Computers & Education xxx (2009) xxx–xxx

Contents lists available at ScienceDirect

Computers & Education

journal homepage: www.elsevier.com/locate/compedu

An eye-tracking study of how color coding affects multimedia learning ☆

E. Ozcelik [a], T. Karakus [b], E. Kursun [c], K. Cagiltay [b,]*

[a] Department of Computer Engineering, Atilim University, Ankara 06836, Turkey
[b] Department of Computer Education and Instructional Technology, Middle East Technical University, Ankara 06531, Turkey
[c] Ataturk University, Erzurum 25240, Turkey

A R T I C L E I N F O

Article history:
Received 24 February 2008
Received in revised form 3 March 2009
Accepted 4 March 2009
Available online xxxx

Keywords:
Multimedia learning
Split attention
Color coding, Eye-tracking

A B S T R A C T

Color coding has been proposed to promote more effective learning. However, insufficient evidence currently exists to show how color coding leads to better learning. The goal of this study was to investigate the underlying cause of the color coding effect by utilizing eye movement data. Fifty-two participants studied either a color-coded or conventional format of multimedia instruction. Eye movement data were collected during the study. The results indicate that color coding increased retention and transfer performance. Enhancement of learning by color coding was due to efficiency of locating corresponding information between illustration and text. Color coding also attracted attention of learners to perceptually salient information.

1. Introduction

The famous proverb "a picture is worth a thousand words" means that humans learn better with visual aids. Multimedia provides opportunities to create effective learning environments by combining various formats such as text, static pictures, animations, videos, and audio. Prior research has suggested that learners who study expository text with illustrations perform better than students who use only text (Mayer, 1989; Mayer & Gallini, 1990).

When instruction includes both verbal and pictorial information, learning can be enhanced. However, learners are sometimes forced to split their attention between text and illustrations in multimedia instructional materials. When a learner's attention is split, information in a text passage must be kept in working memory while the related element in the illustration is sought. Learners may have difficulties when multiple sources of information, such as text and illustrations need to be integrated at the same time (Kalyuga, Chandler, & Sweller, 1999; Kaplan & Erden, 2008).

One proposed solution to this problem is to spatially integrate related text and pictures (Chandler & Sweller, 1992). This suggestion has been tested with the finding that learning is enhanced when related text and pictures are placed in close proximity within an integrated format, rather than being separated in space (Moreno & Mayer, 1999). Another technique to prevent split attention is to use the same color to associate elements in the text and in the illustration. Kalyuga et al. (1999) demonstrated that learners who studied color-coded formats obtained higher scores on a multiple-choice test than learners who studied conventional formats. Similarly, Keller, Gerjetz, Scheiter, and Garsoffky (2006) found that students had higher knowledge test scores when information visualizations were color-coded than when they were monochrome. These results show that color coding enhances learning.

1.1. The eye-tracking technique for providing cognitive processing measures

Although it is proposed that color coding of associated items in text and illustrations leads to better learning, there is currently little evidence showing how this actually occurs in the medium of multimedia learning. To show how people behave in learning environments, the eye-tracking technique can be used to provide real-time cognitive processing measures, such as the locus of eye fixation, average duration of fixations, number of fixations, and total fixation time (Rayner, 1998). Eye movement data can provide valuable information about

☆ *Author note:* Part of this study was presented at the Annual Meeting of the American Educational Research Association, Chicago, IL, April 9–13, 2007.

* Corresponding author. Tel.: +90 312 210 3683; fax: +90 312 210 7986.
E-mail address: kursat@metu.edu.tr (K. Cagiltay).

0360-1315/$ - see front matter © 2009 Published by Elsevier Ltd.
doi:10.1016/j.compedu.2009.03.002

the cognitive processes of the learner (Ballard, Hayhoe, Pook, & Rao, 1997; Just & Carpenter, 1976). The duration of a fixation is associated with the ongoing mental processes related to the present information (Henderson, 2007; Just & Carpenter, 1976). According to Just and Carpenter's (1976) eye–mind hypothesis, it is not computationally efficient to initiate new processes by directing eyes to new information if the mind is busy with processing the current information. Total fixation time (i.e. cumulative duration of fixations within a region) is considered as a sign of the amount of total cognitive processing engaged with the fixated information (Anderson, Bothell, & Douglass, 2004; Graesser, Lu, Olde, Cooper-Pye, & Whitten, 2005; Just & Carpenter, 1980; Rayner, 1998).

Few eye-tracking studies have been conducted on multimedia learning. Eye-tracking studies demonstrated that total fixation time was higher on text than on pictures, indicating that processing of illustrated materials was mainly driven by the text (Hannus & Hyönä, 1999; Hegarty, Carpenter, & Just, 1991). Similar results were reached in print advertisements (Rayner, Rotello, Stewart, Keir, & Duffy, 2001) and cartoons (Carroll, Young, & Guertin, 1992). Hannus and Hyönä (1999) found that high-ability students allocated more time on relevant information than low-ability students, suggesting that paying more attention to important information and spending more effort to integrate related elements of text and illustrations led to better learning outcomes. They also observed positive correlations between total fixation time on relevant information and recall performance.

The effects of color coding on the integration of text and illustrations were investigated in an eye-tracking study by Folker, Ritter, and Sichelschmidt (2005). They demonstrated that the number of fixations was fewer and the total fixation time on the illustration was shorter in the color-coded material than in the standard material. Folker and her colleagues argued that color coding facilitated the processing of the multimedia material. However, they failed to find a significant difference between the two groups in an applied knowledge test. Their eye-tracking measures (i.e. number of fixations and the total fixation time on illustrations) could be contaminated by unequal study time between groups. Indeed, study time of the color-coded group was significantly shorter than the control group in their experiment.

1.2. Hypotheses on underlying causes of the effect of color coding

Although several studies have suggested that color coding enhances learning, the underlying cause of this effect has not been demonstrated explicitly, yet. Two different hypotheses may explain the color coding effect. First, minimizing unnecessary search processes to relate illustrated and verbal information may be responsible for better performance in color-coded materials (Kalyuga et al., 1999). It may be easier for participants to select relevant information from illustrations or text when their attention is guided and, as a result, they can better integrate corresponding illustrated and verbal information into a coherent mental representation (Mayer, 1997, 2001). Jeung, Chandler, and Sweller (1997) demonstrated that electronic flashing that signaled visual referents of spoken text facilitated learning when it was hard for participants to find necessary information. If this proposal is correct, then participants should find associated elements between text and illustration faster in the color-coded format than in the conventional format.

