Computers & Education 119 (2018) 112–128

Contents lists available at ScienceDirect

Computers & Education

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

Exploring the effect of using different levels of emotional design features in multimedia science learning

Ahmet Murat Uzun[a,*], Zahide Yıldırım[b]

[a] Department of Computer Education and Instructional Technology, Faculty of Education, Afyon Kocatepe University, 03200 Afyonkarahisar, Turkey
[b] Department of Computer Education and Instructional Technology, Faculty of Education, Middle East Technical University, 06800 Ankara, Turkey

ARTICLE INFO

Keywords:
Human-computer interface
Media in education
Multimedia/hypermedia systems
Pedagogical issues

ABSTRACT

The current study aims to examine the influence of incorporating different emotional design approaches into multimedia on 7th grade middle school students' positive emotions, mental effort investments and learning achievements (recall and transfer). To this end, four different instructional materials were created. Each material was identical, with only the extent of emotional design differing. For the Neutral Design (ND) group, none of the emotional design principles were used (monochromatic grayscale). For the Colorful Design (CD) group, attention-grabbing, bright and saturated colors were added to the material. For the Anthropomorphic Design (AD) group, expressive facial expressions (anthropomorphism of lifeless objects and expressive facial expressions of human characters) were incorporated into the material. Finally, for the Anthropomorphic Design and Sound Effects (ADSE) group, interesting sound effects were also used. The study was conducted at a middle school with a group of 106 students. Positive emotions were measured using an emWave emotion detection device. Results revealed that positive emotions generally increased as the amount of emotional design features increased. However, while students who used the Colorful Design invested more mental effort compared to students who used the Neutral Design, students who used the Anthropomorphic Design and Sound Effects (ADSE) invested less mental effort compared to students who used the Colorful Design. In addition, students who used the Colorful Design outperformed students who used the Neutral Design in terms of their recall scores. No significant difference in terms of transfer of learning scores was observed across the groups. Results were discussed with respect to different views in the literature regarding the use of emotional design in multimedia.

1. Introduction

Despite a long history in educational research, affective factors have been somewhat overlooked in research on multimedia learning (Astleitner & Wiesner, 2004; Leutner, 2014; Mayer, 2014a). This may be because traditional multimedia theory is based on a cold cognitivist approach which ignores the regulatory roles of motivation on cognition (Astleitner & Wiesner, 2004; Park, Flowerday, & Brünken, 2015). With recent theoretical developments in multimedia learning (e.g., Moreno, 2006, Cognitive Affective Theory of Learning with Media or CATLM), there is considerable concern about the use of affective design issues in multimedia. Accordingly, emotional design in multimedia learning has been offered by researchers with the expectation of improving learners' motivation to learn (Mayer & Estrella, 2014; Park, Knörzer, Plass, & Brünken, 2015; Plass, Heidig, Hayward, Homer, & Um, 2014; Um, Plass, Hayward, & Homer, 2012).

* Corresponding author.
E-mail addresses: auzun@aku.edu.tr (A.M. Uzun), zahidey@metu.edu.tr (Z. Yıldırım).

https://doi.org/10.1016/j.compedu.2018.01.002
Received 6 June 2017; Received in revised form 25 November 2017; Accepted 5 January 2018
Available online 10 January 2018
0360-1315/

Emotional design in multimedia is a relatively new approach, with studies regarding emotional design in their "infancy" (Mayer & Estrella, 2014, p. 12). Besides, most studies on emotional design compared multimedia materials that used multi-dimensional operationalization of emotional design (i.e., use of color, anthropomorphism and baby-face bias together) with neutral design. Moreover, when measuring emotion, researchers suggest that more direct measurement tools may be utilized (Dong, 2007; Mayer & Estrella, 2014). Therefore, this current study investigates the effect of emotional design on learning in a gradually increased manner rather than considering all emotional design elements together. Furthermore, instead of self-reported paper-and-pencil-based measures, the emWave emotion detection system is used, which is a more direct measure to detect learning emotions.

2. Theoretical framework

2.1. Multimedia learning

Cognitive Load Theory (CLT) and Cognitive Theory of Multimedia Learning (CTML) assumes that the effectiveness of knowledge acquisition is dependent on the optimal use of limited cognitive resources (Mayer, 2009; Paas & Sweller, 2014). The total cognitive capacity is determined by three cognitive processing demands (Park, 2010). *Extraneous cognitive processing* refers to unnecessary processing caused by poor instructional design. *Essential cognitive processing* is based on the intrinsic complexity or difficulty of the material. Finally, *generative cognitive processing* refers to "cognitive processing that contributes to learning. It is caused by challenging or motivating the learner to exert effort toward understanding the material" (Mayer & Moreno, 2010, p. 133). Based on cognitive constraints, the goals of the instruction are to minimize extraneous processing, manage essential processing and foster generative processing (Mayer, 2009). Principles about minimizing extraneous and managing essential processing are well-documented in the literature on multimedia learning. However, little is known about fostering generative processing (Mayer, 2009, 2014a, 2014b; Mayer & Estrella, 2014).

The limitation of a pure cold cognitivist approach is that the mind and thinking are considered as entirely objective, cold, mechanical and rational (Dai & Sternberg, 2004). One can argue that similar problems exist for multimedia learning. Previously, the use of affective factors in multimedia learning was discouraged, since these elements were seen as a source for extraneous cognitive load. However, the potential motivational benefits of affective design elements should not be dismissed, although they can interfere with learning (Park et al., 2015). Therefore, the traditional cold cognition approach should be reframed by adopting a more integrative approach to multimedia learning that considers non-cognitive as well as cognitive parameters (Astleitner & Wiesner, 2004; Park et al., 2015). CATLM (Moreno, 2006) is one prominent theory that combines both cognitive and affective functioning, which extends Mayer's CTML by adding new assumptions such as (a) *affective mediation*, (b) *metacognitive mediation* and (c) *individual differences*. The current study is mainly concerned with the *affective mediation* assumption which argues that "motivational factors mediate learning by increasing or decreasing cognitive engagement" (Moreno, 2006, p. 151).

2.2. Emotions and cognitive processing

Based on previous studies, Park et al. (2015) defined emotions as "individuals' judgments about the world that are evoked as a reaction to and an interaction with certain stimuli" (p. 31). The current study adopted this definition in line with its purpose. Moreover, similar to previous studies on emotional design, the current study does not focus on specific types of emotions, but on the valence (positive or negative) dimension of emotion.

The effect of emotion on cognition could be addressed under two competing hypotheses: suppression and facilitation (Oaksford, Morris, Grainger, & Williams, 1996). According to the suppression hypothesis, memory resources are limited and non-cognitive elements such as emotion can be seen as sources for extraneous cognitive load that can interfere with learning (Astleitner & Wiesner, 2004; Park et al., 2015). Emotional states, either positive or negative, may lead to task-irrelevant thoughts that impair cognitive performance (Seibert & Ellis, 1991). Positive mood can have deleterious effects on cognition when the task requires executive control (Oaksford et al., 1996). On the other hand, facilitation hypothesis postulates that positive emotions facilitate cognitive processing (Fredrickson, 1998) by broadening the scope of attention (Fredrickson & Branigan, 2005), improving creative problem solving (Isen, Daubman, & Nowicki, 1987), facilitating memory encoding and retrieval of affectively valent information (Nasby & Yando, 1982). Besides, positive emotions foster intrinsic and extrinsic motivation (Isen & Reeve, 2005). Positive activating emotions such as enjoyment of learning can increase student interest and thereby their motivation to learn (Pekrun, Frenzel, Goetz, & Perry, 2007).

2.3. Emotional design for multimedia learning

Recent research has shown that multimedia materials could be designed in a way to induce positive emotions (Chen & Wang, 2011; Plass et al., 2014; Um et al., 2012). One approach to inducing positive emotions is emotional design; which is "trying to make the core elements in a lesson more emotionally appealing through giving them human-like features (for example, symmetrical faces with facial expressions) and rendering them in enjoyable colors" (Clark & Mayer, 2016, p. 224).

In order to spice-up an otherwise boring lesson, previous research mainly concentrated on the concept of seductive details, which are defined as interesting, entertaining but irrelevant information that are not directly related to the learning objective (Clark & Lyons, 2010; Mayer, 2009). Studies have provided evidence that interesting design elements do not always result in enhanced learning, especially when only tangentially related to the learning objective. Consequently, due to capacity limitation of the working memory, using seductive details in multimedia has not been encouraged (Harp & Mayer, 1997, 1998). On the other hand, emotional design in multimedia suggests a new way of incorporating motivation into multimedia. Unlike seductive details, emotional design does not add extraneous information (e.g., interesting but irrelevant details) to the material; rather it manipulates the intrinsic features of design elements (i.e., interesting and relevant design features) (Heidig, Müller, & Reichelt, 2015).

Previous research on emotional design in multimedia employed various design techniques such as appealing color combinations (Mayer & Estrella, 2014; Plass et al., 2014; Um et al., 2012), anthropomorphism (attributing human characteristics to non-human entities) (Mayer & Estrella, 2014; Park et al., 2015; Plass et al., 2014; Um et al., 2012), baby-face bias, rounded objects and figures (Plass et al., 2014; Um et al., 2012), and interesting sounds (Moreno & Mayer, 2000).

