Evaluation of multimedia educational materials using eye tracking

Oscar Navarro\textsuperscript{a}\textsuperscript{*}, Ana Isabel Molina\textsuperscript{b}, Miguel Lacruz\textsuperscript{c}, Manuel Ortega\textsuperscript{b}

\textsuperscript{a}Guidance Department of San José de Calasanz School, Ciudad Real 13700, Spain
\textsuperscript{b}Technologies and Information Systems Department, Castilla-La Mancha University, Ciudad Real 13071, Spain
\textsuperscript{c}Pedagogy Department, Castilla-La Mancha University, Ciudad Real 13071, Spain

Abstract

In the last two decades there has been a significant technological development in all areas of society. This technological advance has been transferred to the educational area and now the use of multimedia materials and resources in the teaching and learning process is widespread. The main objective of the study presented in this article is to show the use of an innovative educational materials evaluation tool: the eye tracking technique. The use of this technique makes it possible to evaluate aspects of the format in which such materials are designed as well as aspects related to its usability. An appropriate design of multimedia materials may favour the acquisition of knowledge by pupils. Two experiments involving an eye tracking device in the data collection process were carried out in order to show the application of this evaluation technique. Two homogeneous groups were shown a presentation in which the format varies slightly. The presentations in the first experiment included images and texts as well as color combinations with different contrasts. In the second experiment, a presentation with information in black and white was compared with another in which colors are used. Once analyzed the data it is concluded that a more efficient learning occurs when information is shown using warm colors instead of pastel shades and when color is used to highlight some areas of interest in which we want the pupils to focus their attention.

Keywords: Multimedia resources; eye tracking; Color; Primary education; Assessment; Efficiency

1. Introduction

The important technological advances occurred in recent decades have led to a breakthrough in virtually all areas of knowledge. Within the field of education, it is becoming increasingly common to use multimedia materials and
resources as essential tools in the teaching and learning process. These elements, such as interactive whiteboards, netbooks, and Internet, call for an improvement in the development of multimedia materials and resources. The multimedia material we use is often unnecessarily difficult to understand and complicated to use. The difficulties in handling the software require a need of time and effort that sometimes causes frustration, therefore discouraging the use of these resources. It is very difficult to specify the features that these materials should present but some guidelines can be provided. Besides, the usability of a product depends also on its context of use (Bevan & Macleod, 1994). Cervera asserts that the teaching materials are not good or bad in an absolute sense: depending on the case some are more effective than others. In order to choose appropriately, it is essential that the teacher clearly identifies the goals to achieve, contents to address, how to do so, and the individual situation of teacher and students (Cervera et al., 2010). It can be stated that multimedia materials and resources have “as their main purpose to bring students closer to the reality of their learning, augmenting the information representation”. It will be intended to integrate current technologies in the teaching and learning process of the student to achieve the didactic objectives set (Poveda, 2010). The future lines of work in the education field are transitioning from the traditional materials to the incorporation of new media, such as e-books and other materials in electronic format (Andra et al., 2009). This whole environment requires continuous updating and adjustment to the requirements and demands of the emerging technologies, and fosters an educational effort aimed at assimilating the competence derived from the use of digital technologies. The opposite case implies “becoming a new technologically illiterate” (Area, 2009).

The use of the technique of eye tracking is included in these studies. It refers to a set of technologies that make possible to establish the eye gaze of an individual. This technique allows us to record the number of eye fixations and stabilization of a gaze point, which a student experiences while viewing content displayed on the screen. This device, the eye tracker, uses infrared beams which are first reflected on the pupil of the individual and then recorded. This makes it possible to calculate precisely where the individual was looking at. This allows determining which areas receive more attention, for how long and the order followed in the visual exploration. From these fixations it is possible to calculate a series of metrics. The series of metrics will include the number of fixations and the time until the first fixation on a certain part of the screen, which we call area of interest (AOI), as well as the total time of all fixations made. The AOIs are the parts of the screen that contain the most relevant information of the texts and images displayed. The metrics from AOIs will allow us to evaluate and compare different configurations of the materials displayed. The eye tracking has a great potential for application in a wide variety of disciplines and areas of study, from marketing and advertising to medical research and psycholinguistics, as well as usability studies (Hassan & Herrero, 2007). Although there are fewer studies related directly to the field of education, the interest in this specific field seems to be growing. Most of the existing research focuses on the stages of secondary and higher education. While it is scarce, we can also find recent publications containing pupils of primary education and newborns (Mayer, 2010; Sim, Cassidy, & Read, 2013; Mason, Tornatora, & Pluchino, 2013).