Presenting stimuli close together in time causes formation of associations between these stimuli (Kahana, 1996; Raaijmakers & Shiffrin, 1981). When participants spend more time on processing two items, their association is increased (Kahana, 1996). Participants have a tendency to recall items from a nearby serial position (Kahana, 1996). Intrusion errors (i.e. recall of incorrect items) are more likely to come from adjacent list pairs in a paired associate task (Davis, Geller, Rizzuto, & Kahana, 2008). Inter-response time (i.e. time that elapses between two responses) in free recall is short when items have close study list positions (Murdock & Okada, 1970). These findings suggest when learners locate relevant information from multimedia in a short time, they may develop associations between verbal and pictorial information. This may result in more successful learning. In a meta-analysis of the temporal contiguity effect, Ginns (2006) showed that materials were more effective when mutually referring information were presented close to each other temporarily. For instance, Baggett (1984) suggested that temporal overlap of visual and verbal information facilitated the formation of dual-media associations.

The second hypothesis is that color coding effect results from guiding of attention by salient information that is relevant for the task of the learner. Attention of learners is guided by only low-level visual features (e.g. color) in a bottom-up manner according to a purely bottom-up saliency-based account of visual attention (e.g. Itti & Koch, 2000). Underwood and his colleagues (Underwood & Foulsham, 2006; Underwood, Foulsham, Van Loon, Humphreys, & Bloyce, 2006) showed that perceptually salient information captured attention in a task that required general inspection of whole pictures for a later memory test. However, attention was not drawn to visual salient objects in a search task that required top-down processing. These findings indicated that visual saliency could be overridden by task-related processes. Similarly, Yantis and Egeth (1999) demonstrated that salient features did not attract attention automatically if salient features were not useful for the current task. In sum, these results suggested that in addition to stimulus-driven salience, top-down task requirements also Q2 had an influence on guiding visual attention (Navalpakkam & Itti, 2005). In the current experiment, participants needed to locate corresponding information between the text and the illustration. Using the same color to associate elements in the text and in the illustration might facilitate "attending to [target] locations that are salient and relevant" (Navalpakkam & Itti, 2005, p. 212). The attentional guidance hypothesis expects that participants will pay more attention on the colored elements that are salient and relevant for the current task of the learner, and therefore spend more time on them. If this hypothesis is correct, then total fixation time of colored information, which is an indicator of amount of cognitive processing devoted to colored information, should be higher in the color-coded format compared to the conventional format.

2. Purposes of the study

The goals of this experiment were to investigate the effects of color coding on learning outcomes and to examine why learners performed better when using color-coded material as opposed to monochrome material. If inefficiency of locating corresponding information between text and illustration is responsible for impaired learning, then it is be expected that participants will find associated elements between text and illustration faster in a color-coded format as opposed to a conventional format. If the underlying cause of color coding is the attraction of attention to salient information, then it is expected that total fixation time on colored information will be higher in a color-coded material than in a conventional material.

Please cite this article in press as: Ozcelik, E., et al. An eye-tracking study of how color coding affects multimedia learning. *Computers & Education* (2009), doi:10.1016/j.compedu.2009.03.002

3. Method

3.1. Participants

The participants were 52 undergraduate students (15 female and 37 male), all between 19 and 27 years of age ($M = 19.45$, $SD = 2.00$). All were native Turkish speakers. They voluntarily took part in the experiment for extra course credits. No participant reported color-blindness. Twenty-six randomly selected participants were assigned to the color-coded group, and the rest studied the conventional format.

A one-way analysis of variance (ANOVA) was conducted on prior knowledge to find whether the color-coded and the conventional group differed in their prior knowledge. The results show that the color-coded group ($M = 12.54$, $SD = 5.64$) and the conventional group ($M = 12.65$, $SD = 5.38$) were not significantly different in terms of prior knowledge, $F(1,50) = .01$, $p = .94$. There was also no significant difference in study time between the groups, $F(1,50) = .01$, $p = .92$, indicating that the participants spent similar amounts of time studying the color-coded material ($M = 121$ s, $SD = 44$ s) and the conventional material ($M = 121$ s, $SD = 40$ s).

3.2. Materials

All materials were presented in Turkish. The instructional material focused on chemical synapses. The material included instructions about how signaling occurs in the nervous system, and described the roles of action potentials, calcium ions, sodium ions, neurotransmitters, and voltage-dependent ion channels. This topic was selected, because it was expected that the participants would have little or no prior knowledge of the subject, which might affect their performance. The instructional material consisted of two computer screens. The first screen presented background information about chemical synapses and was presented to all of the participants, whereas the second screen was offered in two different formats. The conventional format (Fig. 1) included an illustration and text which were physically separated from each other. The color-coded format (Fig. 2) was identical to the conventional format except that all of the field-specific terminology (e.g. neurotransmitter) was presented in the same color within the text and in the illustration. All of the materials were presented visually on the computer screens. Auditory information was not supplied. The instructional materials were self-paced.

The subjects' prior knowledge concerning synapses and nerves was assessed by a questionnaire which consisted of five statements, such as, "I know the role of neurotransmitters in signal transmission", "I know the difference between axons and dendrites", and "I know which ions are required for the transmission of signals". The subjects were asked to mark a five-point scale ranging from very little (associated

