for a particular product preference beyond the interface surface level up to the level of what the child finds important (i.e., desired consequences and values). Even though the User eXperience Laddering protocol is not focused on detecting changes in experiences, the interviewer may eventually reveal information that relates to experiences beyond that of the moment, for instance when children verbalize explanations relating to their expectations or previous memories. This contrasts the MemoLine (Vissers et al., 2013), which is, as far as we know, the only method in CCI to date that has been conceptualized with the primarily goal to retrospectively evaluate changes in children user experiences over time. The MemoLine tool consists of a number of timelines that the children colour in, using different colours to depict their changes in experiences over time. It also comes with a subsequent interview in which children are probed about the reasons for these changes. In the original study of the MemoLine by Vissers et al. (2013), the authors emphasised that they conducted an explorative study to develop and introduce the instrument, highlighting that further research is required to understand the MemoLine's appropriateness in different contexts. This article aims to address this call, and therefore sets out to assess the MemoLine in two different contexts. As previous research has shown that user experience is likely to change over time (Karapanos et al.,

2009), and capturing this would yield valuable information to HCI researchers and interaction designers, it is important to understand its potential inasmuch as its limitations.

In what follows, we will first discuss the UX Curve as this retrospective method served as a starting point for the development of MemoLine and then present two case studies of its application, followed by a reflection of the appropriateness of the tool.

3. Developing the MemoLine

3.1 UX Curve as a starting point for MemoLine

Karapanos et al. (2010) developed UX Curve as a paper instrument with the aim to support users in reporting when and why their experiences with technologies changed over a longer period of time. Figure 1 shows an example of a completed UX Curve. The time dimension is represented by the blank horizontal axis, starting from the beginning of use up until the moment the study ends or the interviews are conducted. The vertical axis represents the nature and intensity of the user's experience during this timespan. This vertical axis is divided in the middle by a horizontal line that serves as a line of demarcation between the positive and negative experiences. The task that was given to the participants in the UX Curve study was: 'Try to recall the moment when you started using the product, and draw a curve that reflects how your relationship towards the product has changed from the first time you used it until today'. At the end of this task the expected result is a curve that consists of peaks and troughs, which can be seen as visualisations of the participants' memorable and important changes in experiences. In order to discover the underlying reasoning, participants are also asked to share either verbally or in a written way a brief description of their memorable experiences that are linked to these changes within the curve. In order to complement the general experiences, participants are also asked to consider their experiences from four other points of view (constructs) that are specific to user experience, namely: the perceived attractiveness, ease of use, utility, and degree of usage. A UX Curve template is provided to the participants for each of these viewpoints, meaning that they eventually draw five curves based on their experiences.

[insert figure 1]

Fig. 1. An example of a completed UX Curve (Karapanos et al. 2010)

The use of UX Curve as a research instrument offers insight on two levels (Kujala et al., 2010). Firstly, it enables the detection of what meaningful qualities are perceived as important by the users when using that product on a long-term basis; secondly, this method offers insight into the underlying reasons of 'why' the perception may have changed over time.

3.2 Necessary adaptations towards children

Based on the types of insights that can be obtained using UX Curve, it appeared that this instrument could serve as a starting point for a method that would enable children to recall memories of experiences over a longer period of time. However, in order to increase the instrument's appropriateness for use with children, this adult-oriented instrument underwent some modifications, which resulted in the MemoLine instrument.

The MemoLine instrument was initially developed for use on a project to evaluate educational video games for children with an age between 9 and 11 years old (Vissers, De Bot, & Zaman, 2013). In the latter study the MemoLine was used with a small number of children (n=6) to evaluate a serious game over a 6 month period. The results showed that the children were able to use the instrument and recall past events relating to the game. Given the small sample size and encouraging results, it was felt necessary to conduct further research with the instruments in a variety of contexts. It is conjectured that the MemoLine would be suitable for use with children over the age of 7 because, as highlighted in the literature younger children are unlikely to have developed memory retrieval strategies. In what follows, we will discuss the rationale of the development of the MemoLine instrument in more detail.

3.2.1 Adjustment 1: One dimensional timeline
The first adjustment consisted of replacing the two dimensional curve format by a one dimensional timeline, which visualizes the timespan between using the product for the first time up until the moment of evaluation. However, because of this adjustment, an alternative was needed for the Y-axis that originally represented the type of experience (positive or negative) that occurred over time. Eventually the use of three colours (green, red and grey), by which children were able to mark their experiences over time, served as an alternative for the Y-axis. So, instead of asking children to draw a curve based on changes in their experience over time, they were given three coloured pencils to colour periods on the one dimensional timeline. Periods of positive experiences could be marked with the green colour, periods of less positive or negative experiences with the red colour and periods of non-usage with the grey colour (Vissers, De Bot, & Zaman, 2013).

The underlying reason for the adjustment of dimensions and the accompanying use of colours was based on the specific capabilities of the children for whom the MemoLine was originally designed. Consequently, the researchers relied on their knowledge and skill level as described within the Belgian school curriculum. This curriculum showed that children within this age group of 9 to 11 years old do not yet possess the necessary mathematical knowledge to draw mathematical curves with negative values (Vlaams Ministerie van Onderwijs en Vorming, 2014). Another element that was taken into account for this first adjustment was the possible influence of children's drawing skills when drawing a curve. Birch et al. (1967) have described that when children are asked to draw something, they tend to start at the bottom-left with an upward movement. This tendency would have a direct effect on the results of a curve, as children would be less likely to consider the lower (negative) half of the Y-axis.

3.2.2. Adjustment 2: Temporal recognition cues
Unlike the original UX Curve, which consisted of a blank horizontal axis representing time, MemoLine has been designed in such a way that it consisted of a timeline that was visually divided in weeks or months (according to the elapsed time). In order to facilitate orientation in time even more, visual recognition points based on the child's personal activities that occurred within the evaluation period (like the child's birthday, holidays, or the start of the new school year) were added. These visual recognition points were added to the timeline by the researcher or child prior to and at the start of the evaluation session.

The aim of these adjustments, namely the division of the timeline into weeks or months and the addition of visual recognition points, was to support children in recalling experiences and orientating in time. Similarly, Kujala et al. (2011) have stressed the importance of recognition points in time when

using the UX Curve instrument since they found that difficulties arose even in adults when being asked to describe the exact timeframe of memorable events. These recognition points may help foster Constructive Approaches to recall as the children are chronologically identifying events that occurred in a forward temporal order when adding the recognition points to their MemoLines.

3.2.3. Adjustment 3: Game experience constructs

The original UX Curve instrument allows researchers to examine different user experience constructs, by asking the participants to draw multiple curves, wherein each focuses on a different construct. This principle of division and evaluation of multiple user experience constructs is preserved within the MemoLine instrument for children. However, it was not deemed appropriate to simply apply the same constructs as used in the UX Curve study (i.e., general impression, perceived attractiveness, ease of use, utility, and degree of usage), as the first implementation of the MemoLine focused on the evaluation of games rather than mobile phone usage. Several researchers have attempted to identify the relevant constructs for measuring user experiences with games, but no consensus has been found to date (Desurvire et al., 2004). In a number of studies (Takatalo et al., 2010; Kujala et al., 2011; Desurvire et al. 2004) four user experience constructs were often referred to as contributing to a successful and enjoyable game experience, namely Usability, Challenge, Quantity of Play and General Impression. Therefore, the first MemoLine study concerned with the evaluation of a game, relied on a separate timeline for each of the constructs. These timelines were accompanied by a question explaining the construct, and a legend that explained the meaning of the different colours that could be used to mark periods of time.

What needs emphasizing is that these four constructs are not a fixed value for the MemoLine instrument, as these were specifically chosen to evaluate game experiences. An example of a completed MemoLine, from the school study reported below can be seen in Figure 2 below. In this example 6 of the 9 boxes were coloured in red, with the remaining 3 green. In addition to the 2 cue points provided on the form (Easter and May Bank Holiday) 3 cue points were added, the first two relate to birthdays and the final one the school summer fair.

[Insert figure 2]

Fig. 2. A completed MemoLine for the construct challenge.