Research has shown that colors affect emotions (e.g., Boyatzis & Varghese, 1994; Kaya & Epps, 1998). In general, cool colors such as blue, green and purple are regarded as "restful" and "quiet", while warm colors such as red, yellow and orange are considered "active" and "stimulating" (Ballast, 2002; as cited in Kaya & Epps, 1998, p. 397). Children primarily prefer bright colors to dark colors (Lohr, 2007); attributing positive emotional meanings to bright colors and negative emotional meanings to dark colors (Boyatzis & Varghese, 1994). The use of highly-saturated colors makes images perceived more easily compared to the use of unsaturated colors. In addition, in order to make certain elements more noticeable, instructional designers could consider the use of warm colors (Lohr, 2007).

Another design technique to elicit positive emotion is anthropomorphism or personification, which includes incorporating humanoid features into lifeless objects (Clark & Mayer, 2016). Although animistic thought and "mechanistic, orthodox view of science" conflict with each other, anthropomorphism and animism can be beneficial to promote students' science learning. They are simply seen as a pedagogic strategy used to help teachers explain abstract concepts that are compatible with students' understandings of the concrete physical world (Kallery & Psillos, 2004; Watts & Bentley, 1991, 1994).

Another design approach is baby-face bias of cartoon-like characters, which refers to the tendency of perceiving things with baby-faced features (e.g., large eyes, small noses, and high foreheads) as more likeable and attention-grabbing (Lidwell, Holden, & Butler, 2003). Cartoon-like characters initially draw children's attention and interest to the material and provide enjoyment by stimulating the imagination (Dalacosta, Kamariotaki-Paparrigopoulou, Palyvos, & Spyrellis, 2009).

Adding interesting sounds and music is another way of making a multimedia more interesting. In previous research, interesting sounds or music was used in the form of background music or environmental sounds (Moreno & Mayer, 2000). Research showed that seductive details effect was found when background music was incorporated into the narrated multimedia animations. On the other hand, seductive details effect was not observed, when context dependent sounds were used (Thalheimer, 2004).

The first known study on emotional design was conducted by Um et al., 2012. Emotional design included anthropomorphic images, attention-grabbing color combinations and baby-like elements; whereas, neutral design involved only grayscale coloring. Um et al., 2012 discovered that emotional design induced positive emotions among users. The internally induced emotions, by means of emotional design, enhanced both comprehension and transfer domains of learning; whereas, externally induced emotions only promoted transfer but not comprehension. Results also indicated that externally induced positive emotions resulted in higher levels of mental effort investment than externally induced neutral emotions; however, there was no main mental effort effect for emotional design. Plass et al. (2014) replicated the study of Um et al., 2012 and found that emotional design helped maintain and increase positive mood state. Additionally, they observed that students taught by emotional design reported less difficulty and performed better on comprehension but not on transfer test, while non-significant main effect for mental effort was found. Mayer and Estrella (2014) conducted a similar study with college students and discovered that the emotional design group performed significantly better than the neutral group on retention but not on transfer test. Additionally, the emotional design group invested greater mental effort than the control group when study time was limited (Experiment 1), but no significant difference was observed in terms of invested mental effort and perceived task difficulty when the study time was unlimited (Experiment 2).

Similarly, Brom, Hannemann, Stárková, Bromová, and Děchtěrenko (2016) found that anthropomorphic comical graphics added to a black and white animation only affected retention, but not transfer, state engagement or positive affect. Qualitative analysis of the interview data revealed that anthropomorphic graphics may serve as memory cues for students to remember parts of the instruction.

In order to test the effectiveness of emotional design, early studies compared multimedia materials which used multiple forms of emotional design features (i.e. use of color, anthropomorphism and baby-face bias together) with neutral design (i.e. no emotional design). However, when the effect of different design elements was decomposed, different results were obtained. For example, Plass et al. (2014) [Experiment 2] demonstrated that while anthropomorphic rounded objects used independently from appealing colors induced positive emotions, use of colors alone did not. It was also shown that both color and anthropomorphic rounded objects enhanced comprehension. Transfer, on the other hand, was facilitated only by anthropomorphic rounded objects used together with neutral colors. Results revealed no main or interaction effect for mental effort investment. Contrarily, Park et al. (2015) showed that anthropomorphism had an attention-capturing effect, yet it had no significant effect on positive emotions or learning achievement compared to the use of color on its own. In addition, no main effect for mental effort investment was found between groups taught with and without anthropomorphism. These findings indicate that different emotional design elements lead to different results.

2.4. Problem statement

Previous research on the affect of emotional design in multimedia have been criticized for the following reasons. First, most of the studies on emotional design were conducted with college students. Second, experiment times were quite short (Thalheimer, 2004) due to the limitations of laboratory conditions. In order to increase ecological validity, longer experiment times within more authentic environments may be needed (Mayer & Estrella, 2014; Park et al., 2015). Third, in most of the previous research, self-reported paper-and-pencil-based measures (e.g., Positive Affective Scale of PANAS, Watson, Clark, & Tellegen, 1988) were administered to detect emotions (Plass et al., 2014; Um et al., 2012). For further studies, researchers recommend using more direct measures (Dong, 2007; Mayer & Estrella, 2014). Finally, as indicated, the discrimination of different design features and their distinctive effects on positive emotions and learning was not at the expected level (Heidig et al., 2015). Therefore, the purpose of this current study is to explore the effect of different emotional design features incorporated within multimedia instruction on positive emotions, mental effort and academic achievement. Four multimedia materials were created, each of which differed in terms of emotional design extent and complexity. The Neutral Design (ND) included only grayscale coloring; whilst in the Colorful Design (CD), grayscale coloring was exchanged with bright, saturated, attention-grabbing colors. The Anthropomorphic Design (AD) involved expressive anthropomorphism and positive exaggerated facial expressions of human characters. Finally, for Anthropomorphic Design and Sound Effects (ADSE), interesting sound effects were added to AD.

3. The current study

3.1. Research questions and hypotheses

The following research questions and hypotheses were formulated based on the purpose.

Research Question and Hypothesis 1. How do the use of different levels and complexity of emotional elements in multimedia affect positive emotions? Based on the previous research, it was hypothesized that students' positive emotions are significantly affected by the use of different levels and complexity of emotional design in multimedia. More specifically, it was hypothesized that as the amount of emotional design elements increase, positive emotions will also increase.

Research Question and Hypothesis 2. How do the use of different levels and complexity of emotional elements in multimedia affect students' mental effort ratings? Based on the previous research, it was hypothesized that students' mental effort ratings are significantly affected by the use of different levels and complexity of emotional design in multimedia. More specifically, it was hypothesized that as the amount of emotional design elements increase, mental effort will also increase.

Research Question and Hypothesis 3. How do the use of different levels and complexity of emotional elements in multimedia instructional materials affect students' learning achievement, controlling for the effect of prior knowledge and working memory capacity? Based on the previous research, it was hypothesized that learning achievement is significantly affected by the use of different levels and complexity of emotional design in multimedia, controlling for the effect of prior knowledge and working memory capacity. More specifically, it was hypothesized that as the amount of emotional design elements increase, learning achievement scores will also increase.

4. Method

4.1. Participants and design

Participants of the study were 113 7th grade students from a public middle school in an urban area of Turkey. The school had a total of 14 classes for the 7th grade; of which, four classes were selected for the current study on the basis of their teacher's having volunteered to participate in the study. Due to school regulations and a three week duration of the study, random assignment of subjects to the conditions was not feasible. Instead, conditions were randomly assigned to the four classes. Seven students were absent during some parts of the study. For this reason, their data were excluded from the statistical analysis. In total, data was gathered from 106 students (46% female and 54% male) for the study, with 28 students in the ND group (40% female and 60% male), 25 in the CD group (48% female, 52% male), 28 in the AD group (54% female and 46% male), and 25 students in the ADSE group (44% female and 56% male). All of the students had previously taken the basic information technology and software course during their 5th and 6th grades.

The study adopted a pretest/posttest quasi-experimental design. The dependent variables of the study were positive emotions, mental effort ratings, and posttest achievement scores. The independent variable was the treatment condition. Pretest achievement and working memory capacity were selected as covariates.

4.2. Design and development of multimedia instructional materials

Four types of multimedia learning software for four treatment conditions were developed within the scope of the study. Multimedia materials were developed for the "Work, Energy, and Energy Conservation" subtopics of the 7th grade science curriculum (Milli Eğitim Bakanlığı [Ministry of National Education], 2015). Learning outcomes determined by the Ministry of National Education were regarded as a framework in developing the materials. In designing multimedia for the groups, Cognitive Theory of Multimedia Learning (CTML) (Mayer, 2009) together with Cognitive Load Theory (CLT) (Paas & Sweller, 2014) were used as theoretical frameworks. For the design of storyboards and scenarios related to the materials, five science teachers, two subject area experts, and an instructional designer were periodically consulted concerning their views and evaluations. For the development of the multimedia software, an iterative process of *development-evaluation-feedback-development* was followed.

Neutral Design (ND) material included only grayscale coloring and served as the base version for all materials. A sample screenshot from the ND material is provided in Fig. 1.