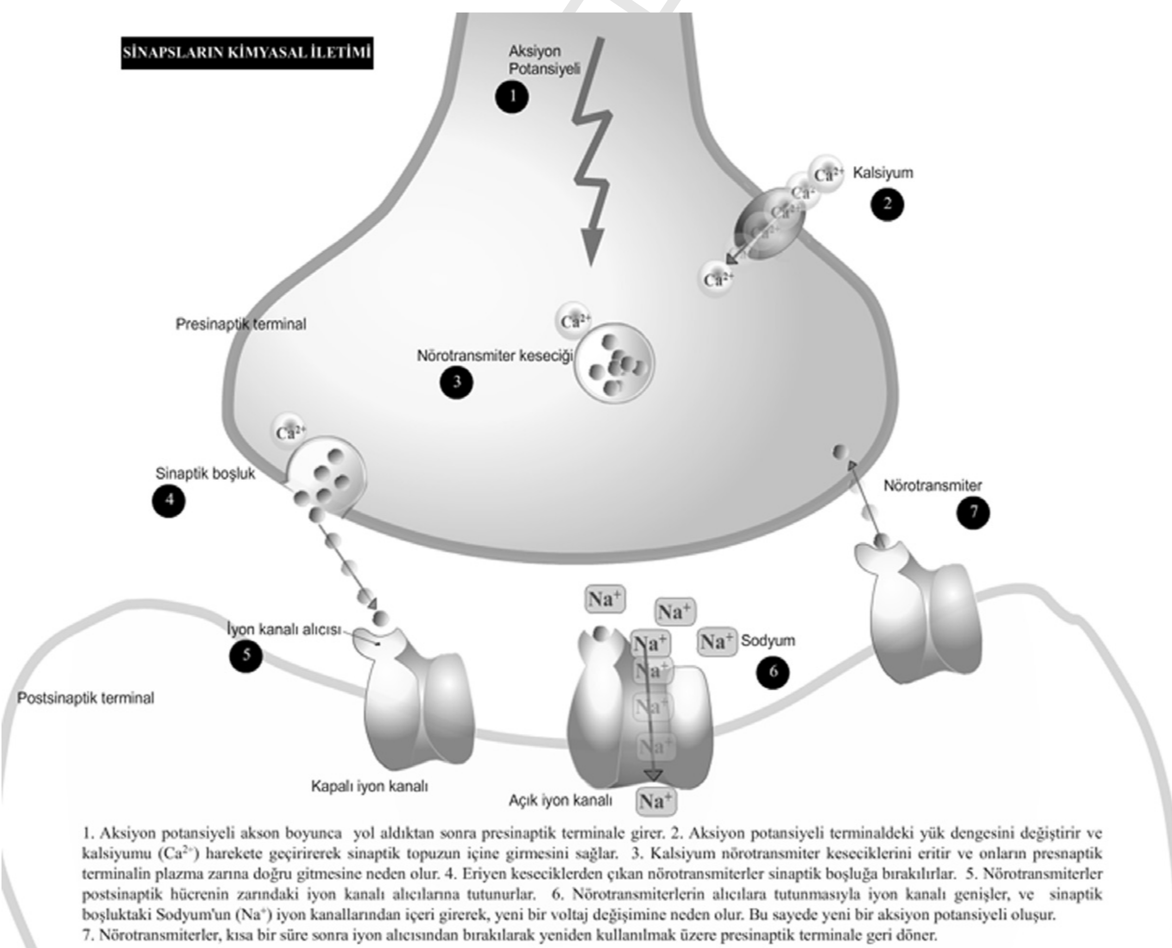

Fig. 1. The conventional format.

Please cite this article in press as: Ozcelik, E., et al. An eye-tracking study of how color coding affects multimedia learning. *Computers & Education* (2009), doi:10.1016/j.compedu.2009.03.002

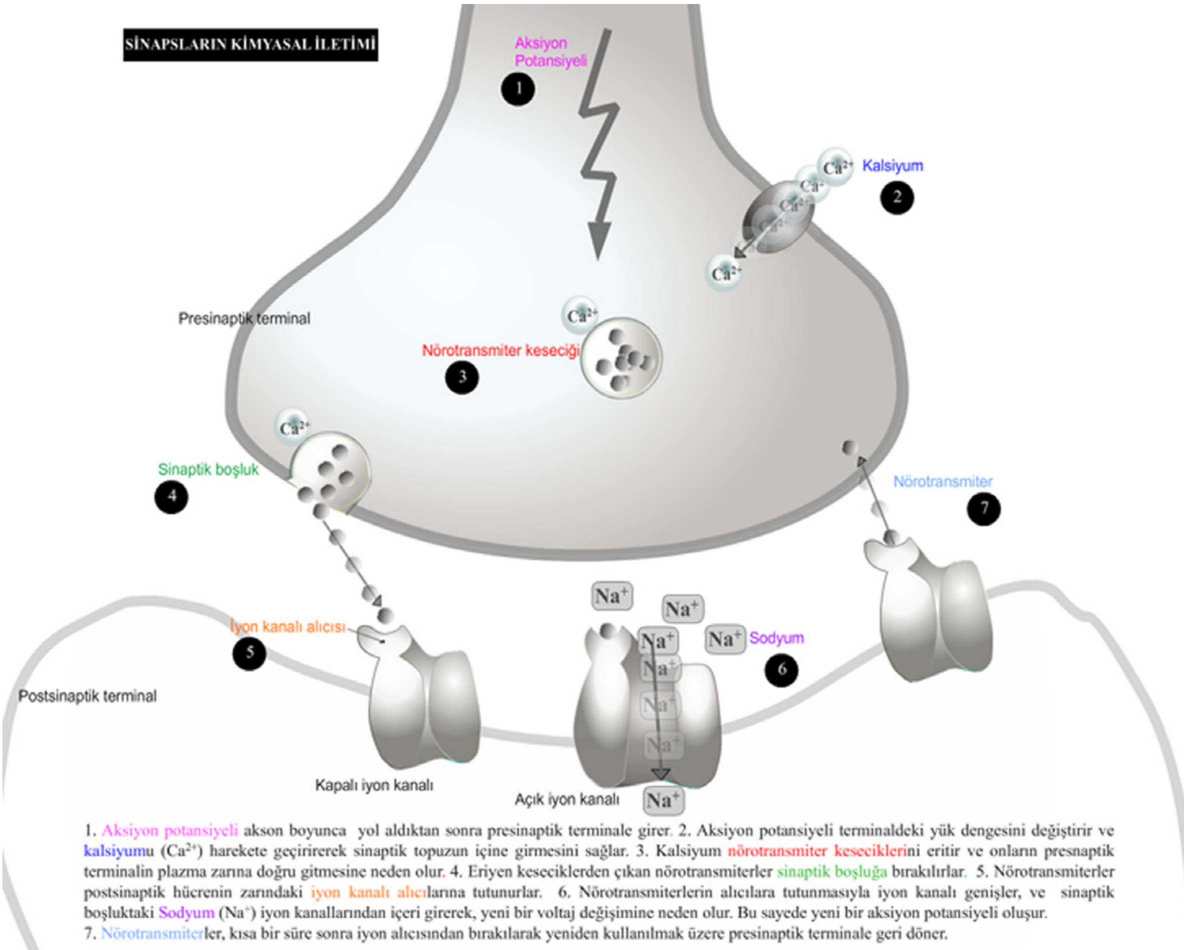

Fig. 2. The color-coded format.

with score 1) to very much (associated with score 5) for each of the statements. A similar questionnaire on a different topic was also used by Moreno and Mayer (1999).