3.3 MemoLine as a retrospective tool

The MemoLine is a tool to facilitate retrospective recall through a series of interviews and as such is potentially prone to issues that have been identified in other research areas. Retrospective methods are common within other contexts and research areas outside of CCI, for example, the reliability of children's testimonies within a court of law (Brainerd, 2014). Previous research on children's retrospective thinking has not yield univocal results. Some studies have reported on the problematic nature of children's memory reports (Johnson & Foley, 1984; Howe, 2007; Ceci, Loftus, Leichtman, & Bruck 1994) especially with younger children (Bruck & Ceci, 1999). Whilst other studies have emphasised children's capability to remember personal experiences, answer repeated questions accurately and maintain their account of events despite misleading suggestions (Cassel & Bjorklund, 1995; Rudy & Goodman, 1991).

To understand these contradictory findings and assist children in recalling and verbalizing past experiences, it is important to understand people's memory system. First of all, with each interaction with the world, the initial stimulation received by the senses is hold in a sensory buffer. Then, when people do pay attention to this initial stimulation, the sensory register transfers this information to the short-term memory. At this moment, people can be questioned about that particular interaction, and prompted to verbalize that particular momentary experience. However, as short-term memory has a limited capacity, people will easily forget this momentary experience, unless the information is transferred into long-term memory. This will happen if people engage in remembering strategies, e.g. by recounting the experience several times or attaching meaning to it. It is this long term memory that can be potentially measured in a retrospective study. The long-term memory is structured with a number of separate components and distinction is made between episodic and semantic memory (Schaffer, 2000). Within the context of capturing children's experiences, researchers are particularly interested in retrieving episodic memories as these relate to children's personal experiences, whilst semantic memory stores facts and knowledge of the world. Episodic memory typically relates to the experiences of events located in a particular time and place (Tulving, 1983); it is likely to be autobiographical in nature, and thus personally relevant (Pathman, Samson, Dugas, Cabeza, & Bauer, 2011). In general, from the age of seven, children have developed memory retrieval strategies which they can effectively use to recount long-term memories (Schaffer, 2000; Bjorklund & Harnishfeger 1987).

There are different retrieval strategies for memory recall including *Constructive Approaches* and *Value-Account Approaches*. Constructive Approaches assumes that the reconstruction of past events occurs in a forward temporal order, with the recall of an event triggering the reconstruction of additional events and contextual information surrounding these (Barsalou, 1988). In contrast *Value-Account Approaches* (Betsch et al. 2001) assumes that the reconstruction occurs in a top-down approach, with people not recalling the exact details of the event but the emotional *experience*. It is anticipated that the MemoLine tool would foster recall through Constructive Approaches due to the fact it encourages reflection of past events in a forward temporal order.

4. The Studies

Two studies were used to investigate the appropriateness of the MemoLine tool for capturing children's long-term user experience when playing digital games. The two studies used MemoLine in two different contexts in two different countries: at home within Belgium and at school within the UK. In the home study, the context was replicated but a different type of game was evaluated. The goal was to study whether MemoLine could be used for different applications. This implicates changes regarding content or experiences that could or could not be measured using the same tool. In the school context, the same game was used as in the original study but the context changed. Here, we aimed to examine how MemoLine could be utilised in other environments in which children access technology on a regular basis. There are a large number of studies within CCI that are run within schools (Sim, G., Cassidy, B., & Read, J.C. 2013; Sim & Horton 2012) and it is important to understand whether the tool is appropriate within this environment. Based upon the case studies presented, the MemoLine will be critiqued on its use for different applications and guidelines will be synthesised for the use of MemoLine in a variety of contexts.

4.1.1. Home Study

The case study in which the MemoLine was used in the home setting, within Belgium, was part of a larger project aiming at developing a new educational music app for children (8 to 12 years old) making use of game principles. Early in this project, a proxy technology assessment (PTA) was carried out (Pierson *et al*. 2006). In this PTA, the MemoLine was used to evaluate an existing music game, Piano Dust Buster 2, with children in order to understand their experience of a music game and to identify what children find important when playing educational music games.

4.1.2. Evaluated game: Piano Dust Buster 2
Piano Dust Buster 2 is a music education application for iOS aimed at novice piano players. Players learn to play the piano by helping a grandma character 'dust off' her piano. Following the lead of this character, touching the correct piano keys is turned into a game while players simultaneously play a song on the piano.

4.1.3. Participants
The researchers addressed their personal and professional network to recruit participants. In total 16 children (distributed over 10 families), between 7 and 12 years old, with an interest in learning music participated in this study. The children had a maximum of 2 years experience with formal music education. Family characteristics (such as socio-economic background and family composition) were not considered. All children participated out of personal interest. They did not receive any incentive for their participation.

4.1.4. Materials
The four game experience constructs that were included in the MemoLine for this study were translated into four questions: (1) did you like/dislike playing the game?, (2) did you like/dislike the songs you could play?, (3) how challenging/not challenging was playing the game?, (4) how often did you play the game (often/not often)? The questions presented to the children differed from the original and school study, to address research needs within the project. However, the questions were framed within the same constructs.

4.1.5. Procedure
During a first home visit, the children downloaded the game on an iPad together with the visiting researcher. The children were told they could play the game as often as they liked during the four-month test period. In addition, the children were invited to be 'co researchers', and to critically assess the Piano Dust Buster 2 game, while the researchers said they had too many games to assess and needed the children's help with this.

When the researcher returned for individual MemoLine sessions, about four months later, the children were set at ease by firstly demonstrating to the researcher (who claimed not to know the game) how the game should be played and talked aloud while doing this. The researcher then explained the concept of the MemoLine and helped the children to master colouring timelines by filling in a training timeline about the child's favourite toy. To this end, the researcher coloured in the first two weeks in consultation with the children and asked the children to complete the remainder of the timeline themselves. After this training timeline, the children coloured the four remaining timelines regarding the game. Prior to colouring each timeline, the researcher read the question and answer options to the children and made sure they understood the question. During or after colouring, the researcher inquired about the motivations behind the transitions between the different colours.

After finishing colouring all MemoLine timelines, the children answered some additional questions regarding their performance, their appreciation of specific game features, the game's learning outcomes etc. The questions were presented on a sheet of paper. The researcher used the blank space between the questions to paraphrase the answer of the child. Each session took between 14 and 65 minutes (average: 40 minutes). This includes the introduction, talk aloud exercise, presentation of the method, colouring the timelines, answering additional questions and wrap-up. The entire sessions were audio recorded. The variation in duration depends on the amount of information the child provided and the follow-up questions presented by each individual researcher. In order to minimize distraction, the children and researcher looked for a quiet space in the house (e.g., office, dining area, child's bedroom). This room was picked in consultation with the child and parent. Since the child had met the researcher before, the child felt comfortable to sit alone with the researcher. In a few cases, the living room was picked as a quiet place, which meant parents overheard the conversation and added simple anecdotes to complement the children on rare occasions. The four researchers wrote a methodological report about their experiences with the implementation and feedback from the children after each visit.

4.2.1 School Study

For this study it was decided to replicate the MemoLine using the same constructs and serious game relating to maths that were used within the original study (Vissers et al., 2013) with the major change being a language conversion into English and location of the study. Within the original study, four constructs were measured and these were: often, challenge, understanding and enjoyment. The study was conducted within the school environment whereas the original study was conducted within children's homes. It was important to understand the versatility of the tool and to understand how it can be used within a range of context, so it was felt necessary to use the method within a school. This led to a slight variation in the look of the MemoLine as the Easter Break was highlighted due to the school being closed for this period and therefore the children would not have had access to the game.

4.2.2. Monkey Tale Game

The game 'Monkey Tales: The Castle of Draconian' for the PC was selected, as it had been previously used in the original MemoLine study and was judged to be suitable for use within school as it covered aspects of the Math curriculum. The main objectives of the game are to complete all the levels within the castle by solving puzzles whilst trying not to get killed by monsters. Within the levels are mini math games that can be played by challenging a monkey. These games enable the children to learn math concepts such as fractions, addition and multiplication.

4.2.3. Participants

A total of 32 primary school children from a UK school participated, aged 10 to 11. The school had previously worked with the researchers on other projects prior to this study. Before the study commenced, the teacher examined the suitability of the game and the math content before agreeing to enable the researchers to install the software on the school machines and to run the study. The teacher and two researchers were involved in the study.