There are a number of works which cite the Principles of Multimedia Learning by Richard Mayer (Mayer, 2005) (Mayer, 2009). We will evaluate one of the principles in this study, which can be found and further described in the previously stated article. The Signaling Principle refers to the fact that learners learn more deeply when cues are added to highlight the organization of the essential contents. This research aims to investigate the use of color in multimedia materials. According to some authors, in a multimedia presentation it is important to use colors that are easily differentiated between and to also use a high contrast to distinguish them, such as the saturation, brightness and tone (Ware, 2008; Johnson, 2010). Primary and warm colors are recommended for use in materials designed for children (Romero, 2003). There are other studies with adults that used eye tracking techniques to analyze the different possibilities of color to promote the search for relevant information from text and/or illustrations. This methodology has been used to analyze the response of individuals when they look at different combinations of the most representative colors such as black, white, yellow, red, blue and green (Garcia-Hernandez, 2008). Also, it has been compared the use of color against white to mark an area highly relevant for the participant (Boucheix & Lowe, 2010). In the same line as this research, Ozcelik et al. study color coding to promote a more effective learning, expecting the participants to find linked items between text and illustration in a color format as opposed to a conventional one (Ozcelik, Karaku, Kursun, & Cagiltay, 2009).
2. Purpose of the study

The goal of this experiment was to investigate the effects of color in a multimedia presentation at the stage of primary education. This research was conducted in a school of primary education at Ciudad Real (Spain). The school board and tutors of the second and sixth grade of primary education collaborated in this study. These two groups were chosen since they represent two key evolutionary moments with very pronounced differences from the developmental and psycho-evolutionary point of view. This study intends to analyze and assess the use of color in a multimedia presentation at the stage of primary education. Our interest is evaluating the most efficient and effective settings and presentation formats of educational materials and resources involved in the teaching process. The two experiments use different color combinations. Our specific research hypothesis was:

\[ H_1: \text{Greater efficiency in content retention will occur if the color in a multimedia presentation is properly used.} \]

3. Methods

3.1. Participants

Initially, 89 primary school students in second and sixth grade were involved. They were able to speak and understand Spanish perfectly. Due to visual limitations of 10 of the participants, we only considered the data of 79 students (42 males and 37 females), with an average age of \( M = 7.43 \) (SD = 0.31) in second grade and \( M = 11.61 \) (SD = 0.38) in sixth grade. A preliminary study involving 8 students was carried out in order to adjust the length of the presentations so they reflected the appropriate duration. Also, several calibration tests for the eye tracking device (eye tracker) were run with second grade students, since these students tend to present a more dispersed attention.

3.2. Materials

The materials presented to the students consisted of slides prepared using different software (Adobe Flash CS5, PowerPoint 2007, Gimp and Paint). The materials required to conduct every experiment and verify the hypothesis consisted of pictures with different color combinations. At the end of the presentation, the participants completed an activity intended to determine the level of assimilation of the contents displayed. These contents are further explained in annex II of the Royal Decree 1513/2006 of 7 December, which establishes the core teachings in primary education in Spain. The contents mentioned above belong to the third block of contents in the area of Mathematics (Geometry), which refers to "the identification and description of flat shapes, their parts, spatial location in terms of angles and spins, and the recognition of symmetries and regularities" (\textit{RD 1513/06, sec. BOE} 293). There are other studies dealing with the area of Mathematics (Andra et al., 2009; Moreno & Duran, 2004) and some of them include explicit references to geometric figures (Garcia-Hernandez, 2008). The block of contents referred to in the materials displayed for both grades is the same. However, they are different in nature due to problems that emerged in the data collection of 6th grade and therefore it was necessary to replace the initial presentation.

3.3. Apparatus

A Tobii device model X60 was used for the eye tracking, which includes specific software (Tobii Studio 3.0.5.301) required for calibration, presentation design, collections of data and the subsequent calculation of the different metrics. This study was conducted in four of the classrooms of the school.