For the ND group, neutral facial expressions were used for humans. For instance, in the original version of the material, the

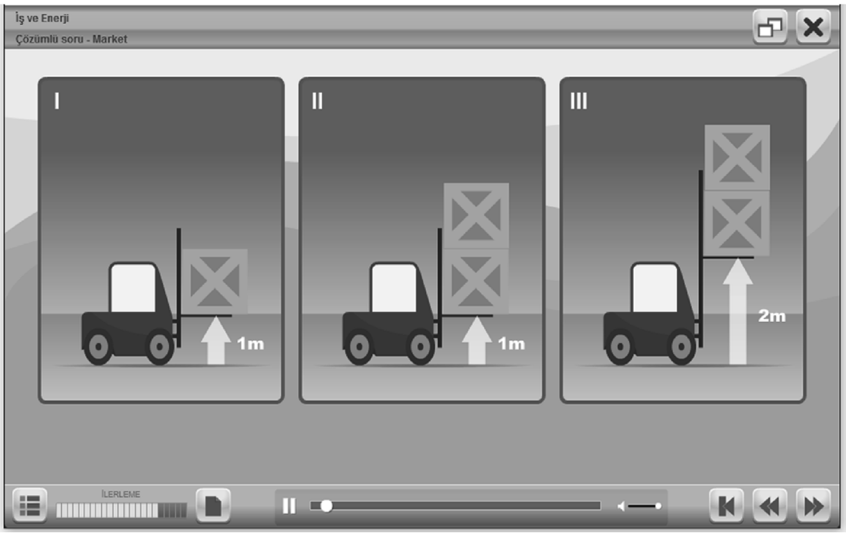

Fig. 1. Sample screenshot from the ND material.

character's facial expression was neutral and did not change whilst lifting weights (see Fig. 2).

The second version adopted the *Colorful Design* (CD) approach. For this group, saturated, bright, and attention-grabbing colors were used instead of grayscale coloring. Both warm and cool colors were used in the design. As Lohr (2007) indicated, warm colors tend to advance, whereas cool colors tend to recede. Based on this suggestion, warm colors were used for the main objects in the screen, which were required to be focused upon. In this way, it was aimed to highlight the more important parts of the instruction. A sample screenshot from the CD material is given in Fig. 3.

For the CD group, the facial expressions of human characters were the same as for the ND group. Neutral facial expressions were used for human beings, whereas no facial expressions were used for lifeless objects. Fig. 4 shows a sample screenshot depicting a human character used for the CD group.

Fig. 2. Sample screenshot of a human character from the ND material.

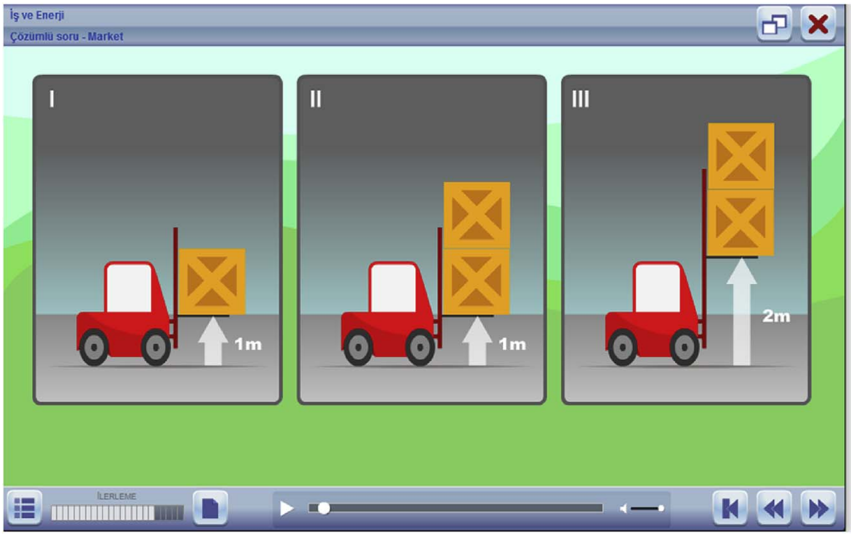

Fig. 3. Sample screenshot from the CD material.

Fig. 4. Sample screenshot of a human character from the CD material.

The third material adopted the Anthropomorphic Design (AD) approach in which expressive anthropomorphism was used for lifeless objects. For example, when an object's energy increased, the facial expression of the related object became more positive, depicting a smiling face (see Fig. 5).

For the AD group, to increase positive emotions and decrease negative emotions, human characters were depicted as having positive facial expressions with smiling faces. However, the facial expressions used for humans were expressive rather than stable. In other words, characters' facial expressions changed based on what they were doing. Fig. 6 represents a human character used for the AD group:

The last material adopted the *Anthropomorphic Design and Sound Effects* (ADSE) approach. In addition to expressive facial expressions, interesting context-dependent sound effects were also used. For example, when an animation depicted the force of friction

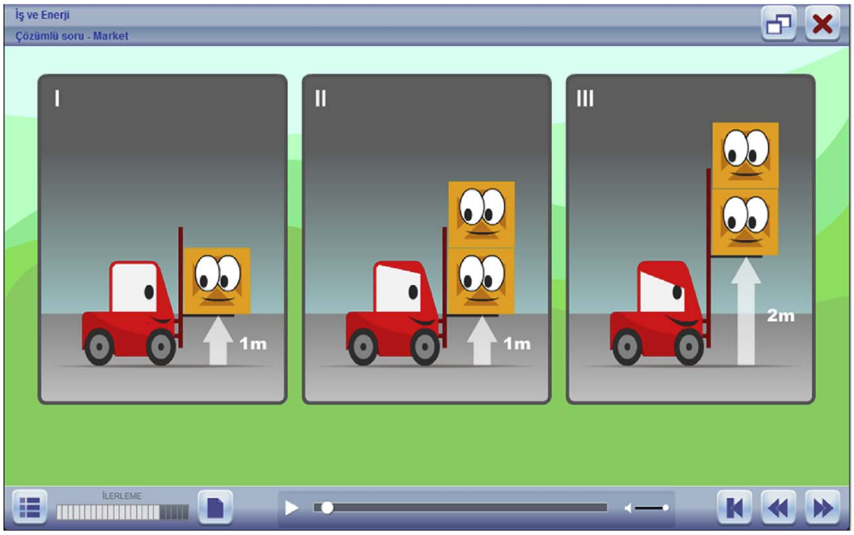

Fig. 5. Sample screenshot from the AD material.

Fig. 6. Sample screenshot of a human character from the AD material.

on a rough surface, a friction sound effect was also added. As the screenshot for ADSE would visually match that used for AD, no screenshot for ADSE is provided.

4.3. Data collection instruments

4.3.1. emWave desktop Emotion Recognition Technology

There are many new approaches such as Cardiovascular Measures, EMG, fMRI, EEG and EDA to identify human emotions (El-Nasr, Morie, & Drachen, 2010). The current study employed emWave Desktop Emotion Recognition Technology (see Fig. 7), one of the cardiovascular measures of emotion detection. emWave was originally a stress detector developed by the HeartMath Institute. The system comprises of an ear pulse sensor, a USB module and software that plots the heart rhythm in real time. The sensor measures

Fig. 7. emWave Emotion Recognition Technology and Students using Multimedia Instructional Materials.

the speed of blood flow reaching to the ear lobe per unit of time and calculates Heart Rate (HR) and Heart Rate Variability (HRV) values.

The emWave system works according to changes between the sympathetic and parasympathetic nervous systems of the Automatic Nervous System (ANS). The sympathetic nervous system accelerates the HR by stimulating cardiac muscles. Conversely, the parasympathetic nervous system decreases the HR and is responsible for the maintenance and control of the internal organs when at rest (Hasan, Bègue, & Bushman, 2013). Incoherence between the sympathetic and parasympathetic nervous systems is observed when the subject feels negative emotions (Childre & Cryer, 2004; as cited in Hasan et al., 2013). In contrast, when a subject feels positive emotions, increased synchronization or harmony is observed between the two components of the ANS (Bradley et al., 2010). This situation is also termed as the coherence state (Rozman, McCraty, & Tomasino, 2008). Coherence refers to the psychophysiological state "reflected by a smooth, sine wave-like (coherent) pattern in the heart rhythms" (Rozman et al., 2008, p. 197). Instead of HR, the emWave system uses HRV, which ignores momentary variations in the HR when the average HR is measured (Rozman et al., 2008). The interface screen of the emWave software program is shown in Fig. 8.

Fig. 8 shows a student's heart rhythm pattern and a set of values related to this pattern while studying multimedia instructional material. The heart rhythm pattern can be seen at the top of the screen. The bottom-left of the screen shows the accumulated coherence score of the subject, and the bottom-right displays the coherence ratios. The coherence ratios show the subject's psychological coherence score based on HRV analysis. Along with the session, the coherence ratios scores are updated every 5 s in association with the heart rhythm coherence. There are 3 bars in the coherence ratios graph. The red bar indicates low coherence (Fig. 8 shows value = 12), the blue bar indicates medium coherence (Fig. 8 shows value = 24), and the green bar indicates high coherence (Fig. 8 shows value = 64).

Each bar shows the amount of time spent in low-, medium- and high-coherence zones by ratios, proportionate to the total time spent in the complete session (emWave Pro Library, 2016). The red bar indicates low coherence between sympathetic and parasympathetic activity. This zone also indicates the amount of "negative emotions" experienced throughout the session. The blue bar

Fig. 8. Heart Rhythm Pattern of Student who tends to Experience Positive Emotions.

indicates medium coherence between sympathetic and parasympathetic activity. This zone also reflects the amount of "peaceful emotions." Finally, the green bar indicates high coherence between sympathetic and parasympathetic activity. This zone indicates the amount of "positive emotions" experienced in the entire session (Chen & Wang, 2011). The sum of low-, medium- and high-coherence ratios always equals to 100%. The positive emotion is calculated by adding medium coherence and high coherence values and found to be 88 in this experiment (24 + 64). Based on the coherence values, it could be argued that the subject mostly experienced positive emotions (total coherence value = 88) during the session. In addition, Heart Rhythms Patterns of the student were consistent with the calculated coherence value, showing a smooth, consistent, sine-wave-like pattern.