4.2.4. Materials

The three game experience constructs that were used in the original MemoLine study (Vissers et al. 2013) were translated from Flemish into English. The three questions were: (1) Did/didn't you enjoy

playing the game?, (2) How challenging/not challenging was it to playing the game?, (3) Did/didn't you find it clear what you had to do and how you needed to do this?

4.2.5. Procedure

Before commencing the study the Monkey Tale game was installed on 15 identical computers within the computer lab at the primary school. The teacher ensured that the children played the Monkey Tales game in the ICT class every Friday for approximately 30 minutes and for a period of three months. The children would play the game individually but occasionally they sought assistance from the teacher or one of their peers. After that period, two researchers who both had considerable experience of carrying out evaluations with children conducted the MemoLine activity and interviewed the children. The decision was made to have all the children complete the MemoLines at the same time within their classroom, with two researchers and the teacher present to offer assistance if required. An initial training activity, lasting for about five minutes, was conducted with the children to familiarise themselves with the MemoLine prior to them completing it individually. One of the authors of the paper displayed a blank MemoLine on the whiteboard and also showed the children a printed version. The 'training' MemoLine was then annotated with important dates and events over this period, including birthdays, exam weeks and holidays. The researcher then stated a computer game that he had been playing during this three-month period. Based upon the construct of enjoyment, the Memoline was coloured in to represent the experiences over this period. After this demonstration, the children were asked to highlight by drawing on their MemoLine any key dates, family or school events and think of a toy or game they have been playing over this period. Once all the children in the class had identified a toy or a game they were asked to complete the training MemoLine. The two researchers and teacher offered assistance to ensure that the children understood the process. It was noted that only two children asked for help at this stage. Once they had all completed the 'training' MemoLine, the children had an opportunity to ask any questions and the process was briefly explained to them again including the four constructs. The children were then given the four MemoLines to complete based upon their experience of playing the Monkey Tales game. Once the children had completed the MemoLines, the teacher and researchers collected them. The whole process lasted approximately 40 minutes.

Following the completion of the MemoLines normal teaching activities had resumed and the children left the classroom in pairs to be interviewed individually by one of the two researchers. The researchers took notes during this interview and the interviews were recorded using a MP3 recorder. To mitigate against any primacy or regency effect based upon their most recent experiences, the order the MemoLines were presented to the children were counterbalanced. For example child 1 would have the order of MemoLine constructs ABCD, whilst the second child would have BCDA. At the start of the interview, the researchers stated the number for example child 4, to ensure that the data could be aggregated to the correct MemoLines prior to analysis. The interviews lasted approximately 5 minutes and all the children were willing to participate. However, the total time each child was engaged with the researchers was similar across the two studies, with the school study taking approximately 45 minutes and the home 40 minutes.

4.3 Analysis

For both the school and home study, the same data analysis methods were applied. The analysis contains three parts. The first part relates to the methodological reports, the second part to the different timelines, and the third part to the interview data.

As the game was played within the school at a set time, there was no variability in children's response to the construct 'often'. The decision was made to exclude this construct from the analysis as within this context it offered no insight into long-term user experience. Whilst all the four constructs (i.e. enjoyment, content, challenge and frequency) were analysed within the home study.

4.3.1. Field notes
The field notes were reviewed to analyse the remarks, from both children and researchers that influenced the implementation process. The experiences of the children that were not included in the field notes but emerged from the interview data (see below) were also included.

4.3.2. Timelines
To determine how the children's experience altered throughout the course of the two studies the positive experiences represented within the MemoLines were quantified. In the school study, the game was available for 11 weeks, however, over this period 2 weeks was the Easter holidays in which the school was closed. Therefore, to establish how the children's experience changed over this period a score for each construct was calculated over a 9 week period. For the home study, the score was calculated for a period of 13 weeks. The calculation of the score was based on counting the green blocks on the timelines. In the school study, each week was coloured in with one colour. In contrast, in the home study, there was more variation in size of coloured blocks or vertical lines. To determine the change in user experience over the four-month period, several rules were applied to reduce the coloured blocks into 13 weeks. Coloured blocks that did not amount to an entire week were still counted. In the few cases where timelines contained too many changes (more than 13), the calculation was made backwards (starting in week 13) or a returning pattern was excluded. When a child had crosshatched red and green, this period was regarded as a red block (since not entirely positive experience). Some children indicated half lines or small coloured blocks within a coloured block; these were not included. In both studies, when children indicated they had not played the game (white or grey blocks), those blocks were not counted. The coloured blocks were quantified using numbers (0 for grey or white, 1 for green, 2 for red) and put into a spreadsheet. A percentage score was then calculated based on the number of children who had coloured a green block and the number of children who played the game that week. This quantification allowed us to visualize the change in user experience over time that could be identified using the MemoLine for both studies.

In addition, to provide additional insights into changes over time, the data was coded using an adaptation of the coding method used by Karaponos et al. (2010) who used a four-point scale. The decision was made to distinguish between a positive and negative linear curvature in order to enable an understanding of the change in experience. The code Discontinuous was removed as this represented a slope that was significantly higher than an average and this was not viable within the MemoLine tool. The MemoLine sections were thus classified under five categories: 1) *Constant (c)* signifying no change in participant's opinion over a certain period of time, 2) *Linear Positive (lp)* signifying a positive change in experience over time, 3) *Linear Negative (ln)* signifying a negative change in experience over time, 4) *Non-Linear (nl)* indicates that there was no grounds for a linear change either positive or negative, 5) *No Play* (np) when the child did not play the game for a particular period of time. Coding was executed for every change reported on the timelines. In the school study, this amounted to 9 changes per timeline. In the home study, the number of changes varied from timeline to timeline since the children did not

colour blocks that represented entire weeks. Based on the example MemoLine shown in Figure 3, starting from a neutral position there are 10 time intervals where an experience can be recorded, an example of the coding the coding therefore would be lp,c,ln,np,lp,c,np,c,ln,np,lp,np.

[Insert Figure 3]

Fig 3. Example of a timeline from the home study. The time intervals (10) are not directly related to the number of weeks (13). The coding took into account the changes independent of the amount of time it indicated.

4.3. Interview data

To analyse the data from the interviews in the school study, content analysis was used for each of the constructs included within the MemoLines. As the focus of the interviews was on the sections of the MemoLine that were not constant, the decision was made to use the same coding scheme used to analyse the timelines. Content analysis would be applied to the interview data for each of the constructs using the following three categories Linear Positive (LP), Linear Negative (LN) and Non-Linear. The data was analysed over two iterations to establish core themes and then establish sub-themes. The comments were initially analysed by reading each comment and interpreting the underlying meaning. This enabled the four core themes to be established that were applicable to the evaluation of the Monkey Tale game and music game. On the next iteration the sub-themes were identified, the frequency of each theme was then counted. Within the school study, the construct enjoyment was analysed first, followed by understanding and then finally challenge. If the data could not be classified to an existing theme then an additional sub-theme was created.

In the home study, the audio recordings were paraphrased for analysis using the same content analysis technique as in the school study. One of the codes referred to methodology in order to incorporate the child's experiences (cfr. procedural results). The codes represent the sub themes related to the sub-themes, as identified in the school study, and were found to correspond with the identified core themes.

5. Results

5.1 Procedural results

The procedural results are presented based on the researchers' experiences of using the method deduced from the field notes and the observed experience of the children using the method. The results are presented for both the home and school study.

5.1.1 Researchers' notes

In both studies the researchers felt that the training served well to address the children's questions and to familiarize the children with the colouring of the timelines. The time that children took to complete each timeline varied for all the children. For example within the home study, the researchers felt waiting for the child to finish colouring was quite boring when children coloured very precisely. The school study differed in the training approach compared to the home study as it was set in a classroom context with all the children colouring in at the same time. In this context, some of the children had to wait for their peers to finish completing their timelines before the group could move on to discussing the next construct. With a large group of children, it was important to have a number of researchers available to

assist any children who were having difficulty to ensure the children were not waiting too long before moving to the next construct.

In the home study, for the construct challenge, researchers noted that not all children understood what it meant. After explanation (i.e., something is hard but fun) children could proceed without any problem. However, two boys changed the colour legend to their understanding (i.e., red for hard (challenging) and green for easy (not challenging). This issue was not raised in the school study but there was evidence that this confusion may have occurred when the researchers analysed the interview data. At some point, one researcher felt that four colour options would be necessary to distinguish between not challenging enough, not challenging, challenging, too challenging.

