Middle school children fail to benefit from virtual historical fly-throughs

Liliya Korallo
l.korallo@mdx.ac.uk
Psychology Department, Middlesex University, UK

Abstract

Virtual reality environments (or virtual environment, VE, technology) has been applied successfully in visualisation and in spatial training, where an individual has to navigate among landmarks or in virtual shopping malls. This suggests that it ought to have benefits in imparting temporal-spatial information, and that this might be applied in the learning of historical chronology (or other academic disciplines with a strong chronological component). In several previous studies, attempts have been made to use VEs (as historical fly-throughs, in the form of a time machine) to teach epochs of history to primary school children, mediaeval history to middle school children, and epochs of history for an imaginary planet to undergraduate students. Primary children and undergraduates benefited from the VEs, primary children only when also given adequate familiarisation with the VE medium. Only the middle school pupils failed to benefit. The present studies followed up those findings. In the Ukraine, where pupils had much less computer familiarity, primary and middle school children gave the same results as in the UK, when learning from a virtual fly-through representing Ukrainian history. Middle school children also failed to benefit from using the VE, actually remembering more from PowerPoint use in some cases. Reviews of earlier studies and literature revealed that the inability of middle school children to benefit may relate to the documented changes that begin, specifically at 11-13 years, in their cognitive processing related to space, time and history.

Keywords: virtual reality, chronology, middle school

Introduction

Teaching history has always presented many problems, both for those who teach it and for those who have to learn, understand and remember it. Yet it is important to find ways of putting across historical concepts so that learners can easily grasp them. The UK government is, at the present time, very keen to encourage good citizenship (see the paper by Stewart Martin, in the present conference), but it is hard to know how an individual who cannot fully appreciate the history and development of their country and culture can fulfil their role as a good citizen. History is regarded as a subject that is fundamental to all facets of the educational process. As Davies (1998, p.8) stated, “History … is a lively, challenging, indeed thrilling subject which deserves – and indeed I would say has – to be at the centre of any well-balanced curriculum… the primary purpose of education is to produce well-rounded and sensitive human beings. If that is indeed our belief history must be central in the education of our children”.

Before conducting the present research, our team in Middlesex (myself, Nigel Foreman, Steven Boyd-Davis, Magnus Moar, Mark Coulson and David Newson) carried out some revealing questionnaire studies, aimed at assessing what people remember about what they learned during history classes in school, and how well they could place a number of very important and significant historical events in the correct historical chronological order.
A questionnaire conducted with primary school teachers revealed that pupils have particular difficulty in ordering and sequencing events within their lifetime or within one generation. They may be aware of time periods (such as the Tudor, or Victorian eras) but they do not know where these fit into the bigger historical picture (Korallo, 2010). Understanding chronology (the spatial-temporal relationships among events) should enable children to make sense of their broader historical knowledge, to make it less kaleidoscopic (as commented by the UK Department of Education and Science, 1985). It has been claimed that understanding chronology among events related in time and space is more important than the ability to retain specific dates and time periods (Haydn et al., 2001).

Various strategies have been tried, to attempt to improve children’s chronological thinking skills, both in able-bodied children and also those with learning difficulties of various kinds. Artefacts can also encourage curiosity and visualization of abstract concepts, making them more relevant to children; they can literally feel history, by examining objects and dressing up in period costume, for example (Hoodless, 1996; O’Hara & O’Hara, 2001; Wood & Holden, 1995; Cooper, 2000). Such results suggest that children may be better able to relate to chronology if they have imagined themselves in historical “places”, transported back in time via the imagination, prompted by objects, or transported via tevisual electronic media.