4.3.2. Mental effort test

In the current study, Subjective Rating Scale was used to measure the subjects' mental efforts. The scale was originally developed by Paas and Van Merriënboer (1993), and adapted to Turkish culture by Kılıç and Karadeniz (2004) (Cronbachs' $\alpha = 0.78$). Since the current study lasted for a period of three weeks, the scale was implemented three times. At the end of each session, students rated their perceived invested mental efforts through a single-item, nine-point scale ranging from "1 (very, very low mental effort)" to "9 (very, very high mental effort)." Lower scores indicate lower levels of perceived mental effort investment, while higher scores indicate higher levels of perceived mental effort investment. Since the experiment required a three-week duration, the mean score of mental effort investment was calculated by using each participants' scores obtained over the three weeks, and Cronbachs'α found to be 0.79 for the scale in the current study.

4.3.3. Working memory measures

In the current study, working memory capacity was included in the analysis as covariate. Two measures were used (i.e., Mr. Peanut and Backward Span) in order to detect students' working memory capacity and a single measure was obtained by combining the two test scores. The reason why both Mr. Peanut and Backward Digit Span tests were used is twofold. First, by combining the two measures, it was expected to perform as a more reliable measurement of working memory capacity. Second, the researchers thought that using the Mr. Peanut Test only may rely too heavily on visuospatial skills alone. Therefore, the researchers decided that using both a measure with an audial element (Backward Digit Span) and one visuospatial would be a more reliable solution. Moreover, the researchers also believed that the central executive component of working memory capacity may be involved in the Backward Digit Span test, since the participants are required to recall items in reverse order (Kessels, van den Berg, Ruis, & Brands, 2008).

4.3.3.1. Mr. Peanut (Mr. Cucumber) test. Mr. Peanut or Mr. Cucumber Test is a Working Memory Capacity (WMC) test developed to measure the visuospatial dimension of the working memory. The test was originally developed by Case (1985) to detect short-term visual memory in a younger population. In this test, a clown figure (Mr. Peanut) who decorates himself with colored stickers is shown to the participants. The task of the participants is to recall the locations of the stickers by showing the position of the stickers on the clown's body. In the current study, a computerized version of the test (Inquisit Scripts, 2016) was used. Validity analysis of the Mr. Peanut Test was conducted by Jager (2004). Among the types of the validities, concurrent validity was used. The scores of the participants' responses to the Mr. Peanut Test were correlated with the scores of the participant responses to the WMC dimension of the Wechsler Intelligence Scale for Children (WISC-IV), and a high, positive and significant correlation was found ($r = 0.51$) (Jager, 2004). Spearman–Brown split-half coefficients reliability analysis of the test was conducted by Morra and Camba (2009) and found to be 0.81. For the current study, Spearman–Brown split-half coefficient reliability of the test was calculated based on a dichotomy of odd–even items and was found to be 0.87.

4.3.3.2. Backward Digit Span test. In order to explore verbal WMC, Backward Digit Span Test of the Psychology Experiment Building Language (PEBL) (Mueller & Piper, 2014) was employed. In the Backward Digit Span Test, participants are shown a series of digits. The participant's task is to recall the digits in reverse order to the order that was shown to them. The highest numbers of digits correctly typed in reverse order equals to the score of the subject's WMC (Morra, 1994). Reliability of the Backward Digit Span Test was calculated by Morra and Camba (2009) based on the Spearman–Brown split-half coefficients and was found to be 0.78. Moreover, citing the work of Wechsler (1981), Waters and Caplan (2003) reported test–retest reliability of the Backward Digit Span Test as 0.83. For the current study, the Spearman–Brown split-half coeffient reliability was calculated as 0.80.

4.3.4. Work, energy, and Energy Conservation achievement test

The achievement test consisted of two parts; recall and transfer of learning. The test was used as a pretest before the experiment to measure students' prior knowledge about the topic, and was repeated as a posttest after the experiment in order to measure the students' level of achievement. The test was developed in line with the "Work, Energy, and Energy Conservation" subtopics' learning outcomes as determined by the curriculum (Milli Eğitim Bakanlığı [Ministry of National Education], 2015). During the development of the achievement test, a committee of three science teachers and two subject-matter experts were periodically consulted in order to determine the content validity, difficulty levels and congruence of the test items with the learning outcomes.

There were a total of 20 multiple-choice questions in the recall test (Cronbachs' $\alpha = 0.72$). The recall test items were developed on the basis of the "remember" dimension of Bloom's taxonomy, "When the objective of instruction is to promote retention of the presented material in much the same form in which it was taught, the relevant process category is remember" (Mayer, 2002, p. 228). One of the recall test items states, "Which of the followings does gravitational energy depend on?" One point for each correct answer and zero point for incorrect answers were awarded; hence, the participants' total scores obtained from recall test could vary from a minimum score of zero to a maximum score of 20.

There were a total of 14 open-ended items in the transfer test. According to Mayer (2002), dimensions other than remember (Understand, Apply, Analyze, Evaluate, and Create) correspond to the transfer of learning which mostly focuses on measuring the degree of how much meaningful learning occurs. A sample item used for the transfer test is as follows: “You can see a sparrow, an eagle and a pigeon sitting in the branches of a tree at the same height. Compare these animals’ gravitational potential energies and explain your conclusion.” One point for each correct answer and zero points for incorrect answers were awarded. Hence, the scores of the participants obtained from transfer test could vary from a minimum of zero up to a maximum of 14. Two scorers independently scored student responses to the open-ended items. As indicated by inter-rated reliability, there was a strong level of association between the two scorers (ICC = 96). After the scoring procedure was finalized, the two scorers met to compare their results by discussing items which they had scored differently. The discussion concluded after the two scorers reached consensus on the scoring of the specific item.

4.4. Procedures of the study

The experiment was conducted during the fall semester of the 2015–2016 academic year. The curriculum suggested a three-week instructional period to cover the topics of “Work, Energy, and Energy Conservation” (Milli Eğitim Bakanlığı [Ministry of National Education], 2015). Accordingly, the experimental investigation was carried out for a period of three weeks (one session per week). However, the researchers spent an extra 2 weeks at the school before the actual start of the experiment. For one week, a pilot study was implemented with volunteer students in order to test whether or not any technical problems existed, in terms of either hardware or software, that had the potential to hinder the participant students from properly using the emWave Emotion Recognition Technology (ERT) or the multimedia instructional materials. After that, an orientation study was conducted for the students who were to participate in the actual study in order to demonstrate the correct usage of the ear sensor and multimedia instructional materials. During the orientation week, the researchers also collected data to determine the entry level characteristics of the students. First, in order to examine the participants' prior knowledge levels, a pretest regarding the topics of “Work, Energy, and Energy Conservation” was implemented. Second, for the detection of the students' working memory capacity levels, the Mr. Peanut (Mr. Cucumber) and Backward Digit Span tests were applied to the students. During the subsequent three weeks, the actual experiment was conducted. Students learned the topic with the aid of multimedia instructional materials individually, created based on different emotional design approaches. Throughout the experiment, the students’ positive emotions while using the different types of multimedia instructional materials (ND, CD, AD, and ADSE) were detected and recorded using emWave Emotion Recognition Technology (ERT) based on their Heart Rate Variability (HRV). Randomly selected classes were treated as groups. Therefore, each student belonged to one of the groups (ND, CD, AD, or ADSE) and participated in learning and emWave sessions within the same groups.

Before the commencement of the experiment, the authorized researcher at the HeartMath Institute was contacted via email and Skype. The HeartMath researcher suggested not using the emWave system throughout the whole session, since recording whole sessions could produce noisy data, due to maintaining control of younger students during the experiment would be more challenging compared to other age groups. Instead, the HeartMath researcher suggested that the emWave system should be applied for a certain amount of time during the first and last periods of a session. Accordingly, for the first two weeks, the emWave system was used for the first and last 15 min of each session. For the third (final) week of the experiment, the emWave system was used for the first and last 5 min of the session, due to the learning time duration for the material being shorter compared to the material in the first two weeks.

In summary, two datasets were collected each week, with a total of six datasets collected during the experiment (2×3 weeks = 6 datasets). After the experiment concluded, the same achievement test was administrated as a post achievement test. Finally, in order to measure how much mental effort the students exerted during the learning process for each topic, a Subjective Rating Scale for mental effort was implemented.

4.5. Data analysis

Both parametric and non-parametric techniques were implemented for the analysis of the data. The alpha significance level for each quantitative analysis was set as 0.05. Before proceeding with the analysis, potential confounding variables (covariates) were investigated. Selecting covariates depends on a variety of factors. First, the theoretical suggestions of previously published research should be taken into consideration. Second, covariates should significantly correlate with the dependent variable (Pallant, 2010; Stevens, 2009). Considering the first factor, the literature regarding Cognitive Load Theory (CLT) and Cognitive theory of multimedia learning (CTML) revealed that prior knowledge and working memory capacity could be confounding variables for multimedia learning (Plass, Kalyuga, & Leutner, 2010; Rey, 2012; Sanchez & Wiley, 2006). For the second factor, inter-correlations between the variables were assessed (see Table 1).

Taking WMC and prior knowledge as potential confounding variables, it appears from Table 1 that these variables are not significantly correlated with positive emotions and mental effort. On the other hand, both WMC and prior knowledge are significantly correlated with post achievement scores (recall and transfer). Therefore, when comparing between-group emotional states (Hypothesis 1) and mental efforts (Hypothesis 2), both WMC and prior knowledge were excluded as confounding variables (covariates).

To test Hypothesis 1, after each session, the students’ emotion data file (emwave.emdb) was extracted from the computers they had used. Since the duration of the experiment time corresponds to three weeks and data was collected twice a week, a total of six datasets ($2 \times 3 = 6$) was obtained. For each student, the average positive emotion was calculated by summing all of the datasets and dividing that sum by the number of the datasets ($n = 6$). Since normality analysis showed that the data did not distribute normally, Independent Samples Kruskal-Wallis test was used.