The researchers involved in the home study reported that the colouring procedure was clear for the majority of the children, however, the children sometimes needed encouragement or the question rephrasing to enable them to continue the process. Only a few children required the researcher to walk the child through every timeline, this was facilitated through the use of prompts such as *'And then, did you play? Did you like playing?'.*

One of the adaptions from the UX Curve was the inclusion of temporal recognition cues in which children had to record personal events pertinent to them along the x-axis of the MemoLine. The children appeared to understand this process and were able to add these temporal recognition cue points to their timelines. The number of cue points recorded ranged from 1 to 4 and overall the children recorded similar events, notably birthdays of friends and family members, which is likely due to the information provided in the training session. There were a few other types of events that were recorded such as summer fairs and sporting events.

5.1.2. Feedback from the children

Some of the children indicated in the home study that colouring the same scheme multiple times was boring, or that colouring took too long. This was not picked up in the school study. Although some of these children were waiting for their peers, they did not experience long time delays or show signs of boredom or frustration at having to complete another timeline.

In the home study, due to the timeline's layout (i.e., divided in monthly intervals), MemoLine did not serve well to indicate occasional occurrences but this was not an issue in the school study. Children would visualize half-day or one-day events by arbitrarily marking colour cubes, or drawing vertical pencil lines in a contrasting colour. In both studies, the colour options (i.e., two colours to express experience) limited children in making nuances in their experience (e.g., a little bit challenging). Sometimes, children crosshatched red and green in order to express more nuance. Also, both the colouring and the explanations for the periods of non-play usually did not vary between the different timelines, which demonstrates the reliability of their answers. The negative effect of this is that some children found they were repeating what they were saying, which may be perceived to be rather unnecessary.

Some children expressed finding the process difficult in the home study. They found it especially difficult to remember all the events or to verbalize their experiences. However, in the school study, only one child out of the 31 children who participated struggled to recall any important past events. This may have been attributed to the additional delay between the completion of MemoLine and the interview.

Within the school the children were interviewed by one of two researchers up to two hours after completing the original MemoLine. This meant that the children towards the end had to recall their thought process from when they completed the MemoLine, this was not the case in the home study who were interviewed immediately after or during completion of the timelines. In the home study, a couple of children needed some encouragement from the researcher in order to formulate a response. For example, they would first say "I don't know". After friendly encouragement of the researcher all children eventually formulated a response.

5.2 Results of the two studies
The results from each of the two studies are reported below.

5.2.1. Home study
The children were able to complete all timelines. They added few new recognition points, since the most important ones were marked on the timeline already. However, the recognition points related to Christmas and New Year were mostly talked about as a period (i.e., Holidays or pre-vacation tests) instead of one-day events.

Figure 3 depicts the changes in experience with regards to the different constructs. With regards to play frequency, week 7 marks the first week of the Christmas Holidays, which might explain the sudden drop in play frequency. Also, it was noticed that of all constructs, play frequency decreases most (from 93% to 54%). Figure 4 shows that challenge is greater towards the end of the test period and highlights the continued fluctuation in the constructs over the period of time.

[Insert Figure 4]

Fig 3. Change in the constructs over time for the home study.

Table 1 shows the classification type for the changes in experience over time. Most codes relate to *No Play* or *Constant*. On several occasions, the children's experience (positive or negative) would remain constant after a period of no play. The timelines illustrate a varied play experience, with an average of 9 changes per timeline. Non-linear experiences are mostly used to express enjoyment during game play.

	Classification Type				
Construct	**C**	**LP**	**LN**	**NL**	**NP**
Content	42 (31%)	28 (20%)	22 (16%)	1 (1%)	44 (32%)
Challenge	35 (28%)	26 (21%)	22 (17%)	3 (2%)	40 (32%)
Enjoy	41 (29%)	28 (20%)	18 (13%)	7 (5%)	48 (34%)
Frequency	34 (24%)	32 (23%)	27 (19%)	3 (2%)	43 (31%)

Table 1. Classification of MemoLine changes over time C = Constant, LP = Linear Positive, LN= Linear Negative, NL=Non-Linear and NP = No Play.

Based on the interview data of the children's responses to the changes within the MemoLines, the number of changes in positive and negative experiences reported by the children are shown in table 2

below. The number of experiences does not directly relate to the amounts of comments made by the children, as one child might provide several reasons for the change in experience.

	Learning			Play Experience			Game Play			Social		
	LP	NL	LN	LP	NL	LN	LP	NL	LN	LP	NL	LN
Content	0	(1)	0	4	0	6	21	0	18	0	0	0
Challenge	6	1	10	16	1	0	22	1	10	0	0	0
Enjoy	9	0	0	15	6	14	20	1	13	0	0	9
Frequency	0	(3)	0	20	0	24	0	0	2	0	0	0

Table 2. Linear Positive, Linear Negative and Non-Linear responses to the themes for the Piano Dust Buster 2 game.

For the construct challenge, the main sub themes focused on *first time* or *new songs*, introducing the importance of a novelty factor throughout game play. First time experiences are challenging, e.g. *"because I didn't know very well how to do it then"* or *"because I had just got it and I had never tried it before"*. Also, children reported on a positive sense of challenge when trying out the advanced mode offered in the game: *"I already knew all the songs (…) and put it in the hardest version"*. Additionally, children would come up with additional exercises to increase challenge. For instance, one girl played together with her brothers: *"we selected songs for each other and then you didn't know which song it was going to be. That was fun, because you didn't know if the song was going to be long or short, hard or difficult."* Another boy said *"I played a large part of a hard but fun song on the keyboard"*. However, some children also said that *"when we got used to it, it wasn't challenging anymore"*. When this happened, the children also felt that they would stop learning new things. Related to *new songs*, we learned that challenge increased as new songs were introduced in higher levels *"and then it was a higher level again, and then it became more difficult again"* and *"those were songs I did not know"*. Children reported a low sense of challenge when they had to *repeat songs*, because *"I knew all the songs by heart"* or *"it was a little too easy. [even if] I wanted to play easy songs, I didn't find it very nice"*. With regards to the non-linear experiences, this usually relates to the fact that within one song or level there were things that were more easy or difficult to master: *"in one song there are easy and hard parts"*. Exceptions are the children that expressed they wanted to do everything right or excel: "*there is one part in the song 'La Bamba' that I always get wrong, so I'm not happy with my performance*."

The construct enjoyment was mostly related to the fact that children felt they had mastered the songs and exercises offered in the game. They then used these skills to progress: *"we alternated turns and tried to get new songs as quick as possible"* or *"sometimes I play really easy songs, songs that are like nothing, as if I have been playing them for years. We get points without it being difficult."* When children felt they were good at it, this still instilled the need to learn more and make progress in the game: *"I wanted to get higher and higher"*. In contrast, children reported a low level of enjoyment when the exercises were too *hard*. For instance, when children are confronted with their limits and could not proceed in the game. *"We have been practicing 'Spring' by Vivaldi for weeks now and we can't finish it. We don't get enough points to proceed to a new song."* The non-linear experiences are quite diverse. One girl felt enjoyment while playing but since she couldn't play as often as she wanted, this influenced

her experience negatively. Similarly, one girl felt she could not enjoy playing to the fullest because her father told her she had to play. Two children did not like the game that much, but played it anyway.

The results for the construct content are related to the type of songs offered in the game. The sense of enjoyment and challenge was related in a positive and negative way to novelty (new songs to practice) or familiarity (after practicing) of the songs. However, some comments also related to the type of song *"some songs were too childish"* or *"I don't like the Mozart songs"*. Rarely, a known song children knew from the radio motivated the children to progress more rapidly *"I didn't like the songs I was playing, but I needed to play them in order to finally play 'Waka Waka'."*

Comments related to the construct frequency were related to other (play) activities that children engage in *"after the test period, we had to play again to catch up on the time we lost"* or *"at the time, it was the only nice game we could play with"*. The non-linear comments related to frequency of play were coded under no-play. The reasons were also attributed to external factors; such as low battery of the iPad or the withdrawal of certain privileges like playing on the iPad by the parent or guardian as a punishment.