Many history teachers (see Wood, 1995) feel that the collection of items and photographs of the local environment, or studying a particular building or street that have changed over time, or sequencing and justifying artefacts are useful exercises, facilitating the understanding of time and appreciating that time has passed. A discussion group can help children to understand the concept of time in a deeper way, assisting them with expressing historical concepts verbally, making them less remote and abstract. Reading stories and discussing the content of the story, introducing the children to the passages of time through the story, writing essays, newspapers, inviting grandparents to participate in group discussions, talking about their experiences of being school children, discussing and comparing their old photographs with new ones, undertaking a research project that involves visiting sites, archives and other historical places – all these bring a child closer to history (reviewed by Korallo, 2010). These activities can enable children to understand the mechanisms involved in creating a historical process, showing that the child himself or herself is actually a part of the mechanism that “forms” history.

Timelines can help children and adults to learn chronology in a more exciting way. Diderot wrote a descriptive account of the chronological machine invented by Jacques Barbeu-Dubourg that “imagines a combination of several component charts brought together to form a single large one” (Diderot, also cited in Ferguson, 2002). The most influential timelines produced in the eighteen century were the Chart of Biography (1765) and New Chart of History (1769) invented by Joseph Priestly (1733-1804, cited in Rosenberg, 2001). Since timelines have featured as particularly useful ways of introducing history, the present paper focuses on timelines, but in the context of electronic media and multimedia.

The application of effective technologies to improve education in the humanities is urgently needed, yet very few studies have so far attempted to do this (NCAC, 2000). The present study explores the benefit of using VEs in a specific educational setting, with middle school children. Attempts have been made previously to employ engaging multimedia techniques to facilitate learning, such as in the study of Masterman and Rogers (2002). They advocated the use of Interactive MultiMedia (IMM), in which a 2-D time line was depicted, like a road twisting and winding down a computer screen from top to bottom. Children had to follow the route and mouse-click on any image en route to display information about that
specific historical event – it might be a battle, a king or queen, or an invention. Teachers and children said they had enjoyed the experience, although, like many novel classroom applications, there was no formal assessment made that supported the effectiveness of using VEs in terms of learning enhancement. No comparison was made with other available procedures and media. It is important to do this to justify the expense of importing new technology in to the classroom and training teachers in their uses.

Borkowski (2002) claimed that ICT in an educational setting specifically allows the exploration of scientific research and the exploration of real phenomenon. Phenomena can be demonstrated virtually. It arguably improves on the acquisition of knowledge through simply reading textbooks, which many students (faced with electronic media, video games and the like, in their daily lives) find uninspiring (Borkowski, 2002; Eggarsou & Psycharis, 2007). These days, 3-D graphics software and hardware have advanced to the point where semi-realistic views of real world and abstract ideas/concepts can be presented in such a way that they resemble the actual environment (see Calitz and Munro, 2001). One advantage of presenting multi-dimensional historical information in a 3-D environment is that it increases interactivity: “Computer simulations and virtual reality are potentially powerful learning technologies by themselves, offering teachers a means to concretize abstract concepts for students” (see the National Centre on Assessing the General Curriculum report, 2003). Eggersou and Psycharis’ (2007) results lent strong support for using 3-D environments in educational settings. They had participants exploring the Erechtheum in Athens. A 3-D environment allows pupils not only to learn about the dimensions, layout and structure of a famous building, but also explore it. Virtual reality can potentially enable students to visualize new information in a new way and thus helping them to understand abstract concepts in a more concrete way. Generally, 2-D timelines do not give that sense of relationship between historical events, since they are rather presented in categories that might not necessarily have any interlinks with each other. Besides, semantic information can also be visualized in a 3-D format, such as when depicting the history of photography (Kullberg, 1995). It is likely that spatial memory is invoked by such experiences – long routes and long sequences of landmarks can be memorized relatively easily, rather like remembering a row of shops in a frequently navigated street (or, in more complex arrays, a series of parallel streets; see Korallo, 2010). This is likely to be because humans acquire their spatial knowledge, for example of new environmental spaces — large ones not perceivable from a single viewpoint — by travelling through them (see review papers in Foreman and Gillett, 1997). Locations that are represented in terms of spatial memory can be encoded in relation to other locations rather than from a particular standpoint (Hartley et al., 2004). Another advantage is that, once acquired, the spatial relationship between objects can last longer and stay stable over a long period of time when held within a spatial memory store (Hartley et al., 2004), rather as we remember the layout of a familiar town. The hippocampal place system (in the temporal lobe) appears to be responsible for many aspects of topological spatial memory, and this system is apparently invoked when a VE is navigated. There are sometimes disadvantages in using VEs; for example, distances tend to be underestimated (Witmer & Klein, 1998), especially in females (Foreman et al., 2004) but this is unlikely to affect the retention of information gathered from a simple, linear, sequential fly-through. Overall, the many studies conducted using VEs indicate a good correspondence between the spatial knowledge obtained from the VE and equivalent knowledge of a real environment, so that spatial information gathered virtually is likely to be remembered just as successfully (Arthur et al., 1997; Regan & Yardick, 1994;
Schwienhorst, 2002). Magnus Moar will be talking about the applications of real world spatial memory to memory for sequences, elsewhere in this meeting.