Table 1
Inter-correlations between variables.

	1.	2.	3.	4.	5.	6.	7.
1. Positive Emotions	1						
2. WMC	.01	1					
3. Mental Effort	.06	.01	1				
4. Pretest Recall	.03	.30**	-.03	1			
5. Pretest Transfer	.06	.25*	-.07	.48**	1		
6. Posttest Recall	.04	.42**	.04	.34**	.45**	1	
7. Posttest Transfer	-.09	.40**	.02	.39**	.56**	.63**	1

Note. **$p < .01$, *$p < .05$, WMC = Working Memory Capacity.

In order to test Hypothesis 2, first, the average mental effort exerted by the students was calculated by summing each week's mental effort score and diving the score by the number of weeks (n = 3). Independent Samples Kruskal-Wallis test was conducted as the data did not meet normality assumption for the mental effort.

To test Hypothesis 3, ANCOVA test was conducted twice; both for recall and transfer of learning. Students' prior knowledge (recall and transfer) and working memory capacity (WMC) were taken as covariates. WMC score was calculated as the sum of students' scores obtained from the Mr. Peanut and Backward Digit Span tests.

5. Results

5.1. Positive emotions

Independent Samples Kruskal-Wallis test was conducted to ascertain whether or not positive emotions were significantly affected by use of different levels and complexity of emotional design in multimedia. Table 2 displays the results of the test. Results revealed a statistically significant difference in positive emotion scores among treatment conditions, H (3) = 9.86, p = .020, indicating that positive emotions were significantly affected by the use of different emotional design approaches in multimedia. Mean group ranks were 48.59 for ND, 41.02 for CD, 58.02 for AD and 66.42 for ADSE. Another way to explore the group differences is to look at the box plots generated for the groups (Field, 2013).

As Fig. 9 displays, medians related to positive emotion scores tend to increase across groups, as levels and complexity of emotional design increase (i.e., from ND to ADSE). Follow-up tests with adjusted alpha values (Bonferroni correction) were conducted to investigate group differences in detail. The effect size (r) was calculated by using the $z/\sqrt{N}$ formula (Field, 2013). According to Cohen (1988, 1992, as cited in Field, 2013), the effect sizes of 0.10, 0.30 and 0.50 indicate small, medium and large effects respectively.

Table 3 shows the results of pairwise comparisons. Results showed that the only significant difference with respect to positive emotions was discovered between the ADSE and CD groups (p = .020), while other differences were insignificant. The mean rank was 66.42 for the ADSE and 41.02 for the CD group respectively. This result could be interpreted as using ADSE in multimedia produced significant positive emotions in students compared to using CD.

Considering the effect sizes, the largest was observed in the CD-ADSE comparison, indicating a large effect (r = 0.41). One interesting point was that although the ND-ADSE comparison was not significant, the size of the effect was nearly medium (r = 0.29), showing that the use of ADSE in multimedia may have a practically moderate effect that contributes to the induction of positive emotions in students compared to the use of ND. In the same manner, although the CD-AD comparison was not significant, the effect size for this comparison was nearly medium (r = 0.28), indicating from the viewpoint of practical significance that use of AD in multimedia may have moderate effect on inducing positive emotion in students compared to the use of CD. Other effect sizes were small in terms of magnitude, with values of 0.12 for CD-ND, 0.14 for AD-ADSE and 0.15 for the ND-AD groups. Note that the smallest effect (r = 12) was observed between the CD and ND groups, which may be interpreted as using CD in multimedia may have the least effect on inducing positive emotions in students compared to the use of ND. In summation, consistent with Hypothesis 1, the use of different levels of emotional design in multimedia contributed significantly to the induction of positive emotions in students.

Table 2
Results of the Kruskal-Wallis test regarding students' positive emotion scores.

Group	*N*	Mean Rank	*df*	$\chi 2$	*P*
ND	28	48.59	3	9.86	**.020**
CD	25	41.02			
AD	28	58.02			
ADSE	25	66.42			

Note. ND = Neutral Design, CD = Colorful Design, AD = Anthropomorphic Design, ADSE = Anthropomorphic Design and Sound Effects. Boldface indicates significant difference.

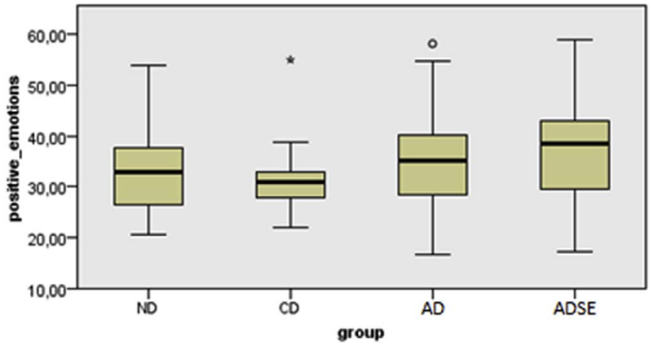

Fig. 9. Boxplots showing the change of positive emotions across groups.

Table 3
Results of pairwise comparisons between groups regarding students' positive emotion scores.

Sample 1-Sample 2	Test Statistic	*SE*	*z*	*p*	Adj.*p*.	*R*
CD-ND	7.57	8.46	0.90	.37	1.00	.12
CD-AD	−17.00	8.46	−2.01	.04	0.27	.28
ND-AD	−9.43	8.22	−1.15	.25	1.00	.15
CD-ADSE	−25.40	8.70	−2.92	.00	**0.02**	.41
ND-ADSE	−17.83	8.46	−2.11	.04	0.21	.29
AD-ADSE	−8.40	8.46	−0.99	.32	1.00	.14

Note. ND = Neutral Design, CD = Colorful Design, AD = Anthropomorphic Design, ADSE = Anthropomorphic Design and Sound Effects. Boldface indicates significant difference.

5.2. Mental effort ratings

Independent Samples Kruskal-Wallis test was used to explore whether or not mental effort ratings were significantly affected by the use of different levels and complexity of emotional design. Table 4 shows the results of the Kruskal-Wallis test. The test was significant, H (3) = 9.823, p = .020, showing that mental effort ratings were significantly affected by the use of different levels and complexity of emotional design in multimedia. Mean ranks were 45.11 for the ND group, 67.84 for CD, 56.73 for AD, and 44.94 for the ADSE group. A boxplot graph was provided to explore the initial picture regarding the difference among groups (see Fig. 10).

As Fig. 10 illustrates, mental effort ratings tend to increase from the ND to CD groups. However, from CD to AD and ADSE groups, students' mental effort ratings seem to decrease linearly, as the levels and complexity of emotional design increases. To explore the exact nature of the differences between groups in detail, pairwise comparisons with adjusted alpha values (Bonferroni correction) were conducted (see Table 5).

As shown by Table 5, two significant differences were observed between the ND-CD and ADSE-CD comparisons, while other differences were insignificant. Students who were taught with Colorful Design (CD) exerted higher levels of mental efforts than students taught by Neutral Design (ND) (p = .04), with the respective mean rank values of 67.84 and 45.11. In addition, students who were taught by Colorful Design (CD) exerted higher levels of mental effort compared to the students who were taught by Anthropomorphic Design and Sound Effects (ADSE) (p = .050), with the respective mean rank values of 67.84 and 44.94. Considering the effect sizes, the largest effect sizes were observed for the ADSE-CD and ND-CD comparisons (r = 0.37), indicating medium to large effects. Small to medium effect sizes were observed for the ND-AD, ADSE-AD and AD-CD comparisons respectively (r = 0.19, r = 0.19, r = 0.18). The effect size for the ND-ADSE comparison was quite low (r = 0.00), showing nearly zero effect.

According to the results, consistent with Hypothesis 2, it was discovered that students' mental effort ratings were significantly affected by use of different levels and complexity of emotional design in multimedia.

Table 4
Results of the Kruskal-Wallis test regarding students' mental effort ratings.

Group	*N*	Mean Rank	*df*	χ^2	*p*
ND	28	45.11	3	9.823	**.020**
CD	25	67.84			
AD	28	56.73			
ADSE	25	44.94			

Note. ND = Neutral Design, CD = Colorful Design, AD = Anthropomorphic Design, ADSE = Anthropomorphic Design and Sound Effects. Boldface indicates significant difference.

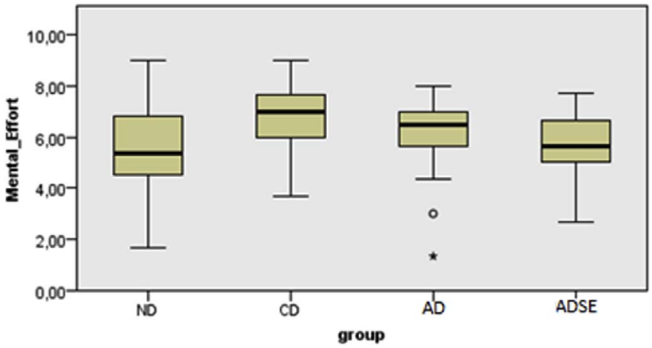

Fig. 10. Boxplot showing the change of mental effort ratings across groups.

Table 5
Results of pairwise comparisons between groups regarding students' mental effort.