5.2.2 School Study

All the children managed to complete the MemoLines for each of the constructs. There was very little variation in the temporal recognition cues that were added with 6 categories being identified, these being: computer games (4), summer fair (28), birthdays (30), exams (4), sporting event (2) and holidays (6). In total 74 cues were added to the MemoLine with birthdays and the schools summer fair accounting for 81% of all recognition cues.

Figure 4 shows the percentage of children who coloured the MemoLine in green each week and how their responses changed over time.

[Insert Figure 5]

Fig. 5. Change in the constructs over time for the home study.

Figure 5 highlights the fact that the responses to the constructs change over time. For the construct enjoyment this peaked at 84% for the first two weeks and dropped as low as 39% in week 5. In contrast, the challenge construct appeared to peak at the end with the highest percentage being reported in week 9 at 76%.

Table 3 shows the changes in experience over time for the 3 constructs under investigation by the children in school playing the Monkey Tales game.

	Classification Type				
Construct	**C**	**LP**	**LN**	**NL**	**NP**
Clear	164 (57%)	52 (18%)	44 (15%)	10 (3%)	18 (6%)
Challenge	157 (55%)	53 (18%)	49 (17%)	10 (3%)	18 (6%)
Enjoy	123 (43%)	69 (24%)	57 (20%)	12 (4%)	27 (9%)

Table 3. Classification of MemoLine changes over time C = Constant, LP = Linear Positive, LN= Linear Negative, NL=Non-Linear and NP = No Play.

For a large proportion of time the children's experience stays constant. For these constructs it would appear that *Enjoyment* had a greater level of variance in the children's experiences. Their initial experiences were high with 25 of the 32 children reporting a positive experience in the first week and this remained constant in the 2nd week, however 11 children then reported a negative experience in week 3 and there was considerable variance after this point.

Based upon the analysis of the interviews, following the completion of the MemoLines, table 4 below shows the results from the content analysis for each of the three constructs. The 3 rows in column 1 represent the 3 constructs investigated using the MemoLines and the underlying themes relate to this construct.

	Learning			Play Experience			Game Play			Social		
	LP	NL	LN	LP	NL	LN	LP	NL	LN	LP	NL	LN
Clear	5	0	1	3	0	5	18	8	20	1	0	0
Challenge	1	0	5	10	0	7	11	8	18	1	0	1
Enjoy	0	0	2	16	4	22	10	1	17	3	0	0

Table 4. Linear Positive, Linear Negative and Non-Linear responses to the themes for the Monkey Tales game.

For Clear, the majority of the comments related to the theme Game Play and there were several linear positive comments relating to the fact the game was initially hard but they could quickly make progress and thus the *Easy to learn* sub-theme emerged accounting for 14 of the 18 comments. Children's comments included "*after 1st time it got a bit easier*'" and "*first two weeks I did not know what to do".* In contrast to the positive aspects there were linear negative comments within Game Play relating to the sub-themes *Lack of Clear Instruction* and *Game Complexity*. For the sub-theme Lack of Clear Instructions children's comments included "*didn't know what to do I could not get across to other side in game"* and '*trap monsters confusing what opens what'*. Whilst for Game Complexity comments included "*last week got hard again*" and "*week 4 couldn't understand how to do the level I did know how to shut ghosts in gates.*" The Non-Linear comments all related to Game Play and the *Lack of Clear Instructions* in certain sections for example "*week 2 overall clear but some in game instructions unclear*" and "*week 2 tutorial I did not understand how to push the boxes*".

For challenge the main linear positive sub-themes centred upon *Increase Game Complexity* within Game Play and *Learning to Play* within Play Experience, with statements such as "*levels simple in the first few weeks, in week 8 really hard missing timings to get to special place*." Whilst the majority of the linear negative fell within Game Play and the sub-theme *Hard*. The main reasons for these changes related to the complexity of specific levels, for example children stated "*block level really hard*" and "*week 11 on the bridge level really hard you sink if you fall off*". All the Non-Linear elements also fell within Game Play and the complexity of the levels with statements such as "*last level really hard, kept doing it until I finished*" and "*some of it was easy and some of it was hard*".

Finally for enjoyment, the majority of children's comments were linear negative, primarily focusing on the fact the children perceived the game play as *boring* within Play Experience. This notion of being bored fluctuated over the time period, with children reporting comments such as on week 3 *"- stuck on one level - got boring";* and on week 4 *"new level that was more exciting".* This highlights how the children's experience altered over time. Other negative comments focused on the Math elements with *"bored doing maths"* and *"got harder could not pass the maths parts"*. For the comments that were categorised as linear positive these were mainly within the Play Experience theme, under the sub-theme of *Positive Experience.* These include comments from the children relating to a variety of different aspects including *"better mini games", "levels got more interesting"* and *"week 10 got excited getting onto the last level"*. Unlike the constructs clear and challenge, for enjoyment the majority of non-linear comments related to Play Experience and the *boring* sub-theme, and there was 1 instance of the game being *Hard.* The boring comments focused on the repetitive nature of the game with comments such as *"bored of challenges same as previous ones and not difficult"* and *"started enjoying but levels too easy so didn't enjoy"*. For *Hard* the comment was *"After Easter, enjoyed but kept getting killed which made less enjoyable".*

6. Discussion

The aim of this study was to investigate whether the MemoLine is an appropriate tool for capturing changes in user experience for games over time with children. The tool was critiqued based on the ease of use within a research context and the data derived from the method. Many of the evaluation methods within CCI have been designed to work with children. Therefore ease of use was felt important to ascertain. If the children or researchers struggled to explain or use the method, then the appropriateness may be questioned. If children could use the method but the data derived from the study offered very little insight into their experience and could not be used to inform future designs, then the method may be deemed inappropriate as well.

Overall, within the two study settings, the participating children were able to complete their individual MemoLines after the initial training. The training was carried out differently between the two studies. In the home study, the training was conducted on a one-to-one basis with the researcher assisting the child to colour in the first two weeks. Once the child understood the process, (s)he was asked to complete the remainder of the timelines. In contrast, all the children received the training at the same time within the school, and were assisted by the teacher and researchers if they encountered any difficulties. Even with a difference in individual guidance during training the children were able to follow instructions and complete the activities, as has been evidenced before from participatory design sessions involving groups of children (Van Mechelen et al., 2014; Sim, Read, Gregory & Xu, 2015). Therefore the MemoLine could be completed either individually on a one-to-one basis or as part of a larger group.

The children were able to complete the MemoLines for the various constructs investigated in the two studies. In line with other studies (Karapanos et al., 2010), the MemoLines produced showed that children's experiences changed over time for all constructs within both studies. These changes are essential to facilitate the interviews, acting as a visual prompt to invoke a discussion. If children were not able to identify changes in experience, it would not be possible to capture the reasons *why* user experience changes over time with the technology in question.

The procedure of the MemoLine encourages a Constructive Approach to recall (Barsalou, 1988) by having children reflect on their experiences in a forward temporal order. Recognition cues were added before the children completed the MemoLines in order to provide triggers for the reconstruction of events and experiences. During implementation, children were also given the opportunity to add their own recognition cues prior to or while colouring in the timelines. In the school study all the children added at least one recognition cue, with birthdays being the prominent type. This was not the case for the home study, with the majority of children not adding any additional cues on their timelines. Within the analysis of the interview data there is very little evidence of additional recognition cues influencing the discussion. However, children referred to the cues within an holiday period (i.e., Christmas and New Year) suggested by the researcher. In the school study, only a few children referred to the recognition cues they added. For example one child recollected their experiences around his or her birthday and stated *"had birthday wasn't in the mood",* whilst a few children referred to the Easter Holidays, one stating *"week 5 - got stuck and it got hard. Still hard after break.".* In both studies there is evidence of the children reflecting on their experience around a holiday period. Holidays seem to be an important point of reference that differs to a big extent from regular school days.