The perceived difficulty of the learning materials was measured by a subjective rating scale (Kalyuga et al., 1999). This was a seven-point Likert-type question in which the participants were requested to rate how easy or difficult it was for them to understand the instructions. They were expected to select options from extremely easy (associated with score 1) to extremely difficult (associated with score 7).

The retention test includes seven multiple-choice questions. Each question consists of a question stem and five options. One of the options is the correct response. Each question assesses whether a fact that can be directly found in the text has been remembered or not. An example of a question in the retention test is, “What is the role of calcium in signaling between nerve cells?” and the options are: “(a) to fuse neurotransmitter vesicles”, “(b) to create action potential”, “(c) to open ion channels”, “(d) to close ion channels”, and “(e) to destroy neurotransmitters”. The matching test includes the same illustration in the study phase without the labels of the elements. The participants were asked to find and circle five provided elements in the illustration and write a corresponding number of the element (e.g. two for neurotransmitter) next to the circle. The transfer test consists of five open-ended questions examining to what extent learners apply the presented material to novel problems that cannot be directly addressed in the material. An example of a question in the transfer test is, “It is known that drugs block ion channels. Considering this information, which processes in a synapse are responsible for weakening of senses when an individual takes drugs?”.

3.3. Procedure

The participants were tested individually at the HCI laboratory. First, the participants were given the prior knowledge test and a questionnaire about their demographic information. Next, every subject underwent an automatic eye-tracking calibration. Then, participants were asked to study the materials and they were informed that they would be given tests after the study session to assess their learning. Each group of subjects was given either color-coded or conventional materials. The participants’ eye-tracking data were collected by Tobii 1750 EyeTracker while they were studying the instructional materials. The duration of eye fixations, and the number of fixations data were obtained with the aid of Tobii’s Clearview software. When each subject finished studying the instructional materials, a perceived difficulty of the learning materials question, a retention test, a transfer test, and a matching test were administered. There were no time limitations, either for studying the instructional content or for answering the tests.

Please cite this article in press as: Ozcelik, E., et al. An eye-tracking study of how color coding affects multimedia learning. *Computers & Education* (2009), doi:10.1016/j.compedu.2009.03.002

3.4. Scoring

The rater who assessed all the tests was blind to the group of the answer sheet to eliminate rater bias. In order to minimize measurement errors that can arise from rater judgment in scoring responses to open-ended questions (Penny, Johnson, & Gordon, 2000), a four-point scale rubric was developed to evaluate the transfer test. A rater scored the transfer tests of all subjects and an additional two raters independently scored transfer tests of six randomly selected participants. Inter-rater reliability, which shows the degree of agreement among the raters, was examined. The intra-class correlation coefficient showed an agreement of .89 among the three raters. Correct answers on the retention and the matching tests yielded 1 point, so the maximum points that could be obtained on these two tests were 7 and 5, respectively. Afterwards, the scores of all the tests were converted to percentages.

4. Results

4.1. Learning outcomes

The ANOVA revealed an effect on the retention performance of the participants with a small effect size, $F(1,50) = 4.34$, $MSE = 2.35$, $p = .04$, $\eta^2 = .08$. The participants who studied the color-coded material ($M = 82.42$, $SD = 17.29$) performed better than the subjects who studied the conventional format ($M = 69.78$, $SD = 25.66$). The matching scores of the two groups were not statistically significant, $F(1,50) = .07$, $p = .79$, indicating that the color-coded group ($M = 70.77$, $SD = 33.57$) and the conventional group ($M = 73.08$, $SD = 28.25$) performed equally well in the matching test. On the other hand, the effect of the material format on transfer performance was found to be significant with a medium effect size, $F(1,50) = 10.27$, $MSE = 12.90$ $p = .002$, $\eta^2 = .17$. Accordingly, those subjects who received the color-coded material ($M = 63.85$, $SD = 19.66$) had higher scores on the transfer test than the learners who were given the conventional material ($M = 47.88$, $SD = 16.07$). There was no significant difference in the perceived difficulty of the learning materials between the color-coded group ($M = 3.68$, $SD = 1.49$) and the conventional group ($M = 3.65$, $SD = 1.49$), $F(1,46) = .04$, $p = .95$. It should also be noted that 4 subjects (1 from the color-coded group and 3 from the conventional group) did not provide answers to the difficulty ratings.

4.2. Eye-tracking measures

Performance on locating mutually referring information between the text and the illustration was determined by using eye movements data. The analysis of the eye movements data showed that 8 subjects (4 from the color-coded group and 4 from the conventional group) studied the material without shifting between the illustration and the text. Thus, they did not attempt to find a technical term (e.g. neurotransmitter) between the text and the illustration. These participants were excluded in the analysis. This resulted in the inclusion of 44 participants in the current analysis. Rectangular region of interests (ROIs) were drawn manually around each technical term in the text and in the illustration by Tobii's Clearview software. The eye tracker provided the actual time when a fixation landed on a specific ROI on the screen. The efficiency of search performance was measured by calculating how quickly subjects found the corresponding technical terms both from the text to the illustration and from the illustration to text. In order to assess the efficiency of search performance, the average time elapsed starting from the first fixation landing on a technical term in the text (or in the illustration) until the corresponding term is first fixated in the illustration (or in the text) was calculated for all attempts of searching corresponding information of a participant. The results suggest that participants spent less time locating the corresponding terms when the studied material was color-coded ($M = 4.16$ s, $SD = 2.72$ s) than when the studied material was conventional ($M = 9.85$ s, $SD = 4.53$ s), $t(42) = -5.05$, $p < .001$.