Sample1-Sample2	Test Statistic	SE	*z*	*p*	Adj.*p*.	*r*
ND-CD	−22.73	8.44	−2.69	.01	**0.04**	.37
ADSE-CD	22.90	8.68	2.64	.01	**0.05**	.37
ND-AD	−11.63	8.20	−1.42	.16	0.94	.19
AD-ADSE	11.79	8.44	1.40	.16	0.97	.19
ADSE-ND	0.17	8.44	0.02	.98	1.00	.00
AD-CD	11.11	8.44	1.32	.19	1.00	.18

Note. ND = Neutral Design, CD = Colorful Design, AD = Anthropomorphic Design, ADSE = Anthropomorphic Design and Sound Effects. Boldface indicates significant difference.

5.3. Learning achievement scores

The final aim of the study was to test whether or not achievement scores were significantly affected by the use of different levels and complexity of emotional design in multimedia. Two separate ANCOVA tests were conducted for recall and transfer of learning. Prior knowledge and working memory capacity (WMC) were regarded as covariates.

With regard to recall of learning, it was found that both the pretest recall, F (1, 100) = 7.13, p = .009 and WMC, F (1, 100) = 13.27, p = .000, were significant covariates. Considering the main effect, it was found that after adjusting for pretest recall and WMC scores, there was a significant mean difference in students' posttest recall scores across the treatment groups, F (3, 100) = 3.61, p = .016. That is, use of emotional design in multimedia significantly affected students' post-recall scores. The strength of the relationship between treatment conditions and post-recall scores was medium, as indicated by a partial eta-squared of 0.10 (Green & Salkind, 2005). In other words, the use of different levels of emotional design in multimedia accounted for 10% of variance in students' post-recall test scores. Post-hoc comparisons with Bonferroni correction were performed for multiple comparisons to reveal which groups differed significantly (see Table 6).

As Table 6 displays, the CD group has the highest adjusted mean score (M = 14.31) of all treatment conditions. The adjusted mean scores of the ADSE, AD and ND groups are 14.06, 13.86 and 11.85 respectively. Among all the pairwise comparisons, only the difference between the CD and ND groups was significant; with differences between groups for other pairwise comparisons found as not significant. The CD group (M = 14.31) significantly outperformed the ND group (M = 11.85) in terms of their posttest recall scores.

Considering the transfer of learning, results indicated that both the pretest transfer, F (1,100) = 37.82, p = .000 and WMC, F (1,100) = 10.83, p = .001, were significant covariates. Considering the main effect, the treatment effect was not significant, F (3,100) = 0.94, p = .422. Since the main effect regarding post-transfer scores was not significant, pairwise comparisons were not performed.

Table 6
Means, adjusted means and adjusted means difference regarding posttest recall scores.

Group	*M*	*Adjusted M*	1.	2.	3.	4.
1. ND	11.61	11.85				
2. CD	14.44	14.31	**2.46***			
3. AD	13.93	13.86	2.01	-.45		
4. ADSE	14.12	14.06	2.21	-.25	.20	

Note. ND = Neutral Design, CD = Colorful Design, AD = Anthropomorphic Design, ADSE = Anthropomorphic Design and Sound Effects. Boldface indicates significant difference.

Given the results as reported, it could be argued that the results are partially congruent with Hypothesis 3. Among all emotional design elements, only the use of Colorful Design (CD) approach in multimedia significantly affected recall of learning compared to the use of Neutral Design (ND); whereas, none of the emotional design elements affected transfer of learning significantly.

6. Discussion and conclusion

The purpose of the current study was to investigate the effect of various emotional design approaches in instructional multimedia on positive emotion, mental effort and learning achievement.

6.1. Positive emotions

Based on the cardiac coherence values obtained from HRV analysis through emWave device, this study's results demonstrated that positive emotions differed significantly with respect to the type of emotional design. That is, materials which adopted different emotional design approaches (ND, CD, AD, and ADSE) contributed significantly to the induction of positive emotions among students. Generally, as the extent of emotional design increased, positive emotions also tended to increase. This result is consistent with the literature on emotional design, suggesting that affective qualities of learning materials increase positive emotions during learning (Plass et al., 2014; Um et al., 2012). Although the overall effect was significant, the only significant difference among groups was found between the CD group that adopted Colorful Design and the ADSE group that adopted Anthropomorphic Design and Sound Effects. ADSE material produced more positive emotions in students compared to the CD material. Given this finding, it seems that multimedia material in which the combination of interesting sound effects, facial expressions (anthropomorphism of lifeless objects and facial expression of human characters) and attention-grabbing colors was found to be much more effective in producing positive emotions than multimedia material that used attention-grabbing colors only.

Although the ADSE group, which included all emotional design approaches, resulted in the highest level of average positive emotions, this group also saw the largest deviation. This may be because the ADSE design may have attracted students' attention and curiosity that in turn resulted in enhanced positive emotions. On the other hand, the design may also have disturbed or annoyed those students, especially those who are intrinsically motivated and do not need any additional emotional design elements in order to learn (Clark, 1982).

It is interesting to note that although the difference was not significant, considering the effect sizes, it was observed that the ADSE-ND and AD-CD comparisons were nearly medium. The size of the effect was smallest between the CD-ND groups, meaning from the viewpoint of practical significance that use of CD in multimedia has the least effect on induction of positive emotion compared to the use of ND. Taking into consideration all statistical and practical differences, the results are consistent with the current literature on emotional design. That is, as expected, more positive emotions were induced among students as the amount of emotional design elements increased (Dong, 2007; Plass et al., 2014; Um et al., 2012).

6.2. Mental effort

With regard to mental effort, the results showed that multimedia materials with different levels of emotional design affected the amount of mental effort. However, it was surprising that mental effort did not change in parallel with the extent of emotional design. Specifically, students taught by CD material spent more mental effort compared to the students taught by ND; whereas, students taught by ADSE spent less mental effort compared to students taught by CD.

The limited literature found on emotional design shows that the use of emotional design elements in multimedia decreased the level of perceived task difficulty compared to the neutral design (Plass et al., 2014; Um et al., 2012), suggesting that the use of emotional design elements in multimedia might have resulted in students perceiving learning tasks as easy tasks to do. In line with this finding, for the current study, students in the ADSE group might have exerted less effort to learn the subject since they might have perceived the task as easy to do due to the use of emotional design elements (i.e., anthropomorphism of inanimate objects, expressive facial expressions of human beings and sound effects). Albeit indirectly, this result is also congruent with the literature showing that affective design elements of a material reduced the amount of time devoted to a task (Haaranen, Ihantola, Sorva, & Vihavainen, 2015; Sitzmann & Johnson, 2014). From these results, it may be plausible to argue that mental effort to be devoted to a task can be influenced by the structure of a learning environment. Learners manage their mental effort according to the perceived task difficulty of courses (Pintrich, 2000). The perception of "being difficult to learn" of the instructional messages that are delivered by the learning environments might actually lead students to exert a substantial amount of effort in order to make sense of those instructional messages (Sitzmann & Johnson, 2014). In the current study, the decrease of mental effort investment from the CD group to other groups may be due to the use of interesting emotional design features in multimedia being perceived by the students as "easy to learn" because of their "entertaining appearance." Considering the entertaining appearance of the AD and ADSE materials, students might have made inaccurate judgments regarding the amount of effort they must exert in order to achieve (Clark, 1982). A similar pattern of results was also found in Salomon's (1984) study who found that "easy" materials or those considered to have an "entertaining appearance" may result in less mental effort investment.

In summary, inconsistent with Mayer and Estrella (2014) [Experiment 1], in the current study, students who used AD and ADSE did not show a higher level of mental effort compared to students who used CD. On the other hand, the use of CD material resulted in more mental effort investment in students compared to the use of ND material. Considering grayscale color scheme structure of the ND material, as expected, it was probable that the students would not give their best effort to learn the contents of ND material due to underestimation of the design.

6.3. Learning achievement (recall and transfer)

The third finding of the current study was that the use of different emotional design approaches in multimedia affected recall but not transfer of learning. With regard to recall of learning, it was discovered that students taught by CD material outperformed students taught by ND. What can be deduced from these results is that only the use of attention-grabbing colors affected recall of learning scores. Facial expressions and sound effects seemed to affect neither recall nor transfer of learning. These results are consistent with the study conducted by Park et al. (2015), showing that expressive anthropomorphism does not contribute to learning compared to colorful design, and also the study of Haaranen et al. (2015) in showing that anthropomorphic graphics does not lead to higher achievement scores compared to abstract graphics. On the other hand, the current study's results are inconsistent with some other studies (e.g., Mayer & Estrella, 2014; Um et al., 2012), which showed that a combination of attention-grabbing colors and anthropomorphism was much more effective than neutral design. However, such studies adopted a “two groups comparison approach” by using multi-forms of emotional design approaches for the treatment group and none for the control group. In such studies, the significant difference may be due to weaknesses of the neutral design, rather than the strength of the emotional design, because for a neutral design that adopted grayscale or black and white coloring, it may be difficult for learners to discriminate and understand the essential and important parts of the instruction. Apart from the aforementioned studies, the current study investigated the effect of emotional design in a gradually increased manner in which the extent of emotional design increased gradually. In this way, the researchers had the chance to explore the graduating effect of distinct design elements on positive emotions and learning outcomes; which revealed different results patterns when compared to previous studies (e.g., Mayer & Estrella, 2014; Um et al., 2012). Subsequent studies investigated the effect of discriminating various design issues and their distinctive effect on learning, and found that different results were observed when those design elements were not considered together but one by one (Heidig et al., 2015; Park et al., 2015; Plass et al., 2014).