3.4 Procedure
Figure 1 shows the schema of the research and consists of a graph detailing the process followed in the two experiments. First, we used the method of Quota Sampling (Kish, 1972) due to the fact that the individuals and their personal characteristics were previously known. Intelligence Quotient (IQ) and Learning Style were taken into account in order to carry out this process. The Kaufman Brief Intelligence Test was required to know the IQ of the students. The test was performed individually with an approximate duration of twenty minutes. This test allows for the differential measurement of verbal and non-verbal intelligence. Besides, it was employed the Felder-Silverman Learning Style Model with the students. This test consists of 44 questions which provided information on their learning style. The language used was adapted to the age of the students and explanations were given when a question was not fully understood, something required mainly in the case of second grade students. The objective was to establish two homogeneous samples from two tests (K-BIT and Felder Test) and for each of the grades selected (second and sixth). From these two samples was set an experimental group (Group 2), in which we applied formats with a good combination of colors, and a control group (Group 1) with a poorer color quality combination. After this and now in an individual way, the students answered a questionnaire about previous knowledge of the contents displayed. The questionnaire, consistent with other similar practices (Mason, Tomatora, & Pluchino, 2013), consisted of open-ended questions where pupils were asked to draw and write their knowledge of the contents. The eye tracker calibration phase began after the questionnaire was completed. It proved very favorable in virtually all the cases obtaining measures of quality of sampling above 90 %. After this the presentation of the content shown on the monitor began. Students completed a posttest after the content presentation in order to assess the knowledge acquired. The activities involved had the same structure and characteristics as the ones the students were used to at the school. The students were given instructions before the material display, always in the same order and using simple language and explanations. More precise instructions were given second grade students since their evolutionary development requires a different treatment than sixth grade children. Figure 1 shows the dependent and independent variables involved in the research.

The first experiment focused on the evaluation of the use of color in a presentation of images and text. The experimental group (Group 2) observed an image with colors that contrast well, which favors the perception of contents (Johnson, 2010; Ware, 2008). Primary and warm colors were used since they are preferable to create content aimed at children than pastels such as those presented for the control group (Group 1) (Romero, 2003). It was taken into consideration Richard Mayer’s Signaling Principle, which refers to the need to add signals that highlight the organization of relevant elements. The hypothesis intended to test in this first study are as follows: \( H_1 \) - There is greater efficiency in content retention if we include a presentation format with colors that contrast well.

In the second experiment, the Control Group is shown a black and white image and text whereas some important
areas of the same slide are highlighted with color for the Experimental Group. When students spot links between text and illustrations, a better retention of content is achieved. The use of color instead of black and white helps the participants to find the required information (Ozcelik, Karakus, Kursun, & Cagiltay, 2009). The Signaling Principle, previously mentioned, is also taken into account to carry out this experiment. The hypothesis intended to test in this first study are the following: H2 - Greater efficiency in content retention will occur if an area is highlighted with color in a black and white background instead of using solely black and white.

3.4. Data analysis

The data collected in the different tests and eye tracker metrics were analyzed with Excel and SPSS for statistical analysis. A unilateral contrast of averages was conducted to test the hypothesis, as well as the analysis of variance and the calculation of the Student's t-distribution. The confidence levels applied were from 95% to 99% and, due to the fact that the sample was not very large, the results were considered statistically relevant with a level of 90%.

The data obtained in the posttest and the time spent by students was compared with the data obtained with the Tobii device. There are a variety of metrics that measure attention, complexity or usefulness of the information contained in an AOI (Jacob & Karn, 2003; Poole & Ball, 2005; Birkett, Galpin, Cassidy, Marrow, & Norgate, 2011; Mason et al., 2013; Bojko, 2013), among which the ones in Table 1 were selected. Results can be obtained from several metrics depending on the device and software used. However, due to the diversity and number of available metrics, the choice of the most convenient ones proves difficult since there is no unanimity on whether to use one or another (Hyönä, Lorch, & Rinck, 2003). The eye tracking device allows the registration of other metrics such as the total length or number of fixations in the A OIs, duration of the first fixation or the number and duration of the visits to each AOI. However, only metrics that specify whether there is a greater or lesser efficiency in learning have been taken into account for this study.

Table 1. Areas of interest (AOI). *Detail of Fixations on images (Im), text (Tx) and taking into account both (all).

<table>
<thead>
<tr>
<th>Eye-movement measure</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time to First Fixation (TFF)*</td>
<td>Time to first fixation on an AOI</td>
</tr>
<tr>
<td>Fixations Before (FB)*</td>
<td>Number of fixations prior to first fixation on an AOI</td>
</tr>
<tr>
<td>Other measure</td>
<td>Definition</td>
</tr>
<tr>
<td>Fixation Count (All-Sc)</td>
<td>Number of all fixations on screen</td>
</tr>
<tr>
<td>Fixation Count (FC/All-Sc)</td>
<td>Percentage of Fixations on-target</td>
</tr>
</tbody>
</table>