When analysing the data, it became clear that the differences in lay-out and presentation of the timelines greatly influence the way children colour the timeline and, hence, the results these produce. First, the division in smaller time blocks influences the width of the coloured blocks. Within the school study, the MemoLines were separated out into weeks and children were more likely to fill in the entire week with a solid colour. In contrast, within the home study, children tended to indicate more change within a period by having multiple colours represented within a block. This was most evident for periods of no play. This may have been due to the fact that the majority of children in the school study seemed to only play the game during the allocated time and it may have been expected that they would just have one block of colour for that session. In contrast, at home, they could play the game as often as possible which resulted in very small colour blocks, like pencil lines. In both studies the coloured blocks varied in size of colour blocks stretching over several weeks. Second, some children applied nuances by hatching in both green and red, or applying contrasting colour cubes within a colour block. It may be questionable whether having only two colours is sufficient to represent their experiences, in a small number of cases children appeared to try and compensate for this by combining the two colours. It is recommended that the MemoLine maybe adapted based upon the anticipated frequency in which the product will be used by the children. For example, if the children are expected to interact with the software or product on a daily basis then more blocks within the MemoLines may be required to enable the children to express their changes in between short periods of time. This may not be required or practical in all instances, for example they may interact with a new tablet or interactive TV on a daily basis and it may not be practical to try and have children recall daily events. With regards to colour it may be necessary to ensure the appropriateness of the colours used due to cultural differences, for example red is perceived as a positive colour in countries such as China. In addition more colours may be required to overcome the limitations of a forced response, thus enabling children to represent a greater range of experiences beyond simple positive or negative experience.

Another point concerns recall of events and the possibility for children to provide reasoning behind the change in experience. In both studies, some children struggled to recall experiences and provide the explanation behind their choice of colour. It is hard to evaluate why these children struggled. For example, it might be down to personality traits such as the child being shy. However, we did not observe

difficulties that eventually impeded finishing the exercise; encouragement and guidance from the researcher throughout the colouring and recall process aided the children to think again. In order to ensure this type of guidance, it may be sensible where feasible to use the MemoLine on an individual basis - as in the home study. This has the added advantage of easing the cognitive demands on the children in having to recall events twice: once when completing the MemoLines and again in the interviews. This occurred in the school study where, in some instances, there was a gap of several hours between the completion of the MemoLines and the actual interviews. To help minimise this delay, it may be necessary to involve more researchers, enabling multiple interviews with the children to take place at once. This suggestion is only feasible when sufficient quiet locations are available in a school to accommodate all the separate interviews.

The post-hoc interviews are meant to understand the reasons for changes in user experience over time. In order to probe into the underlying reasons and values, we recommend relying on child-friendly interview techniques such as User eXperience Laddering (Zaman & Abeele, 2010). In the school study the researchers did not probe this far. For example, in the statement "*levels became difficult*" it would have been useful to establish the actual level within the game that this occurred and why it became difficult. Without this information, it would be difficult for developers to understand the challenge faced by the child and redesign the game. This level of detail did appear in another child's statement in the MemoLine, as made with regards to week 11 "*on the bridge level it got harder sink if fall off*". For the home study, the underlying reasons were a little more detailed than the school study, with comments like "*We have been practicing 'Spring' by Vivaldi for weeks now and we can't finish it. We don't get enough points to proceed to a new song.*" The increased level of depth might be explained by the aim of the home study, namely to inspire further design decisions as opposed to the school study which did not was to understand what they liked or disliked about the game.

In both studies, children provided both positive and negative comments about their game experiences. The reliability of children's verbalization of memories has been questioned in literature (Howe, 2007), pointing to the fact that children are susceptible to making up false events (Bruck & Ceci, 1999). To account for this, we triangulated the results from MemoLine with those of the interviews, and where possible also verified whether children's statements about game elements also occurred within the actual game. In the home study, we aimed to facilitate recall by encouraging the children to 'show and tell', using the game to support their rationale. This instruction facilitated children in reporting their experiences in relation to concrete design properties. In this respect, our findings are in line with previous research in which it has also been shown that it is useful to show screen shots of specific design elements to aid children's memory in retrospective think aloud (Van Den Haak, De Jong, & Jan Schellens, 2003). In the school study, on many occasions, children reported on their interaction with game attributes to explain changes in experiences, even without being explicitly instructed to point to design elements through the use of visual cues. For instance, one child pointed to the difficulty of the level with the cannon, and another described the difficulty on a level in which players had to cross a bridge, both of which appeared in the game.

Our results further showed that MemoLine yields information that other long-term user experience instruments do not capture. First, MemoLine can be used to determine when children stop playing a game, and why. For example, within the home study, children became bored of listening to the same songs. By using MemoLine, the approximate time period could be identified, enabling developers to

make informed decisions about when to release content updates. Determining when gamers stop playing can also be captured through analytics and data mining techniques (Guardini & Maninetti 2013), but it would be difficult to infer causality of why the child stopped playing. The latter can be achieved through complementary verbalization methods such as the MemoLine. Second, reasons for lack of engagement with the game were also identified in the school study with statements such as *week 3 "stuck on one level - got boring".* It may be more difficult to identify some of these issues with other approaches used to capture changes in user experience over time; such as cross-sectional or longitudinal research design. For example, using a pre and post-test with survey tools, as in the study by Barendregt, Bekker, Bouwhuis, & Baauw (2006), it would be difficult to capture specific events that might trigger boredom or identify the challenges faced by the children in higher levels. Barendregt et al. used the Smileyometer to capture children's experiences and this limits the type of data that can be captured. The use of diaries may yield more insights but because of the repetitive use, they are susceptible to drop outs.

In relation to the use of constructs, we found that each construct elicited different insights from the children. The construct 'frequency' was only examined in the home study in order to study if children would continue to play, and, hence, continue learning. The results from the home study indicate that children stop playing and then re-engage with the game after a while. This data allowed for exploration of what makes the child go back to the game after a period of no play in the home context. We see that there are clearly external influences such as other games and activities competing for their time. The reasons might differ in other contexts, such as schools. However, the insights related to play frequency might not be desirable in every research or design project. Furthermore, our data suggest that explanatory insights could be elicited from the interviews, and that the constructs are not independent of each other. For example, in the home study under the construct challenge a child stated *"when we got used to it, it wasn't challenging anymore"* and this would affect enjoyment and frequency of play.

7. Conclusion

The aim of this study was to investigate the appropriateness of the MemoLine as a retrospective tool for capturing changes in long-term user experience of games with children. The appropriateness was analysed based on the ease of use established through analysis of the procedure and the data derived from the interviews.

All the children could complete the MemoLines for the constructs investigated. The coloured timelines could be used by the researchers to carry out the interviews and explore the reasons why children's experiences change over time. From the studies presented above, that highlights the appropriateness of the MemoLine for use in either a home or school context, we can make recommendations for other researchers to use this instrument. We suggest providing training for the colouring exercise, since it would probably not be possible for the children to complete the MemoLine without guidance. After training, the children can proceed with the exercise independently, but researchers should be nearby to aid in case of problems. Also, it might be important to limit the time between the colouring exercise and the interview in order to limit cognitive efforts. Incorporating child-friendly validated interview techniques such as the User eXperience Laddering method could enhance the interviews. However, it may be that the use of User eXperience Laddering as a closing part of the MemoLine procedure increases the cognitive demands on children and this requires further research. The use of screen shots of the game may be used to try and assist the children recall past events. Furthermore, the layout of the

MemoLine may need to be adapted depending on expected frequency of use of the product being evaluated, depending on the context and type of product. Also, researchers should select the constructs they study carefully based on their design needs. It is probably best to limit the amount of timelines since a small number of children did find the colouring in process a little repetitive. Future work is required to determine the suitable number of constructs and additional constructs that might be added both inside and outside a gaming context, needs to be addressed. The inter-relationship between constructs was not fully explored within the studies reported in this paper and this should be examined in future studies, as well as, how the interviews may facilitate the examination of these relationships. In addition future work will look at the possibility of including more colours to enable children to represent a greater range of experiences beyond simple positive or negative experiences (by using green or red colours).

In conclusion, in the two different contexts the majority of children were able to recall experiences relating to the game and provide data that would be useful for developers to understand how their experiences changed over time. This data can then be used to inform future designs or provide new material or updates to existing products. It is important to understand what motivates and demotivates children when using technology over time to help facilitate the improvement of existing or future products. Without this understanding it would be difficult to ascertain the reasons why they stop playing or interacting with products.

8. Acknowledgments

References

Barsalou, L. (1988) The content and organization of autobiographical memories. In Remembering reconsidered: Ecological and traditional approaches to the study of memory. U. Neisser and E. Winograd, Eds. Cambridge University Press, Chapter 8, 193-243.