A 2 (material format: color-coded, conventional) × 2 (region: illustration, text) mixed factorial ANOVA was performed on average fixation durations. Material format was a between-subjects variable whereas region was a within-subjects variable. The effect of material format was significant, $F(1,50) = 6.61$, $MSE = 4813421.9$, $p = .01$, $\eta^2 = .12$. Average fixation duration was longer for the subjects using the color-coded material ($M = 228$ ms, $SD = 27$ ms) than for those using the conventional material ($M = 203$ ms, $SD = 42$ ms). There was no significant difference between average fixation duration between the text and illustration regions, $F(1,50) = 1.32$, $p = .26$. The interaction between material format and region was not significant, $F(1,50) = 1.62$, $p = .21$.

If long average fixation duration is associated with extensive processing of the instructional material, then transfer scores indicating deep processing of the learning material should be predicted by average fixation durations in a linear regression analysis. On the other hand, if long average fixation duration is related to the difficulty of the instructional material, then subjective difficulty ratings should be predicted by average fixation durations. In order to examine these possibilities, two separate linear regression analyses were conducted. The results of the regression analyses indicate that average fixation durations accounted for 22% of the variance in the transfer scores, $F(1,50) = 14.05$, $p < .001$, with a positive correlation coefficient ($r = .47$, $p < .001$). Longer average fixation durations were associated with better transfer performance. However, this eye-tracking measure was not a significant predictor of subjective difficulty ratings, $F(1,46) = .27$, $p < . 61$.

Like average fixation duration, material format also accounted for a significant amount of variance in the transfer measure, $R^2 = .16$, $F(1,50) = 9.94$, $p = .003$. In order to examine whether average fixation duration predicted transfer scores after controlling for the effects of the material format, a sequential regression was performed. Material format and average fixation duration were included in the model as predictors in the first and second steps, respectively. The results show that average fixation duration accounted for a significant proportion of transfer performance after the effects of the material format were statistically eliminated, R^2 change $= .11$, $F(1,49) = 7.32$, $p = .009$. This suggests that participants who made longer average fixation duration tended to have higher transfer scores after controlling for the effects of the material format. The linear combination of these two variables was significantly related to the transfer scores, $R^2 = .27$, $F(2,49) = 9.25$, $p < .001$, indicating that the linear combination of material format and average fixation duration predicted 27% of the variance in transfer performance.

The attentional guidance hypothesis expects that the participants in the color-coded group compared to the participants in the conventional group spent more time on the technical terms, which were the only colored elements in the material. The sum of total fixation durations on ROIs of the terminological items on the text was computed for all participants ($n = 52$). This yielded a total fixation time of the terminological items on the text. The same procedure was applied on the terms on the illustration to find the total fixation time of the terminological items on the illustration. Grand total fixation time of the terms is the sum of (i) total fixation time of the terms on the illustration and (ii) total fixation time of the terms on the text.

To test the predictions of the attentional guidance hypothesis, a 2 (material format: color-coded, conventional) × 2 (region: illustration, text) mixed factorial ANOVA was performed on total fixation time of the technical terms. Neither the effect of material format, $F(1,50) = 2.03$, $p = .16$, nor the effect of region, $F(1,50) = 2.04$, $p = .16$, was significant. The interaction between material format and region was non-significant as well, $F(1,50) = .50$, $p = .83$. These results show that grand total fixation time of the terms was not statistically different between the color-coded format ($M = 13.19$ s, $SD = 6.92$ s) and the conventional format ($M = 10.54$ s, $SD = 6.48$ s). Taken together, these results indicate that participants did not pay more attention to the terms in the color-coded format than the ones in the conventional format. However, non-significant results may have been obtained due to the high variance in this eye-tracking measure since the learners were given unlimited time to study the materials.

To decrease variability in eye movement measures, total fixation time of the technical terms was analyzed by separating the time line into time windows of 10 s. This enabled us to explore fixation patterns as a function of time. The minimum study time of the participants was 49 s. In order to include the data of all participants in the current analysis, only the fixations during the first 40 s after the onset of the study material were taken into account. A 2 (material format: color-coded, conventional) × 2 (region: illustration, text) × 4 (time window: 0–10 s, 10–20 s, 20–30 s, 30–40 s) mixed factorial ANOVA was performed on total fixation time of the terms. While material format was a between-subjects variable, region and time window were within-subjects variables. The effect of region was significant, $F(1,50) = 10.66$, $MSE = 18267144.1$, $p = .002$, $\eta^2 = .18$, indicating that total fixation time of the terms for the first 40 s was higher on the illustration ($M = 3470.44$ ms, $SD = 2807.11$ ms) compared to the text region of the material ($M = 1794.04$ ms, $SD = 1643.44$ ms). Most importantly, the effect of material format was significant, $F(1,50) = 4.97$, $MSE = 180145925.5$, $p = .03$, $\eta^2 = .09$, indicating that grand total fixation time of the terms on the whole material – both on the text and on the illustration – for the first 40 s was higher when the studied format was color-coded ($M = 6086.81$ ms, $SD = 2615.68$ ms) than when the studied format was conventional ($M = 4442.15$ ms, $SD = 2701.20$ ms). The effect of time window was also significant, $F(3, 150) = 8.08$, $MSE = 3182413.6$, $p < .001$, $\eta^2 = .14$. Six paired-samples t tests were performed to follow-up pairwise comparisons for the main effect of time window. In order to control Type I error across the six tests, Holm's sequential Bonferroni procedure was applied. Differences in grand total fixation time of the terms were significant between 0–10 s and 10–20 s, $t(51) = 4.53$, $p < .001$, between 0–10 s and 20–30 s, $t(51) = 4.10$, $p < .001$, and between 0–10 s and 30–40 s, $t(51) = 3.47$, $p = .001$. All the other differences in grand total fixation time of the technical terms were non-significant (all $ps > .50$). These results indicate that grand total fixation time of the terms was higher during 0–10 s ($M = 1834.73$ ms, $SD = 1158.23$ ms) than it was during 10–20 s ($M = 1182.25$ ms, $SD = 961.25$ ms), 20–30 s ($M = 1067.58$ ms, $SD = 996.56$ ms), and 30–40 s ($M = 1179.92$ ms, $SD = 997.52$ ms). The only significant interaction was between time window and region, $F(3,48) = 23.42$, $MSE = 14638784$, $p < .001$, $\eta^2 = .32$. As it can be seen in Fig. 3, in the early time windows (i.e. 0–10 and 10–20 s), total fixation time of the terms was longer on the illustration than on the text, $t(51) = 6.91$, $p < .001$. However, this was not the case in the late time windows (i.e. 20–30 and 30–40 s), $t(51) = 1.07$, $p = .29$. This result indicates that participants started studying the material by looking at the illustration and afterwards they read the text. All the other interactions were non-significant (all $ps > .56$).