The hypothesis that a VE can be an effective medium to teach historical chronology by engaging spatial memory may, therefore need to be qualified depending on participant variables such as age and gender. In particular, in earlier studies, primary school children have been found to perform poorly when given passive fly-through experience, but to perform well (in the same way as undergraduate students) when exposed to a fly-through of historical-sequential events with challenge (Foreman et al., 2008; Korallo, 2010), at least compared with media such as text and PowerPoint. Challenge (anticipation of up-coming items) allowed error free learning from the VE based training. The present study further explores the use of VEs with challenge incorporated, but with middle school children, another age group for which passive fly-through VE training was ineffective in an earlier study (Foreman et al., 2008). Further, pre-training and familiarity with the environment are of interest (Sandamas & Foreman, 2007; Sandamas et al., 2007), since children given training with a VE interface device in advance of training performed better on the training task. Studies with primary school children who were allowed both pre-training experience and challenge incorporated into the environment (Korallo, 2010) showed a better recall of historical information presented chronologically compared with previous groups tested passively (cf., Foreman et al., 2008). The present studies, therefore, looked at the performance of middle school children in the UK and in Ukraine, the latter having less computer familiarity than in the UK, with either one or two sessions of pre-training familiarization with the medium.

Experiment 1: Methods and Procedures

In a first study, thirty (15M, 15F) 12 year-old pupils were drawn from a Ukrainian middle school. The average age of the participants was 12 years old at the time of testing, with good sight. The study was complementary to the Ukrainian National Curriculum so only teacher approval was required. Pupils’ typical computer use averaged 1.5 hours per week. The VE was creating using Virtools Virtual Reality software running on a (Microsoft Windows XP Professional Version 2002 computer. The environment ran in Microsoft Explorer with a Virtools 3-D plug-in. A set of nine pictures was used, representing historical epochs in the history of Ukraine (see Figure 1 for an example). All pictures were named and dated, some depicting well-known historical personalities in Ukrainian history; others showed the famous historical places in the Ukraine. Participants moved forward only in virtual space, and returned to the start by pressing another keyboard key.
For the PowerPoint condition, the same 9 pictures were used, the presentation being as for the VE condition but using successive PowerPoint slides, images separated by blank screens. In the control text condition, the pictures were printed on A4 sized sheets and presented to the child in landscape orientation with text added as in other conditions.

Children were randomly divided into three groups in a Paper group (N=10; 4M, 6F), a PowerPoint group (N = 10; 5M, 5F) and a VR group (N = 10; 5M, 5F) on the advice of teaching staff, so that there was a similar level of ability in history in each of the groups. Participants in the VE group were introduced to the VE one at a time by the Experimenter and shown how to depress the space bar to move through the environment. All pictures were visible when they did this. The participant was then asked to perform the second phase (the training phase), in which they had to anticipate each up-coming picture. The researcher asked, at each blank picture screen, "Which picture is going to appear next?" Responses (which could be a picture name or a description) were recorded as correct or incorrect. If wrong, the participant had to guess again until correct. The number of times that the participants had to pass through the 9 items to reach criterion – two passes without errors -- was recorded, after which they proceeded to the next stage, i.e., the testing stage. On average participants completed four fly-throughs to achieve criterion. For testing, participants were taken to an adjacent desk on which were the 9 items (pictures but without dates or identities) randomly ordered. They were asked to place the printed pictures in correct order along the desk. Participants controlled the appearance of the images in PowerPoint and VE conditions, and the Text condition was as similar as possible to the other two, the same anticipation routine being used in all cases.