Barendregt, W., Bekker, M., & Baauw, E. (2008). Development and evaluation of the problem identification picture cards method. *Cognition, Technology and Work, 10*(2), 95-105.

Barendregt, W., Bekker, M., Bouwhuis, D. G., & Baauw, E. (2006). Identifying usability and fun problems in a computer game during first use and some practice:. *International Journal Human Computer Interaction, 64*, 830-846.

Betsch, T., Plessner, H., Schwieren, C. & Gutig, R. (2001). I like it but I don't know why: A value-account approach to implicit attitude formation. *Personality and Social Psychology Bulletin* 27(2), 242.

Birch, H.G. and Lefford, A. 1967. Visual Differentiation, Intersensory Integration, and Voluntary Motor Control. Monographs of the Society for Research in Child Development. 32, 2, 1–87.

Bjorklund, D., & Harnishfeger, k. K. (1987). Developmental Differences in the mental effort requirements for the use of organisational strategy in free recall. *Journal of Experimental Child Psychology, 44*(109-125).

Brainerd, C. J. (2014). Developmental Reversals in False Memory. A new look at the Reliability of Children's Evidence. *Current Directions in Psychological Science, 22*(5), 335-341.

Bruck, M., & Ceci, S. J. (1999). The suggestibility of children's memory. *Annual Review of psychology, 50*, 419-439.

Cassel, W. S., & Bjorklund, D. F. (1995). Developmental patterns of eyewitness memory and suggestibility. An ecologically based short-term longitudinal study. *Law and Human Behaviour, 19*(507-532).

Ceci, S. J., Loftus, E. F., Leichtman, M. D., & Bruck, M. (1994). The possible role of source misattribution in the creation of false beliefs among preschoolers. *The international Journal of Clinical and Experimental Hypnosis, XLII*, 304-320.

Colombo, L., & Landoni, M. (2014). *A diary study of children's user experience with EBooks using flow theory as framework*. Paper presented at the Proceedings of the 2014 conference on Interaction design and children, Aarhus, Denmark.

Couse, L. J., & Chen, D. W. (2010). A tablet computer for young children? Exploring its viability for early childhood education. *Journal of Research on Technology in Education*, *43*(1), 75-96.

Desurvire, H., Caplan, M., & Toth, J. A. (2004). Using heuristics to evaluate the playability of games. *In Extended Abstracts on Human Factors in Computing Systems* (pp. 1509–1512). Presented at the CHI, New York.

Guardini, P., & Maninetti, P. (2013). Better Game Experience Through Game Metrics: A Rally Video Game Case Study. In M. Seif El-Nasr, S, A. Drachen & A. Canossa (Eds.), Game Analytics Maximizing the Value of Player Data (pp. 325-361): Springer.

Horton, M., Read, J. C., & Sim, G. (2011). *Making your mind up?: the reliability of children's survey responses*. Paper presented at the 25th British HCI Conference, Newcastle.

Howe, M. L. (2007). Children's Emotional False Memories. *Psychological Science, 18*(10), 856-860.

ISO. (2010). Ergonomics of human system interaction - Prt 210: Human-centred design for interactive system. (Vol. 9241-210). Switzerland: International Standards Organisation.

Johnson, M. C., & Foley, M. A. (1984). Differentiating Fact from Fantasy: The Reliability of Children's Memory. *Journal of Social Issues, 40*(2), 33-50.

Karapanos, E., Hassenzahl, M. and Martens, J.-B. (2008). *Users' experiences over time.* Proceedings of the 26th of the international conference extended abstracts on Human factors in computing systems (New York, NY, USA, 2010), 3561-3566.

Karapanos, E., Martens, J.-B. and Hassenzahl, M. (2010). *On the retrospective assessment of users' experiences over time: memory or actuality?* Proceedings of the 28th of the international conference extended abstracts on Human factors in computing systems (New York, NY, USA, 2010), 4075–4080.

Karapanos, E., Martens, J.-B., & Hassenzahl, M. (2012). Reconstructing experiences with iScale. *International Journal of Human-Computer Studies*, *70*(11), 849-865.

Karapanos, E., Zimmerman, J., Forlizzi, J., & Martens, J.-B. (2009). *User experience over time: an initial framework*. Paper presented at the Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, Boston, MA, USA.

Karapanos, E., Zimmerman, J., Forlizzi, J. and Martens, J.-B. (2010). Measuring the dynamics of remembered experience over time. *Interacting with Computers*, 22, 5 (2010), 328-335.

Kujala, S., Roto, V., Väänänen-Vainio-Mattila, K. and Sinnelä, A.(2011). *Identifying hedonic factors in long-term user experience*. Paper presented at the Proceedings of the 2011 Conference on Designing Pleasurable Products and Interfaces, Milano, Italy.

Kujala, S., Roto, V., Väänänen-Vainio-Mattila, K., Karapanos, E., & Sinnelä, A. (2011b). UX Curve: A method for evaluating long-term user experience. *Interacting with Computers, 23*(5), 473-483.

Vlaams Ministerie van Onderwijs en Vorming. (2014). Primary Education – Mathematics – Attainment. Retrieved on 13 February 2015 from:http://www.ond.vlaanderen.be/curriculum/basisonderwijs/lager-onderwijs/leergebieden/wiskunde/eindtermen.htm.

Markopoulos, P., Read, J. C., MacFarlane, S., & Hoysniemi, J. (2008). *Evaluating Children's Interactive Products: Principles and Practices for Interaction Designers.* San Francisco: Morgan Kaufmann.

Moschou, E., & Zaharias, P. (2013). *UX-Curve revisited: Assessing long-term user experiences of MMOGs*. Paper presented at the CHI, Paris, France.

Norman, D. A. (2009). The way I see it: Memory is more important than actuality. *Interactions, 16*(2), 24-26.

Pathman, T., Samson, Z., Dugas, K., Cabeza, R., & Bauer, P. J. (2011). A "snapshot" of declarative memory: Differing developmental trajectories in episodic and autobiographical memory. *Memory, 19*(8), 825-835.

Pierson, J., Jacobs, A., Dreessen, K., Lievens, B., Van den Broeck, I., & Van den Broeck, W. (2006). Walking the interface: uncovering practices through proxy technology assessment. Paper presented at *EPIC*, 24-26.

Read, J. C. (2008). Validating the Fun Toolkit: an instrument for measuring children's opinion of technology. *Cognition, Technology and Work, 10*(2), 119-128.

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

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

Read, J.C., Sim, G., Gregory, P., Xu, D., & Ode, J.B. (2013). Children Designing Serious Games. EAI Endorsed Transaction on Serious Games, 1(e5).

Reis, H. T. (1994). Domains of experiences: investigating relationships processes from three perspectives. In R. Erber & R. Gilmore (Eds.), *Theoretical Frameworks in Personal Relationships* (pp. 87-110). Mahwah, NJ.: Erlbaum.

Rudy, L., & Goodman, G. S. (1991). Effects of participation on children's reports: Implications for children's testimony. *Developmental Psychology, 27*(527-538).

Schaffer, H. R. (2000). *Introducing child psychology*. Oxford: Blackwell Publishing Ltd.

Sim, G., Cassidy, B., & Read, J.C. (2013). Understanding the fidelity effect when evaluating games with children. Paper presented at 2013 conference on Interaction design and children, New York.

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

Sim, G., MacFarlane, S., & Horton, M. (2005). *Evaluating usability, fun and learning in educational software for children*. Paper presented at the World Conference on Educational Multimedia, Hypermedia and Telecommunications, Montreal.

Sim, G., Read, J.C. Gregory, P. & Xu, D. (2015). From England to Uganda: Children Designing and Evaluating Serious Games. *Human-Computer Interaction*. *30*(3-4), 263-293.

Takatalo, J., Häkkinen, J., Lehtonen, M., Kaistinen, J., & Göte, N. (2010). Presence, Involvement and Flow in Digital Games. In *Evaluating User Experience in Games* (p. 277).

Tulving, E. (1983). *Elements of episodic memory.* Oxford, UK.: Oxford University Press.

Van Den Haak, M., De Jong, M., & Jan Schellens, P. (2003). Retrospective vs. concurrent think-aloud protocols: testing the usability of an online library catalogue. *Behaviour & Information Technology*, *22*(5), 339-351.