4. Analysis and results

4.1. Presentation Format With Contrasting Colors

Table 2. Eye tracking metrics of Experiment 1

<table>
<thead>
<tr>
<th></th>
<th>Im</th>
<th>Tx</th>
<th>All</th>
<th>Im</th>
<th>Tx</th>
<th>All</th>
<th>FC/All-Sc</th>
</tr>
</thead>
<tbody>
<tr>
<td>EXPERIMENT 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Time First Fix.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>G1 2º</td>
<td>0.74³</td>
<td>0.70³</td>
<td>0.41³</td>
<td>6.26</td>
<td>6.68</td>
<td>3.11</td>
<td>57.68 0.68</td>
</tr>
<tr>
<td>G2 2º</td>
<td>0.55³</td>
<td>0.51³</td>
<td>0.22³</td>
<td>6.70</td>
<td>6.60</td>
<td>3.10</td>
<td>49.50 0.70</td>
</tr>
<tr>
<td>G1 6º</td>
<td>0.17³</td>
<td>4.90²</td>
<td>0.15³</td>
<td>0.75²</td>
<td>33.45³</td>
<td>0.55²</td>
<td>81.75³ 0.97</td>
</tr>
<tr>
<td>G2 6º</td>
<td>0.02²</td>
<td>3.86²</td>
<td>0.01³</td>
<td>0.15²</td>
<td>27.15³</td>
<td>0.00²</td>
<td>69.80³ 0.97</td>
</tr>
</tbody>
</table>

Figure 2. Time and Posttest of Experiment 1 ¹ p<0.01 ² p<0.05 ³ p=0.1
Figure 2 shows Group 2 (Experimental), in yellow, which represents the students who are shown images and text with contrasting colors, and Group 1 (Control), in green, with a configuration that makes the perception difficult for the students. As can be seen, the time spent by children of second grade to observe the screen is significantly different. Using the value calculated by the Student's t test with a significance level of 0.05, we get a value of $t=1.82$ ($p=0.041$). It can therefore be asserted that for Group 2 the observation takes significantly less time. The results of the posttest for second grade are higher for Group 2. Nevertheless, after carrying out statistical calculations, it is observed that the values obtained are smaller than the critical value. Therefore it can be stated that no significant differences are found. Similar data is obtained in the case of sixth grade. The time spent by Group 2 is significantly shorter, with a value of $t=1.81$ ($p=0.039$). The results of the posttest are virtually identical and there are no significant differences as in the previous case.

As can be seen in Table 2, the metrics provided by the eye tracker, for second grade, show that there is no significant difference in most of the metrics. Similar results are obtained in the fixations before focusing the gaze in a AOI (FB), the number of all the fixations on screen (All-Sc) and the proportion of fixations in an AOI (FC/All-Sc). However, it can be stated that for Group 2 the duration is lesser for all fixations before focusing the gaze into the AOIs (TFF), applying for images, texts and the total. Nevertheless, these statements can be made with a significance level of only 0.1. There are also another metric without significant differences in sixth grade, the proportion of fixations in an AOI (FC/All-Sc). It is lesser for Group 2, although with different level of significance, the duration (TFF) and the number of fixations (FB) prior focusing the gaze into the AOIs, in the case of the total, images and text. There are also a smaller number of all fixations on the screen (All-Sc) for the Experimental Group.

The first conclusion after analyzing the results is that the scores in the posttest are similar both for second and sixth grade, but the duration is shorter for the Experimental Group (which observes a format with contrasting images and colors, favoring the retention of content). Therefore, in both educational levels a greater efficiency in learning is achieved, as the Signaling Principle of Mayer states. These results are confirmed, in greater or lesser extent, by reviewing the metrics recorded by the eye tracker. For second grade students, Group 2 (Experimental) focuses the gaze in the AOIs in a shorter time. A similar trend is observed for sixth grade. In this case students in the Experimental Group focus their gaze in a shorter time and with fewer number of fixations into the AOIs. In addition, the total number of fixations in screen is smaller for Group 2. These data indicate a greater efficiency in the search for areas of interest at both educational levels.