When training was complete, the participant was taken to an adjacent set of desks on which were placed the 9 test items, in random order. Participants all took about 2 minutes to place these in the correct chronological order as seen in the training stage, and took 5-7 minutes to complete the whole task. They were given fruit rewards. After two weeks the testing was administered again to investigate which condition was more effective in terms of remembering items in their correct order.
Experiment 1: Results

Two scores were obtained during the initial training: (1) The Number of trials (passes) to criterion and (2) A Total error score, summing all errors committed prior to reaching criterion. Two further measures were obtained from the initial post-criterion testing: (3) REM score, or “removed score” [see Foreman et al., 2008] and (4) Correct order [the number of pictures placed in their correct position in the 1-9 sequence]. The further two scores were obtained when testing was repeated 2 weeks after the original training and testing phases: (5) REM1 [removed score], and (6) Correct order 1 scores (5) and (6) being calculated in the same ways as (3) and (4). Also, 7) Serial order effects were examined by recording the number of pictures placed correctly at the beginning of the list (primacy effect: list positions 1-3), in middle list positions (4-6) and in at the end of the list (7-9).

An independent one-way ANOVA was used to analyze the variables: Number of trials, Total error, REM 1, Correct order and Correct order 1. However, no statistical significance was found on any of the measured variables. A non-parametric Kruskal-Wallis test was used to analyze the REM score (the data were not normally distributed) but again, the three groups did not differ significantly.

Gender differences were analysed but also found to be non-significant. When a correlation was performed between the Ability variable (teacher’s ratings) and Correct order 1 (number of items remembered when the same test was repeated after 2 weeks), an almost significant negative correlation was revealed between these two variables $r(28) = -.361$, $p = .059$. No significant correlation was found between REM and Ability scores.

Since there were no significant differences among groups on any measure, we were forced to conclude that either that VR environments need to be further improved, or alternatively that they might just not be effective with the middle school age group. Throughout the research, statistical analysis revealed that participants from the Power Point condition suffered most in terms of remembering events presented in chronological order, in agreement with previous findings (Haydn, 1999). Gender differences were also non-significant, suggesting that female and male benefited equally well from new technologies (cf., Coluccia et al., 2007). Serial position effects also showed no differences among groups.

The following experiment examined the effectiveness of VE use in middle school children in the UK, having greater computer familiarity.

Experiment 2: Procedure

In Experiment 2, forty-nine middle school pupils (12-14 years) from North London, with good vision, were randomly selected by teaching staff to participate. The study was conducted as for Experiment 1 but with materials selected by history teachers in the school, relating to the UK national curriculum history syllabus. Conditions were as for Experiment 1 and the same procedure was applied, children being randomly divided into Paper (N = 16; 9M, 7F), PowerPoint (N = 15; 7M, 8F) and VE (N = 18; 10M, 8F) conditions. The only respect in which the experimental design differed from Experiment 1 was that after an interval of one month participants were asked to repeat the study, undergoing both training and testing stages. The rationale for the additional training and testing was to expose the participants to the active learning of the materials twice, thus hopefully increasing recall and allowing greater differentiation among the experimental conditions. Upon completion of training in each phase, measurements were taken and after the second testing session.
**Experiment 2: Results**

Two variables were analyzed in the initial Phase 1 stage (initial training and testing): Total error scores, and Total number of trials (to meet a “two successive correct” criterion). The Total number of trials variable was analyzed by using an independent one-way ANOVA (the data met all parametric requirements). The result showed that groups differed highly significantly $F(2, 46) = 10.35, p < .001$, but post-hoc tests showed that, surprisingly, those participants trained in the VE and PowerPoint conditions required more, and not fewer, passes through the environment to meet the criterion than the Paper condition ($p$’s $< .001$; Figure 2).