Van Mechelen, M., Gielen, M., Abeele,V.V., Laenen, A., & Zaman, B. (2014). Exploring challenging group dynamics in participatory design with children. Paper presented at 2014 conference on Interaction design and children, Aarhus.

Varsaluoma, J., & Kentta, V. (2012). *DrawUX: web-based research tool for long-term user experience evaluation*. Paper presented at the Proceedings of the 7th Nordic Conference on Human-Computer Interaction: Making Sense Through Design, Copenhagen, Denmark.

Vermeeren, A. P. O. S., Law, E. L.-C., Roto, V., Obrist, M., Hoonhout, J., & Vaananen-Vainio-Mattila, K. (2010). *User Experience Evaluation Methods: Current State and Development Needs*. Paper presented at the NordiChi 2010, Reykjavik.

Vissers, J., Bot, L. D., & Zaman, B. (2013). *MemoLine: evaluating long-term UX with children*. Paper presented at the Proceedings of the 12th International Conference on Interaction Design and Children, New York, New York.

Von Wilamowitz-Moellendorf, M., Hassenzahl, M., & Platz, A. (2006). *Dynamics of the user experience: How the perceived quality of mobile phones changes over time.* Paper presented at the User Experience - Towards a unified view, Workshop at the 4th Nordic Conference on Human Computer Interaction., Oslo, Norway.

Xu, Y., Read, J. C., Sim, G., McManus, B., & Qualter, P. (2009). *Children and Smart Technologies: Can Chldren's Experiences be Interpreted and Coded*. Paper presented at the HCI 2009 People and Computers, Cambridge.

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

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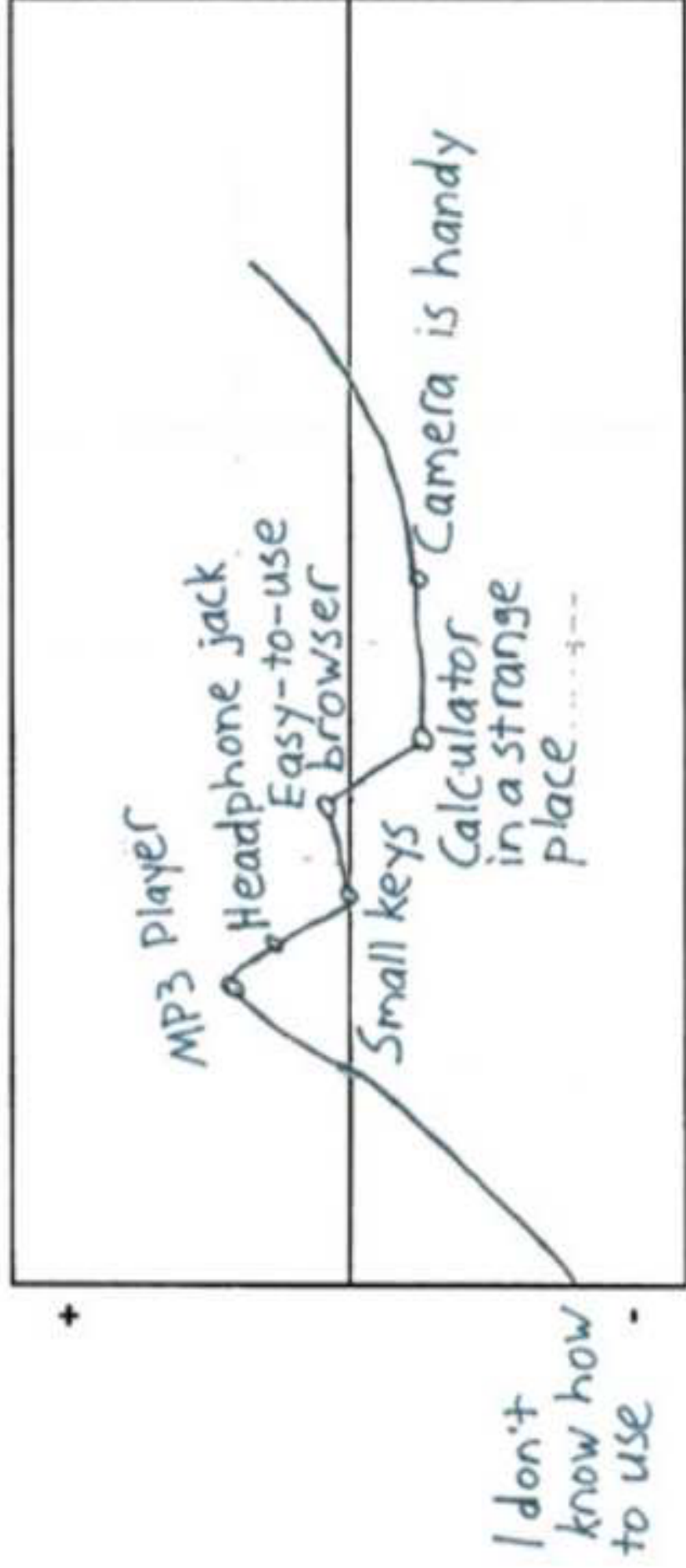
Short description of the changes:

Learning in the beginning

Actual Game- How challenging/not-challenging was it to play the game?

March April May

Easter

May Bank Holiday
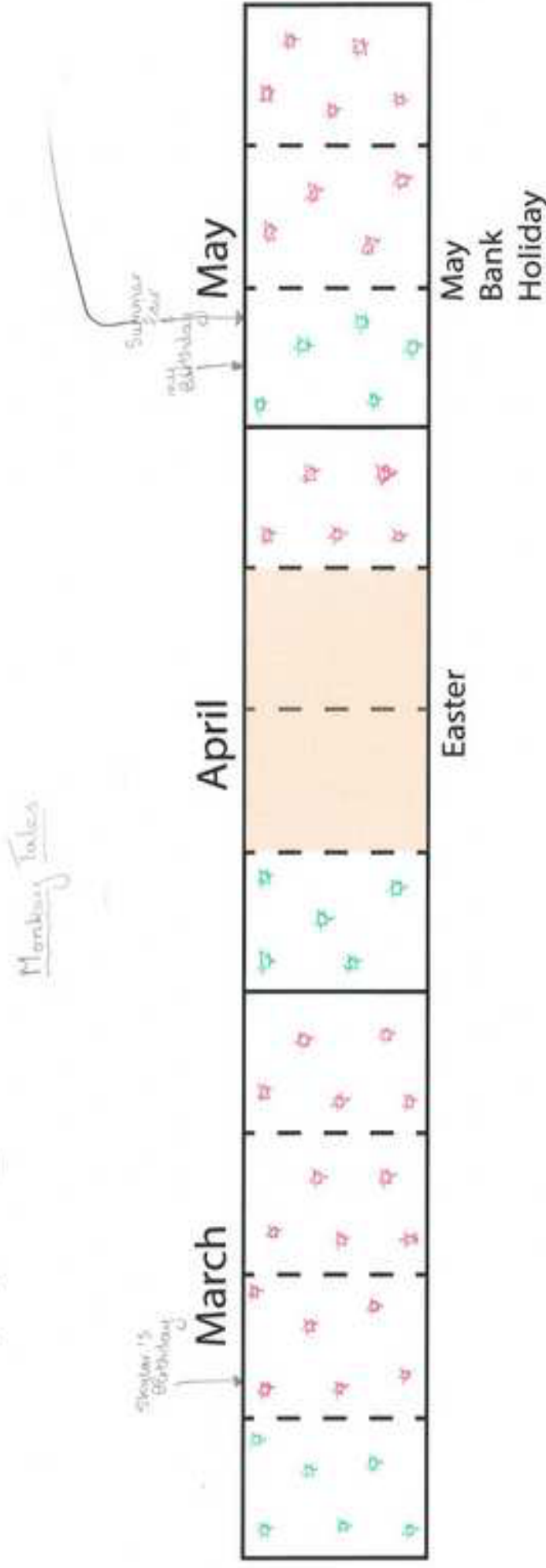
I played the game and I felt challenged.

I played the game but didn't felt challenged.

I did not play the game.

Vond je de liedjes die je kon spelen in Piano Dust Buster 2 leuk/niet leuk?

Changes in constructs over time

Change in MemoLine constructs over time