4.2. Use Of Color To Highlight An Area Of Interest

Table 3. Eye tracking metrics of Experiment 2

<table>
<thead>
<tr>
<th></th>
<th>Time First Fix.</th>
<th>Fix. Before</th>
<th>Other</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Im</td>
<td>Tx</td>
<td>All</td>
</tr>
<tr>
<td>G1 6º</td>
<td>1.24</td>
<td>0.54</td>
<td>0.13</td>
</tr>
<tr>
<td>G2 6º</td>
<td>0.46</td>
<td>0.61</td>
<td>0.12</td>
</tr>
<tr>
<td>G1 2º</td>
<td>2.80</td>
<td>1.78</td>
<td>1.15</td>
</tr>
<tr>
<td>G2 2º</td>
<td>1.48</td>
<td>1.10</td>
<td>0.33</td>
</tr>
</tbody>
</table>

Figure 3 (Group 2, in yellow) represents the students shown a presentation in black and white with some areas highlighted in color. On the contrary, Group 1 (Control, in green) is presented the same slides but in black and white only. For second year students it is observed a clear difference in the total length when looking at the screen. T-value=2.06 ($p=0.023$) is greater than the critical value, with a significance level of 0.05. Therefore, it can be stated that Group 2 observes the screen for a shorter time. With regard to the results of the posttest, the Experimental
Group gets a higher value than Group 1. It also confirms this difference, since the value of the Student's t-test is greater than the critical value t=1.57 (p=0.062), but this can only be asserted with a significance level of 0.1. With regard to 6th grade, it is also noted that the time used by Group 2 is lower than for Group 1. It is confirmed since the values obtained are of t=1.44 (p=0.081) with a significance level of 0.1. As for the posttest, the Experimental Group obtained a slightly higher score, although significant differences are not noted.

Table 3 shows the metrics provided by the eye tracker. In second course it is observed that there are no significant differences in three of them, the average duration of fixation in each AOI (FD) both in texts and in total, as well as the total number of fixations on the screen (All-Sc). On the other hand it can be stated that for Group 2 the duration is lesser (TFF) and the number (FB) of all fixations before focusing the gaze into the AOIs. It is also obtained for Group 2 a higher value in the proportion of fixations in an AOI (FC/All-Sc). For sixth grade most of the metrics do not show significant differences. Lower values for the Experimental group are recorded only in the time spent (TFF-Im) and number of fixations (FB-Im) before focusing on the images.

The Experimental Group represents children who watch a presentation where color is used to highlight some relevant elements. These students spend less time watching the presentations at both educational levels. Besides, second grade students obtain a higher result in the posttest, although only with a significance level of 0.1. Therefore, a greater efficiency in learning is achieved, as the Mayer’s Signaling Principle states. However, this efficiency is more relevant in second grade. After reviewing the data recorded by the eye tracker it is noted that there are some relevant differences between the two education levels. For second grade, students in the Experimental Group focus their attention on the AOIs in less time and with fewer fixations. The proportion of fixations in an AOI is also higher for Group 2. In addition, in some of these metrics the significance level is 0.01. In sixth grade smaller values are only obtained, for the Experimental Group, in the amount of time spent and the number of fixations before focusing the gaze on the images. Therefore, taking into account the total time of observation, the posttest and the eye tracker metrics, it can be stated that the students to whom it is displayed a presentation with some areas highlighted with colors achieve greater efficiency in learning. But with greater relevance in second grade.

5. Discussion and conclusions

Color is an element commonly used in textbooks of Primary Education. A proper encoding can increase the efficiency of information processing and reduce the process of search (Folker, Sichelschmidt, & Ritter, 2005). This study intends to analyze the efficiency in learning when the students are shown a multimedia presentation with different formats that include color. It has been already mentioned the importance of preparing appropriate materials to ensure compliance with the educational objectives. We must pay special attention to the choice of appropriate combinations of colors in order to attract the attention of our students (Garcia-Hernandez, 2008). The combination of colors in which the areas of interest (text and images) contrast well makes it possible to obtain similar results with respect to the contents learnt. However, the visual exploration requires less time in both second and sixth grade. Besides, this format favors an anticipation of the content exploration since the Experimental Group focuses before in relevant elements. These students spend less time watching the presentations at both educational levels. Besides, this format favors an anticipation of the content exploration since the Experimental Group focuses before in relevant elements. These students spend less time watching the presentations at both educational levels.
Although this hypothesis is accepted for both educational levels, it is more evident in the case of second grade. The results of the second experiment make it possible to accept hypothesis H2 for both educational levels, although it is more relevant in the case of second grade.

Acknowledgements

The authors would like to thank Group CHICO of the UCLM (University of Castilla-La Mancha) for facilitating the use of the Tobii X60 device, as well as the school board, tutors, and pupils of second and sixth grade from CEIP San José de Calasanz for having participated in this research. This research has been partially funded by the Ministry of Science and Innovation through the project “Software environment for learning to program in group and its integration, by means of standards, with learning management systems” (REF: TIN2011-29542-C02-02) with evaluation of ANEP, the Thematic Network 513RT0481 and the JCCM Project PPEII-2014012-A GITE-Learn.

References