When participants found necessary information faster, they might spend more time on the key terms that were essential for understanding the subject. In order to examine the relationship between the efficiency of locating relevant information and the grand total fixation time of the terms, Pearson correlation coefficients were calculated in a bivariate correlation analysis. The results show that there was

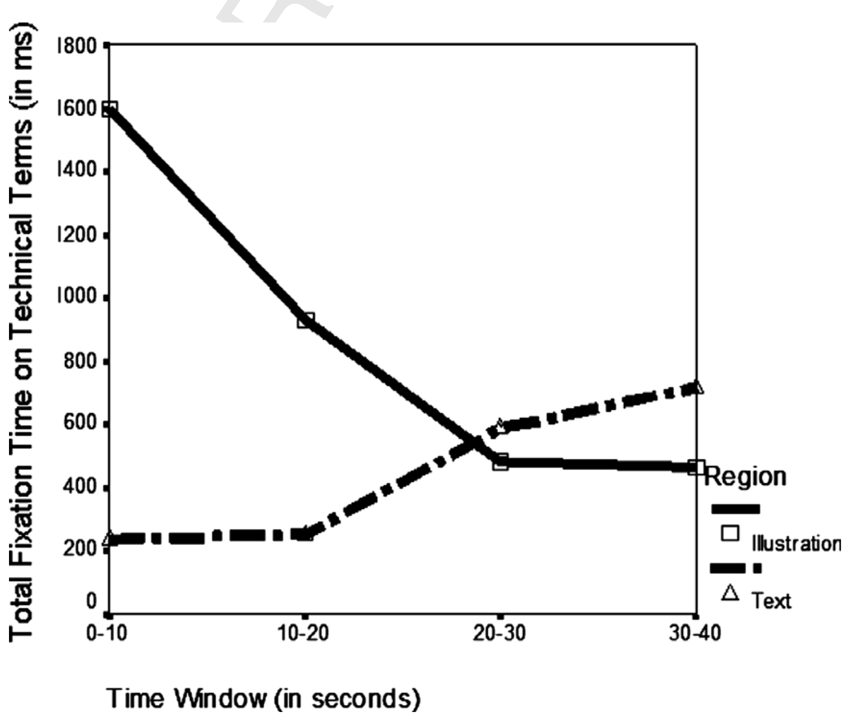

Fig. 3. Total fixation time of the technical terms as a function of time and region.

Please cite this article in press as: Ozcelik, E., et al. An eye-tracking study of how color coding affects multimedia learning. *Computers & Education* (2009), doi:10.1016/j.compedu.2009.03.002

a significant negative correlation between these two variables ($r = -.41$, $p = .006$). This suggests that spending less time finding the corresponding information between the text and the illustration was associated with processing these terms for longer periods.

Efficiency of locating mutually referring information and/or total time spent on the terms may be associated with transfer performance of the participants. In order to examine these possibilities, two separate bivariate correlation analyses were performed. The correlation was statistically significant between efficiency of search performance and transfer ($r = -.33$, $p = .028$), but not between grand total fixation time of the terms and transfer ($r = .08$, $p = .56$). These results suggest that search performance and transfer were negatively related, indicating that shorter search time was associated with better transfer scores, whereas grand total fixation time of the terms and transfer were not related.

If the learners who found necessary information faster tended to spend more time on the terms, then it is expected that the first readings of the technical terms by the color-coded group took place later than the ones by the conventional group. To examine this, a 2 (material format: color-coded, conventional) × 2 (region: illustration, text) mixed factorial ANOVA was performed on average of first fixation times of the terms. The analysis revealed a significant effect of region, $F(1,50) = 22.45$, $MSE = 8{,}032{,}086{,}326$, $p < .001$, $\eta^2 = .32$. Consistent with the previous finding that participants first looked at the illustration and then read the text, first readings of the terms by the participant on the illustration ($M = 33.70$ s, $SD = 32.71$ s) took place earlier than the ones on the text ($M = 50.97$ s, $SD = 35.30$ s). There was no significant effect of material format, $F(1,50) = 1.87$, $p = .18$. A significant interaction between material format and region was observed, $F(1,50) = 10.18$, $MSE = 3{,}641{,}612{,}934$, $p = .003$, $\eta^2 = .18$. A planned comparison was run for contrasting the averages of the first fixation times of the technical terms on the text between conventional and color-coded formats. The independent-samples t test showed that the first readings of the terms on the text took place later in the color-coded materials ($M = 62.90$ s, $SD = 45.03$ s) compared to conventional materials ($M = 39.04$ s, $SD = 14.84$ s), $t(50) = -2.52$, $p = .015$. Another planned comparison for contrasting the averages of first fixation times of the technical terms on the illustration between conventional and color-coded formats revealed that the first reading of key terms on the illustration was not different between color-coded ($M = 32.90$ s, $SD = 43.68$ s) and conventional materials ($M = 34.48$ s, $SD = 17.51$ s), $t(50) = .17$, $p = .87$. In sum, these results suggest that first readings of the key terms on the text took place later for the participants who were given the color-coded material than for the participants who were given the conventional material.

5. Discussion

The main goal of this study was to investigate the effects of color coding on learning by examining learning outcomes and eye movement measures. The results of this experiment indicate that the subjects who studied the material in the color-coded format had higher retention and transfer scores than the subjects who studied the material in the conventional format. Our findings were consistent with the color coding effect (Folker et al., 2005; Kalyuga et al., 1999; Keller et al., 2006).

Participants spent less time finding corresponding items between the illustration and the text in the color-coded material than in the conventional material. These findings provide evidence that color coding facilitated the efficiency of finding corresponding elements between text and illustrations (Kalyuga et al., 1999). This is because color coding alerts learners to where they may detect related items (Greene & Rayner, 2001). Research on visual cognition supports our findings that color coding makes visual search tasks easier, because an object that is unique in color stands out from its surroundings (Green & Anderson, 1956; Treisman & Gelade, 1980).