Considering all of these findings together in terms of learning achievement, from the viewpoint of surface and deep learning, it could be argued that at least one emotional design manipulation (attention-grabbing, saturated or bright colors) affected surface learning (as indicated by recall of learning). On the other hand, none of the emotional design manipulations had an effect on deep learning (transfer of learning), which is in line with the literature. For example, Brom et al. (2016) found that anthropomorphic faces and comical graphics only superficially affected recall of learning scores. Other studies also found that emotional design manipulations were mostly effective for surface learning (Mayer & Estrella, 2014; Plass et al., 2014 [Experiment 1]). As reported by Brom et al. (2016), only the original study of Um et al. (2012) showed emotional design to affect both deep and surface learning. To summarize, the current study found that learning outcomes were not greatly affected by the use of different levels or complexity of emotional design.

This current study employed a more direct measure of emotion recognition (emWave) in order to evaluate the effect of using different emotional design approaches on multimedia learning. To summarize, the results support the idea that positive emotional design may induce positive emotions in learners. However, positive emotions induced by multimedia materials may not greatly affect learning. Moreover, although the learners taught by multimedia with Anthropomorphic Design and Sound Effects (ADSE) exhibited positive emotions during the learning session, they invested less mental effort in order to learn a topic. It is likely that learners did not spend the necessary amounts of mental effort due to the perceived interesting and entertaining appearance of learning environments. Nevertheless, it could not be argued that either anthropomorphism or sound effects are without use, since they neither improved learners’ learning nor hampered it (as seductive details did).

The findings of the current study have a set of considerable implications that may be useful for instructional designers and practitioners for their future practices. First of all, instructional designers and practitioners should be careful in their considerations of using emotional design elements in multimedia. They should drop the myth that “liking” always results in “learning” (Clark & Lyons, 2010), and should be aware of the potential benefits and pitfalls of using emotional design elements in multimedia. For this reason, instructional designers should first give priority to promoting cognitive rather than emotional interest. They should first use cognitive design elements in multimedia congruent with the cognitive theories of learning, and then carefully incorporate emotional design elements into multimedia by considering the limited capacity of cognitive resources and students' perceptions. Simply augmenting emotional interest does not guarantee better learning (Park, 2005). On the other hand, evidence from this current study suggests that although expressive facial expressions along with sound effects did not produce better learning outcomes, they were effective in terms of increasing learners’ positive emotions. On that basis, the potential emotional and motivational benefits of such design elements should not be ignored by practitioners. Emotional design elements may be used for those students who possess low levels of individual interest towards a topic and for those who possess low motivation to learn. In this sense, analyzing the different characteristics of learners may help teachers utilize emotional design elements more efficiently. Instructional designers could design learning environments in a manner compatible with distinct learner characteristics. For example, with the help of personalized learning environments, the unique needs of different individuals could be satisfied.

6.4. Limitations of the study

A number of potential limitations to the current study need to be considered. The first limitation is about the lack of the baseline for the emotional measure. Unlike most previous research on emotional design (e.g., Park et al., 2015; Plass et al., 2014; Um et al., 2012), external mood inducement procedures as a proxy of the baseline could not be used due to the length of the experiment time and difficulty of the classroom management. Besides, learners' motivational profiles (e.g., intrinsic or extrinsic) and emotional states they brought to the experiment were not known specifically. Although addressed by the research limitedly (e.g., Knörzer, Brünken, & Park, 2016), future additional work on the issue would help researchers to better understand how different learner characteristics moderate the effect of emotional design on learning. Future studies may also relate personalized choice based learning environments literature to emotional design (Rey, 2012; Song & Keller, 2001).

The majority of previous studies on emotional design, including the current study, focused on natural sciences (e.g., work and energy, how a virus causes a cold, immunization) as learning domains. Although incorporating emotional design elements into multimedia may be easier and more applicable for the scientific topics, the element interactivity or difficulty level may be higher in science learning than other domains. Additionally, the results of the current study are pertinent to 7th grade students of a particular middle school. As a result, the findings may therefore be limited to those student groups. Hence, further studies on emotional design could target other learning domains (e.g., social sciences) and different study groups.

In order to increase ecological validity, this current study was conducted within the curriculum for a period of three weeks. Accordingly, randomization of each student to the conditions was not possible which prompted the adoption of a pretest/posttest quasi-experimental design. Future studies could replicate this study in more controlled laboratory conditions by using emWave Emotion Recognition Technology and comparing results to the current study. In such studies, it could also be possible to explore students' Heart Rate Rhythm Patterns instead of average coherence scores. In this way, a more accurate and detailed picture of emotional states of students could be monitored.

Acknowledgements

The financial support of Scientific and Technological Research Council of Turkey (TUBITAK) (Project No: 115K390) is gratefully acknowledged.

References

Astleitner, H., & Wiesner, C. (2004). An integrated model of multimedia learning and motivation. *Journal of Educational Multimedia and Hypermedia, 13*(1), 3–21.
Boyatzis, C. J., & Varghese, R. (1994). Children's emotional associations with colors. *Journal of Experimental Child Psychology, 155*(2), 77–85.
Bradley, R. T., McCraty, R., Atkinson, M., Tomasino, D., Daugherty, A., & Arguelles, L. (2010). Emotion self-regulation, psychophysiological coherence, and test anxiety: Results from an experiment using electrophysiological measures. *Applied Psychophysiology and Biofeedback, 35*(4), 261–283.
Brom, C., Hannemann, T., Stárková, T., Bromová, E., & Děchtěrenko, F. (2016). Anthropomorphic faces and funny graphics in an instructional animation may improve superficial rather than deep learning: A quasi-experimental study. In J. Novotná, & A. Jančařík (Eds.). *Proceedings of the 15th European Conference on e-Learning Excel 2016* (pp. 89–97). Reading: Academic Conferences and Publishing. Retrieved from http://artemis.ms.mff.cuni.cz/main/papers/ecel16_facesWasteWater_160625_camera_ready.pdf.
Case, R. (1985). *Intellectual development from birth to adulthood.* New York, USA: Academic Press.
Chen, C.-M., & Wang, H.-P. (2011). Using emotion recognition technology to assess the effects of different multimedia materials on learning emotion and performance. *Library & Information Science Research, 33*(3), 244–255.
Clark, R. E. (1982). Antagonism between achievement and enjoyment in ATI studies. *Educational Psychologist, 17*(2), 92–101.
Clark, R. C., & Lyons, C. (2010). *Graphics for learning: Proven guidelines for planning, designing, and evaluating visuals in training materials.* San Francisco, USA: John Wiley & Sons.
Clark, R. C., & Mayer, R. E. (2016). *e-Learning and the science of instruction: Proven guidelines for consumers and designers of multimedia learning.* Hoboken, USA: John Wiley & Sons.
Dai, D. Y., & Sternberg, R. J. (2004). Beyond cognitivism: Toward an integrated understanding of intellectual functioning and development. In D. Y. Dai, & R. J. Sternberg (Eds.). *Motivation, emotion, and cognition: Integrative perspectives on intellectual functioning and development* (pp. 3–38). Mahwah, USA: Lawrence Erlbaum Associates Publishers.
Dalacosta, K., Kamariotaki-Paparrigopoulou, M., Palyvos, J. A., & Spyrellis, N. (2009). Multimedia application with animated cartoons for teaching science in elementary education. *Computers and Education, 52*(4), 741–748.
Dong, C. (2007). *Positive emotions and learning: What makes a difference in multimedia design?*Doctoral Dissertation. New York, USA: New York University.
El-Nasr, M. S., Morie, J. F., & Drachen, A. (2010). A scientific look at the design of aesthetically and emotionally engaging interactive entertainment experiences. In D. Gökçay, & G. Yıldırım (Eds.). *Affective computing and interaction psychological cognitive and neuroscientific perspectives* (pp. 281–307). Hershey, PA, USA: Information Science Reference.
Field, A. (2013). *Discovering statistics using IBM SPSS statistics.* London: SAGE.
Fredrickson, B. L. (1998). What good are positive emotions. *Review of General Psychology, 2*(3), 300–319.
Fredrickson, B. L., & Branigan, C. (2005). Positive emotions broaden the scope of attention and thought-action repertoires. *Cognition & Emotion, 19*(3), 313–332.
Green, S. B., & Salkind, N. J. (2005). *Using spss for windows and macintosh: Analyzing and understanding data.* Upper Saddle River, New Jersey: Pearson Prentice Hall.
Haaranen, L., Ihantola, P., Sorva, J., & Vihavainen, A. (2015). In search of the emotional design effect in programming. *Proceedings - 2015 IEEE/ACM 37th IEEE international conference on software engineering (ICSE): Vol. 2*, (pp. 428–434).
Harp, S. F., & Mayer, R. E. (1997). The role of interest in learning from scientific text and illustrations: On the distinction between emotional interest and cognitive interest. *Journal of Educational Psychology, 89*(1), 92–102.
Harp, S. F., & Mayer, R. E. (1998). How seductive details do their damage: A theory of cognitive interest in science learning. *Journal of Educational Psychology, 90*(3), 414–434.
Hasan, Y., Bègue, L., & Bushman, B. J. (2013). Violent video games stress people out and make them more aggressive. *Aggressive Behavior, 39*(1), 64–70.
Heidig, S., Müller, J., & Reichelt, M. (2015). Emotional design in multimedia learning: Differentiation on relevant design features and their effects on emotions and learning. *Computers in Human Behavior, 44*, 81–95.
Inquisit Scripts (2016). *Mr peanut test.* Retrieved from http://www.millisecond.com/download/library/MrPeanut/.
Isen, A. M., Daubman, K. A., & Nowicki, G. P. (1987). Positive affect facilitates creative problem solving. *Journal of Personality and Social Psychology, 52*(6), 1122–1131.
Isen, A. M., & Reeve, J. (2005). The influence of positive affect on intrinsic and extrinsic motivation: Facilitating enjoyment of play, responsible work behavior, and self-control. *Motivation and Emotion, 29*(4), 297–325.
Jager, K. A. (2004). *Informal measures of processing speed and working memory and their relationship to the Wechsler intelligence scale for children-IV*Doctoral dissertation. University of Calgary.
Kallery, M., & Psillos, D. (2004). Anthropomorphism and animism in early years science: Why teachers use. *Research in Science Education, 34*(3), 291–311.
Kaya, N., & Epps, H. (1998). Relationship between color and emotion : A study of college students. *College Student Journal, 38*, 396–406.