<table>
<thead>
<tr>
<th>Condition</th>
<th>Mean number of passes to meet criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>Paper</td>
<td>6.00</td>
</tr>
<tr>
<td>PowerPoint</td>
<td>5.00</td>
</tr>
<tr>
<td>Virtual Environment</td>
<td>4.00</td>
</tr>
</tbody>
</table>

![Figure 2. Mean number of passes to criterion in initial training in Experiment 2](image)

A Kruskal-Wallis non-parametric ANOVA showed that in terms of REM (since data were not normally distributed), there was no significant difference observed amongst the three conditions. Participants performed equally well in all three conditions. The Correct order variable was also found to be non-significant. Serial order effects were investigated (comparing blocks 1-3, primacy; 4-6, middle; and 7-9 recency) by using a Kruskal-Wallis test. There was no statistical difference observed on middle and recency position blocks, although the primacy position was found to be significant $X^2(2) = 7.75, p = .021$. Participants placed more items correctly in the PowerPoint condition compared to the Paper condition. The Gender variable showed no significance on any measure.

When children were retrained and retested in the second phase, the Correct order 2 variable (as correct order 1, but for phase 2 data) was found to show a significance, participants trained in the PowerPoint condition placing significantly more items correctly than those trained in the VE condition. Other variables did not yield any statistical differences.

Serial order effects were examined by using the Kruskal-Wallis test. No significance was found in the middle or recency position blocks. However, there was a significant difference in respect of items placed at the beginning of the list, $X^2(2) = 7.31, p = .026$. A further analysis using the Mann-Whitney test showed that the participants who were trained in the VE condition placed more items correctly in early list positions (the 1-3 position block) than their PowerPoint counterparts, $U(N1=17; N2=16) = 87, p < .02$. 
The gender variable was not significant for most variables, although when a t-test was performed on the primacy 2 position block, it showed that male participants placed more items correctly here than the female participants \( t(46) = 1.95, p < .001 \).

## General Discussion

The present study was interesting and enlightening, given the background of research on which it was based. It raises important questions about the use of VEs in an educational setting and with different age groups. First, the present study repeatedly confirmed that for middle school children, VE presentations of temporal-spatial information are not effective. VE participants required more passes through the environment to meet a performance criterion. The Correct order 2 variable showed an interesting feature, in which the participants who were trained in the PowerPoint condition did — contrary to the previous findings — produce a positive effect, insofar as PowerPoint participants placed more items correctly than the VE participants when retrained and retested. The only exception was that when serial position data were analysed, participants who trained in the VE condition placed more items correctly in list positions 1-3 than their counterparts in the PowerPoint condition. Despite the repeated use of the training and testing stages, the results did not show any evidence of greater consolidation of chronological memory after VE experience, implying that there may be other reasons as to why middle school children did not benefit in the way that other age groups (primary, and undergraduate) did.