Our data demonstrate that color coding enabled learners to locate related information quickly between the text and the illustration. This might facilitate formation of associations between verbal and pictorial representations since items processed close in time are associated with each other (Kahana, 1996; Raaijmakers & Shiffrin, 1981). As in our conventional materials, learning might be impaired when there is a delay between processing of corresponding words and pictures (Baggett, 1984). The current results are in line with the meta-analysis of the temporal contiguity effect in multimedia learning, which demonstrates improvement of learning outcomes as a result of temporal proximity of related information (Ginns, 2006).

One of the findings of this study is that average fixation duration was longer when the material was coded by colors. Longer average fixation duration is regarded as an indication of deeper processing (Rayner, 1998), because fixation duration is thought to be determined by current cognitive processes. When it was easy for learners to find relevant information and integrate verbal and pictorial information with each other, learners could more easily engage in deeper processing required for meaningful learning (Mayer, 2003; Mayer & Chandler, 2001; Seufert, 2003). The results of the linear regression analysis showed that the average fixation duration could account for significant amount of variance in the transfer scores. The positive correlation coefficient between these measures indicated that there was a significant relationship between transfer scores and average fixation durations. On the other hand, linear regression analyses revealed that average fixation duration was not a significant predictor of subjective difficulty ratings of the learning material. Taken together, these results suggested that long fixation durations were associated with extensive processing of the instructional material rather than the difficulty of the material. Like average fixation duration, material format was a significant predictor of transfer performance. Average fixation duration added to the prediction of transfer performance after the effects of the material format were controlled for. These findings suggested that longer average fixation duration was associated with better transfer performance even if the effects of the format of the material were eliminated.

Grand total fixation time on the technical terms was not statistically different between color-coded and conventional formats. Even though the statistics were not significant, there was a trend towards longer total fixation time on the technical terms for the color-coded group than the conventional group. The non-significant results may be due to the high variance in these eye-tracking measures in the self-paced learning environment. In order to overcome the problem of high variance, total fixation times of the technical terms were analyzed by separating the time line into time windows of 10 s. Grand total fixation time of the technical terms for the first 40 s was higher in the color-coded format than in the conventional format. Thus, the results support the attentional guidance hypothesis. Accordingly, the participants who studied the color-coded instruction paid more attention to salient colored information than the participants who studied the conventional instruction when fixations during the first 40 s were included in the analysis.

The efficiency of finding associated items between the illustration and the text was negatively related with grand total fixation time of the technical terms. Besides, transfer performance was negatively related with the efficiency of search processes, but not with grand total fixation time of the terms. Taken together, these findings indicate that participants who found related information faster tended to spend

Please cite this article in press as: Ozcelik, E., et al. An eye-tracking study of how color coding affects multimedia learning. *Computers & Education* (2009), doi:10.1016/j.compedu.2009.03.002

more time on the technical terms and to have better performance in the transfer test, but there was no relationship between time spent on the technical terms and transfer performance. It was also observed that learners first inspected the illustration and second they read the text. Since learners located corresponding information faster by the help of color coding and these quick subjects in search processes tended to process the terms for longer periods, first fixations of the technical terms on the text took place later in the color-coded group compared to the conventional group.

Our results explicitly demonstrated that the efficiency of locating mutually referring information was one of the causes of the color coding effect. An alternative hypothesis proposes that enhancement of learning by color coding was due to the fact that the material that would be tested was signaled in the color-coded condition. This hypothesis proposes that the learners in the color-coded group outperformed those in the conventional group, because color coding highlighted sections that would be assessed in the tests. This hypothesis seems unlikely to be correct for three reasons. First, all of the terminological words were signaled in the color-coded format. Second, requiring learners to apply their knowledge to novel (i.e. not previously presented in the materials) situations in the transfer test enables us to minimize possible confounding with material being tested. Third, there was no significant difference in the matching test, which can be considered as the most vulnerable test among the tests employed in the current study to this confound since it included matching of all the terminological terms with the elements in the illustration.

6. Conclusions

This experiment provides evidence for the effectiveness of the use of color coding in multimedia learning. When learners find corresponding items between text and illustrations, they demonstrate better retention and transfer performance. Using eye-tracking data, the present study suggests that color coding aids participants to find corresponding information in text and illustrations and to pay attention to critical information for meaningful learning. The current study also shows that locating corresponding information faster is associated with spending more time on the technical terms and with better performance in the transfer test.

The results of the present study suggest the following implications for practice. Designers should use the same color to associate elements in related text and illustrations. Attention of learners should be directed to important elements in the instructional material and unnecessary search for related information should be minimized in order to enhance learning. Learning environments should be designed considering how humans select relevant information and integrate verbal and visual representations (Mayer, 2001).

7. Uncited references

Q1 Itti and Koch (2001) and Lowe (2004).

Acknowledgment

This study was supported by TUBITAK under grant SOBAG 104K098 and METU Human Computer Interaction Research Group (http://hci.metu.edu.tr). We thank two anonymous reviewers for their constructive comments.

We thank two anonymous reviewers for their valuable and constructive comments on the manuscript.