Kessels, R. P. C., van den Berg, E., Ruis, C., & Brands, A. M. A. (2008). The backward span of the corsi block-tapping task and its association with the WAIS-III digit span. *Assessment, 15*(4), 426–434.
Knörzer, L., Brünken, R., & Park, B. (2016). Emotions and multimedia learning: The moderating role of learner characteristics. *Journal of Computer Assisted Learning, 32*(6), 618–631.
Kılıç, E., & Karadeniz, Ş. (2004). Hiper ortamlarda öğrencilerin bilişsel yüklenme ve kaybolma düzeylerinin belirlenmesi [specifying students' cognitive load and disorientation level in hypermedia]. *Educational Administration-Theory and Practice, 40*(40), 562–579.
Leutner, D. (2014). Motivation and emotion as mediators in multimedia learning. *Learning and Instruction, 29*, 174–175.
Lidwell, W., Holden, K., & Butler, J. (2003). *Universal principles of design.* Gloucester, UK: Rockport.
Lohr, L. (2007). *Creating graphics for learning and performance: Lessons in visual literacy.* Upper Saddle River, New Jersey: Pearson Prentice Hall.
Mayer, R. E. (2002). Rote versus meaningful learning. *Theory Into Practice, 41*(4), 226–232.
Mayer, R. E. (2009). *Multimedia learning.* New York, USA: Cambridge University Press.
Mayer, R. E. (Ed.). (2014). *The Cambridge handbook of multimedia learning.* New York, USA: Cambridge University Press.
Mayer, R. E. (2014b). Incorporating motivation into multimedia learning. *Learning and Instruction, 29*, 171–173.
Mayer, R. E., & Estrella, G. (2014). Benefits of emotional design in multimedia instruction. *Learning and Instruction, 33*, 12–18.
Mayer, R. E., & Moreno, R. (2010). Techniques that reduce extraneous cognitive load and manage intrinsic cognitive load during multimedia learning. In J. L. Plass, R. Moreno, & R. Brünken (Eds.). *Cognitive load theory* (pp. 131–152). New York, USA: Cambridge University Press.
Milli Eğitim Bakanlığı [Ministry of National Education] (2015). *Öğretim programlarını izleme ve değerlendirme sistemi [Monitoring and evaluation system of the curriculum].* Retrieved from http://mufredat.meb.gov.tr/ProgramDetay.aspx?PID=143.
Moreno, R. (2006). Does the modality principle hold for different media? A test of the method-affects-learning hypothesis. *Journal of Computer Assisted Learning, 22*(3), 149–158.
Moreno, R., & Mayer, R. E. (2000). A coherence effect in multimedia learning: The case for minimizing irrelevant sounds in the design of multimedia instructional messages. *Journal of Educational Psychology, 92*(1), 117–125.
Morra, S. (1994). Issues in working memory measurement: Testing for m capacity. *International Journal of Behavioral Development, 17*(1), 143–159. http://doi.org/10.1177/016502549401700109.
Morra, S., & Camba, R. (2009). Vocabulary learning in primary school children: Working memory and long-term memory components. *Journal of Experimental Child Psychology, 104*(2), 156–178.
Mueller, S. T., & Piper, B. J. (2014). The psychology experiment building language (PEBL) and PEBL test battery. *Journal of Neuroscience Methods, 222*, 250–259.
Nasby, W., & Yando, R. (1982). Selective encoding and retrieval of affectively valent information: Two cognitive consequences of children's mood states. *Journal of Personality and Social Psychology, 43*(6), 1244–1253.
Oaksford, M., Morris, F., Grainger, B., & Williams, J. M. G. (1996). Mood, reasoning and central executive processes. *Journal of Experimental Psychology, 22*(2), 476–492.
Paas, F., & Sweller, J. (2014). Implications of cognitive load theory for multimedia learning. In R. E. Mayer (Ed.). *The cambridge handbook of multimedia learning* (pp. 27–42). New York, USA: Cambridge University Press.
Paas, F., & Van Merriënboer, J. J. G. (1993). The efficiency of instructional conditions: An approach to combine mental effort and performance measures. *Human Factors, 35*(4), 737–743.
Pallant, J. (2010). *SPSS survival manual: A step by step guide to data analysis using SPSS.* Berkshire, UK: Open University Press.
Park, S. (2005). *The effects of seductive augmentation and agent role on learning interest, achievement, and attitude*Doctoral Dissertation. Tallahassee, Florida: Florida State University.
Park, B. (2010). *Testing the additivity hypothesis of cognitive load theory*Doctoral Dissertation. Saarbrücken: Universität des Saarlandes.
Park, B., Flowerday, T., & Brünken, R. (2015a). Cognitive and affective effects of seductive details in multimedia learning. *Computers in Human Behavior, 44*, 267–278.
Park, B., Knörzer, L., Plass, J. L., & Brünken, R. (2015b). Emotional design and positive emotions in multimedia learning: An eyetracking study on the use of anthropomorphisms. *Computers & Education, 86*(C), 30–42.
Pekrun, R., Frenzel, A. C., Goetz, T., & Perry, R. P. (2007). The control-value theory of achievement emotions: An integrative approach to emotions in education. In P. A. Schutz, & R. Pekrun (Eds.). *Emotion in education* (pp. 13–36). San Diego, CA, USA: Academic Press.
Pintrich, P. R. (2000). The role of goal orientation in self-regulated learning. In M. Boekaerts, P. R. Pintrich, & M. Zeidner (Eds.). *Handbook of self-regulation* (pp. 451–502). New York, USA: Academic Press.
Plass, J. L., Heidig, S., Hayward, E. O., Homer, B. D., & Um, E. (2014). Emotional design in multimedia learning: Effects of shape and color on affect and learning. *Learning and Instruction, 29*, 128–140.
Plass, J. L., Kalyuga, S., & Leutner, D. (2010). Individual differences and cognitive load theory. In J. L. Plass, R. Moreno, & R. Brünken (Eds.). *Cognitive load theory* (pp. 65–87). New York, USA: Cambridge University Press.
emWave Pro Library (2016). *Science: Heart science overview.* Retrieved from http://cdn.quantumintech.com/emwave-pro-library/science_overview.html.
Rey, G. D. (2012). A review of research and a meta-analysis of the seductive detail effect. *Educational Research Review, 7*(3), 216–237.
Rozman, D., McCraty, R., & Tomasino, D. (2008). It's heart rhythm not rate that counts: Heartmath studies and emwave PC stress relief system. In J. H. D. M. Westerink, M. Ouwerkerk, T. J. M. Overbeek, W. F. Pasveer, & B. de Ruyter (Vol. Eds.), *Probing experience: Vol. 8*, (pp. 195–204). Dordrecht: Springer Netherlands.
Salomon, G. (1984). Television is “easy” and print is “tough”: The differential investment of mental effort in learning as a function of perceptions and attributions. *Journal of Educational Psychology, 76*(4), 647–658.
Sanchez, C. A., & Wiley, J. (2006). An examination of the seductive details effect in terms of working memory capacity. *Memory & Cognition, 34*(2), 344–355.
Seibert, P. S., & Ellis, H. C. (1991). Irrelevant thoughts, emotional mood states, and cognitive task performance. *Memory & Cognition, 19*(5), 507–513.
Sitzmann, T., & Johnson, S. (2014). The paradox of seduction by irrelevant details: How irrelevant information helps and hinders self-regulated learning. *Learning and Individual Differences, 34*, 1–11.
Song, S. H., & Keller, J. M. (2001). Effectiveness of motivationally adaptive computer-assisted instruction on the dynamic aspects of motivation. *Educational Technology Research & Development, 49*(2), 5–22.
Stevens, J. (2009). *Applied multivariate statistics for the social sciences.* New York, USA: Routledge.
Thalheimer, W. (2004). *Bells, whistles, neon, and purple prose : When interesting words, sounds, and visuals hurt learning and performance-a review of the seductive-augmentation research.* Somerville, Massachusetts: Work-Learning Research. Retrieved from http://willthalheimer.typepad.com/files/review_seductive_augmentations_pdf6.pdf.
Um, E. Rachel, Plass, J. L., Hayward, E. O., & Homer, B. D. (2012). Emotional design in multimedia learning. *Journal of Educational Psychology, 104*(2), 485–498.
Waters, G. S., & Caplan, D. (2003). The reliability and stability of verbal working memory measures. *Behavior Research Methods, Instruments, & Computers, 35*(4), 550–564.
Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative Affect: The PANAS scales. *Journal of Personality and Social Psychology, 54*, 1063–1070. http://doi.org/. https://doi.org/10.1037/0022-3514.54.6.1063.
Watts, M., & Bentley, D. (1991). *An instructional “man-trap”: Anthropomorphic and animistic thought in constructivist science education.* Retrieved from http://files.eric.ed.gov/fulltext/ED336269.pdf.
Watts, M., & Bentley, D. (1994). Humanizing and feminizing school science: Reviving anthropomorphic and animistic thinking in constructivist science education. *International Journal of Science Education, 16*(1), 83–97.