The present studies were negative, but they have yielded important conclusions. They are consistent. They have demonstrated that VEs, while apparently effective with children of primary age and with undergraduate students, are notably not effective when used with groups of middle school pupils. At least, we can say reliably that other age groups profited from the use of VEs when challenge and familiarity with the VE were incorporated, but children aged 11-13 years old still failed to benefit from using VEs when learning historical chronological sequences of materials. This is further emphasised by the fact that, in the second study reported here, the participants were allowed to explore the environment for longer by being given familiarity experience at the outset but also by virtue of being trained and tested twice. It was expected that this extra exposure should have enhanced differences among the conditions, because the VE medium might be expected to be even more impressive; we expected that children would have the images of historical events more strongly established when encountering them in a VE fly-through in different spatial locations. However, clearly the VE medium did not provoke the participants to perform better, or to show any lasting effect (i.e., any greater impact) after using a VE. Perhaps surprisingly, given many authors’ general scepticism about the use of PowerPoint (cf., Haydn, 2003), the present study demonstrated that children in the retest part of the experiment actually showed a better understanding of materials learned from a PowerPoint format than from a VE. Finally, it is important to note that the gender variable showed no significance, so that materials learned in a 3-D environment were equally well remembered by both sex groups, in agreement with several previous findings (Sandamas et al, 2007; see Korallo, 2010 for a review). It is not the case that boys’ generally greater experience of computers or familiarisation via computer game formats, makes them more open to learning in a virtual environment, at least not in terms of the learning of historical chronological sequences.

The most curious question is why middle school (11-13 year-old) children fail to benefit from VE training. There are various possibilities suggested by previous studies, in particular that children of middle school age are undergoing important changes in their cognitive
processing that may impact on their use of spatial-temporal processing. This may apply specially to historical chronology. Although temporal understanding develops by 8-9 years (Oakden and Sturt, 1922) and labels such as dates, past, present and future, begin to assume meanings, time epochs are first used as specific descriptors only around 10 years (Levstik and Pappas, 1981). Placement of materials in correct chronological order is only something that begins around 11-12 years and it continues to mature between 12 and 16 years (Oakden and Sturt, 1922). Jahoda (1963) argued that not until age 11 do children understand the implication of historical dates, and similarly Friedman (1982), from classroom studies, argued that a total and global understanding of time begins after 11 years. Although 11 years is a start point, full understanding of chronology is only evident from 16 years (Flickinger and Rehage, 1949; Oakden and Sturt, 1922). From all these accounts it seems as though at age 11, children are undergoing an important transition which is likely to enable them to begin to process truly “chronological” materials, embarking on a different way of processing chronological information. In other words, primary and middle school children are performing the VE task differently and more subject to overload in their available cognitive capacity: primary children are likely to be using a more pictorial representation, but middle school pupils attempt to apply a truly “chronological” analysis, but they do it incompetently at this stage. Harner (1982) made a similar argument: that by 10-11 years, children have mastered the various linguistic structures related to time although there must follow a period in which they must learn to apply the terminology correctly. Such a limitation would lead to what amounts to an information overload, and might relate to hormonal and other changes taking place in children of this age (Cockburn & McKenzie, 2001; Flickinger & Rehage, 1949; Friedman, 1982; Jahoda, 1963; Prangsma et al., 2009).

The relatively small sample sizes used in the present study were not responsible for the absence of significant group differences. In previous studies, significant effects, where they occurred between VE, PowerPoint and Paper/text groups were evident with comparable sample sizes. Moreover, in the broader sense, to recommend the VE procedure to be useful it would need to be demonstrably effective with typical class-sized groups and sets of 10-15 pupils, and therefore research with such group sizes is most appropriate. Moreover, where significances were obtained in the present study, paradoxically these favoured the PowerPoint and Paper conditions and were therefore in a direction opposed to the experimental hypotheses. Importantly, there were no marginal significances in the direction predicted by the experimental hypotheses.

Conclusions

Virtual environment technology has a future in classroom learning of subjects with chronological components, at least in primary and adult age groups. Middle school pupils have specific difficulty in using VEs to remember sequences, perhaps due to their having only recently acquired the use of spatial-temporal codes, which they use inefficiently. Future research should concentrate more on how to present information in VEs in a variety of forms, that would be suitable for children of all ages. This might require children’s own construction of timelines using click and drag facilities, to improve active engagement with the environment, or the use of more spatial multiple parallel time lines. Repeatedly returning to a familiar VE might also be effective, perhaps improving the longevity of memories acquired from historical VEs.
Acknowledgements

The authors are grateful to the Leverhulme Trust for financial support (Grant F/00 765/D)

References