References

Anderson, J. R., Bothell, D., & Douglass, S. (2004). Eye movements do not reflect retrieval processes: Limits of the eye–mind hypothesis. *Psychological Science, 15*, 225–231.
Ballard, D. H., Hayhoe, M. M., Pook, P. K., & Rao, R. P. N. (1997). Deictic codes for the embodiment of cognition. *Behavioral and Brain Sciences, 20*, 723–767.
Baggett, P. (1984). Role of temporal overlap of visual and auditory material in forming dual media associations. *Journal of Educational Psychology, 76*, 408–417.
Carroll, P. J., Young, R. J., & Guertin, M. S. (1992). Visual analysis of cartoons: A view from the far side. In K. Rayner (Ed.), *Eye movements and visual cognition: Scene perception and reading* (pp. 444–461). New York: Springer.
Chandler, P., & Sweller, P. (1992). The split-attention effect as a factor in the design of instruction. *The British Journal of Educational Psychology, 62*, 233–246.
Davis, O. C., Geller, A. S., Rizzuto, D. S., & Kahana, M. J. (2008). Temporal associative processes revealed by intrusions in paired-associate recall. *Psychonomic Bulletin & Review, 15*, 64–69.
Folker, S., Ritter, H., & Sichelschmidt, L. (2005). Processing and integrating multimodal material – The influence of color coding. In B. G. Bara, L. Barsalou, & M. Bucciarelli (Eds.), *Proceedings of 27th annual conference of the cognitive science society* (pp. 690–695). Mahwah, NJ: Lawrence Erlbaum.
Ginns, P. (2006). Integrating information: Meta-analyses of the spatial contiguity and temporal contiguity effects. *Learning and Instruction, 16*, 511–525.
Graesser, A. C., Lu, S., Olde, B., Cooper-Pye, E., & Whitten, S. N. (2005). Question asking and eye tracking during cognitive disequilibrium: Comprehending illustrated texts when the devices breakdown. *Memory & Cognition, 33*, 1235–1247.
Green, B. F., & Anderson, L. K. (1956). Color coding in a visual search task. *Journal of Experimental Psychology, 51*, 19–24.
Greene, H. H., & Rayner, K. (2001). Eye movements and familiarity effects in visual search. *Vision Research, 41*, 3763–3773.
Hannus, M., & Hyönä, J. (1999). Utilization of illustrations during learning of science textbook passages among low- and high-ability children. *Contemporary Educational Psychology, 24*, 95–123.
Hegarty, M., Carpenter, P. A., & Just, M. A. (1991). Diagrams in the comprehension of scientific text. In R. Barr, M. L. Kamil, P. B. Mosenthal, & P. D. Pearson (Eds.). *Handbook of reading research* (Vol. 2, pp. 641–668). New York: Longman.
Henderson, J. M. (2007). Regarding scenes. *Current Directions in Psychological Science, 16*, 219–222.
Itti, L., & Koch, C. (2000). A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision Research, 40*, 1489–1506.
Itti, L., & Koch, C. (2001). Computational modeling of visual attention. *Nature Reviews Neuroscience, 2*, 194–203.
Jeung, H., Chandler, P., & Sweller, J. (1997). The role of visual indicators in dual sensory mode instruction. *Educational Psychology, 17*, 329–343.
Just, M. A., & Carpenter, P. A. (1976). Eye fixations and cognitive processes. *Cognitive Psychology, 8*, 441–480.
Just, M. A., & Carpenter, P. A. (1980). A theory of reading: From eye fixations to comprehension. *Psychological Review, 87*, 329–354.
Kahana, M. J. (1996). Associative retrieval processes in free recall. *Memory & Cognition, 24*, 103–109.
Kalyuga, S., Chandler, P., & Sweller, J. (1999). Managing split-attention and redundancy in multimedia instruction. *Applied Cognitive Psychology, 13*, 351–371.
Kaplan, Z., & Erden, M. (2008). Instructional efficiency of integrated and separated text with animated presentations in computer-based science instruction. *Computers & Education, 51*, 660–668.
Keller, T., Gerjetz, P., Scheiter, K., & Garsoffky, B. (2006). Information visualizations for knowledge acquisition: The impact of dimensionality and color coding. *Computers in Human Behavior, 22*, 43–65.
Lowe, R. K. (2004). Interrogation of a dynamic visualization during learning. *Learning and Instruction, 14*, 257–274.
Mayer, R. E. (1989). Systematic thinking fostered by illustrations in scientific text. *Journal of Educational Psychology, 81*, 240–246.
Mayer, R. E. (1997). Multimedia learning: Are we asking the right questions? *Educational Psychologist, 31*, 1–19.

Mayer, R. E. (2001). *Multimedia learning*. Cambridge, MA: Cambridge University Press.
Mayer, R. E. (2003). The promise of multimedia learning: Using the same instructional design methods across different media. *Learning and Instruction, 13*, 125–139.
Mayer, R. E., & Chandler, P. (2001). When learning is just a clock away: Does simple user interaction foster deeper understanding of multimedia messages? *Journal of Educational Psychology, 93*, 390–397.
Mayer, R. E., & Gallini, J. K. (1990). When is an illustration worth ten thousand word? *Journal of Educational Psychology, 82*, 715–726.
Moreno, R., & Mayer, R. E. (1999). Cognitive principles of multimedia learning: The role of modality and contiguity. *Journal of Educational Psychology, 91*, 358–368.
Murdock, B. B., & Okada, R. (1970). Interresponse times in single-trial free recall. *Journal of Verbal Learning and Verbal Behavior, 86*, 263–267.
Penny, J., Johnson, R. L., & Gordon, B. (2000). Using rating augmentation to expand the scale of an analytic rubric. *Journal of Experimental Education, 68*, 269–287.
Raaijmakers, J. G. W., & Shiffrin, R. M. (1981). Search of associative memory. *Psychological Review, 88*, 93–134.
Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin, 124*, 372–422.
Rayner, K., Rotello, C. M., Stewart, A. J., Keir, J., & Duffy, S. A. (2001). Integrating text and pictorial information: Eye movements when looking at print advertisements. *Journal of Experimental Psychology: Applied, 7*, 219–226.
Seufert, T. (2003). Supporting coherence formation in learning from multiple representations. *Learning and Instruction, 13*, 227–237.
Treisman, A., & Gelade, G. (1980). A feature integration theory of attention. *Cognitive Psychology, 12*, 97–136.
Underwood, G., & Foulsham, T. (2006). Visual saliency and semantic incongruency influence eye movements when inspecting pictures. *Quarterly Journal of Experimental Psychology, 59*, 1931–1949.
Underwood, G., Foulsham, T., Van Loon, E., Humphreys, L., & Bloyce, J. (2006). Eye movements during scene inspection: A test of the saliency map hypothesis. *European Journal of Cognitive Psychology, 18*, 321–343.
Yantis, S., & Egeth, H. E. (1999). On the distinction between visual salience and stimulus-driven attentional capture. *Journal of Experimental Psychology: Human Perception and Performance, 25*, 661–676.

Please cite this article in press as: Ozcelik, E., et al. An eye-tracking study of how color coding affects multimedia learning. *Computers & Education* (2009), doi:10.1016/j.compedu.2009.03.002