The Educational Games Panacea: Measuring Engagement Levels for Educational Games vs. Traditional Text Literature using a Wireless EEG Headset

Marvin Andujar[1], Josh Ekandem[2], Ignacio Alvarez[2], Melva James[2], Jonathan Cox[2], Juan Gilbert[2]

[1]Computer Science Department
Kean University
Union, NJ 07083
andujarm@kean.edu

[2]School of Computing
Human-Centered Computing (HCC) Lab
Clemson University
Clemson, SC 02631
{jekande, ignm, melvaj, jlcox, juan}@clemson.edu

Abstract

This paper examines the correlation of engagement measurements with knowledge increment of the two different modalities: gaming and the traditional paper reading learning styles. The studies were performed using the non-invasive electroencephalographic (EEG) brain-computer interface Emotiv EPOC device. The Emotiv headset, despite its limitations is widely used in consumer BCI applications nowadays, which hopes improvement in the near future. This paper reports a controlled experiment comparing the effective learning of gaming with the traditional paper reading.

In this paper, the educational system and brain-computer interface (BCI) are combined to measure engagement levels and test the participant's knowledge regarding the Lewis and Clark expedition adventure. The results led us to see how the traditional way of learning and the game learning style are not so different in engagement, but also how the traditional way is still more useful than learning by playing games. Furthermore, this discusses the analysis of the device and how it helps us to distinguish the learning performance between users.

1. Introduction

Educational video games claim to be better at attracting students attention and keeping them engaged in the learning process than the traditional text learning styles. This paper investigates the engagement levels produced by educational video games and their effect on the retained information using Brain-Computer Interface (BCI) tools. BCI is a neuroscience paradigm that helps us to understand the functionality of the human brain and how it is able to communicate with machines through neuroheadsets. BCIs consist of three major components which are: the ways of measuring neural signals from the human brain, algorithms for decoding brain states, and methodologies for mapping the decoded brain activity for the intended action or behavior [4].

Studies have been conducted by researchers from various institutions of how games are more engaging and useful than the traditional learning style, but so far there has only been assumptions of how the users are engaged and no objective data has been provided to support these assumptions. Using the Emotiv EPOC we conducted studies to measure the engagement levels and post survey to gauge information retention. Our experiment comprised of a control group that performed the handout task and a experimental group which performed the video game task. Both populations were learning the American History of Lewis and Clark Expedition Adventure.

The game section is an online game from the web site named "Class Brain Games". This game contains a lot of important and valuable information about the expedition. The information of this game was extracted and transferred to a handout, so both populations would learn the same piece of information.

2. Method

This section describes how the experiment was designed and describes all the steps taken to complete the studies, how the participants were walked through the studies, and how the data was gathered.

2.1 Subjects

The sample size for our studies is 26 participants aging from 19 – 25 years old. All subjects could read English fluently and were comfortable with reading any type of text. All the participants read and agreed with the consent form before participating in the studies. Upon agreement with the consent form, a pre-assessment was given to determine their previous knowledge regarding the Lewis and Clark Expedition. Once the assessment was completed, the participants from both groups wore the non-invasive Emotiv EPOC device while performing their perspective tasks. After the learning section, the participants were given a post-assessment and a 10 questions quiz to determine the acquired knowledge. The pre and post assessment questions regarding their knowledge of the expedition ranged from 0 meaning knowing nothing at all and 10 meaning have a lot of knowledge.

2.2 Experimental Design

Participants were divided into two different groups, the experimental group, which received information via the PC online game. The game was chosen from (www.classbraingames.com), it was chosen, so both genders would feel engaged and despite their background they are able play the game. The other group was the control, which received information in the form of a handout. Both groups learned the history of Lewis and Clark Expedition. The game group played a series of mini games, once each mini game section was completed, information regarding what they just played would transmit the educational content. The handout population was just reading the same information located in the game, but in the handout they were no pictures or any other visuals. Each population had 20 minutes to complete the task.

While the participants were learning the Emotiv was capturing the engagement levels according to their connectivity with the scalp, reaction to the reading, and their physical movement such as: scratch their head, raise eyebrows, fall sleep, laugh, or frustration. Each participant had different movements, reaction and scalp connectivity.

2.3 Data Gathering

During the learning process, the engagement levels were recorded using the EEG Emotiv EPOC device. The algorithm used to capture and measure engagement is proprietary to the vendor.

2.4 Capturing Engagement

Each population's engagement levels were captured every five minutes for a period of 20 minutes. The control panel established a connection between the device and the computer. The test bench was used to provide additional visual information as to the amplitude of the signal from each electrode while recording the signal. The most important software used was the modified emostatelogger, this software interacts with an application written in C++ using Visual Studio for compilation. The engagement levels were calculated, and the averages of every 5, 10, 15, and 20 minutes were saved into a text file called log. In the studies, the engagement range was from 0.550000 (min) to 0.684000 (max).

3. Wireless EEG Headset

The Emotiv EPOC used in the studies is a wireless EEG data acquisition and processing device. The device uses 14 electrodes to obtain the EEG signal and these channels are based on the international 10-20 locations which are: AF3, F7, F3, FC5, T7, P7, O1, O2, P8, T8, FC6, F4, F8, AF4. It also has 2 reference electrodes in the P3 and P4 locations. In addition the Emotiv Epoc has a built in gyroscope that detects the change in the x and y position of the device [3]. It connects wirelessly to Windows operating systems machines. In the machines a Bluetooth is connected via USB to receive signal from the device.

3.1 Device Limitations

The Emotiv headset does not work well with users with long and thick hair. The sensors work best when in direct contact with the scalp. When not on the scalp the sensor connectivity is not as strong as if it were directly on the scalp. In light of this limitation, engagement and excitement levels could be recorded in our experiment. Another limitation noticed in the studies is the shape and size of the head. If the user has a small head, then it is harder for the two reference electrodes to make contact with the head. Both reference electrodes end up being in the air in which it is necessary for them to make contact. By adjusting the device more, the two references are able to make contact with the head, but the front channels end up more to the front.

4. Results

The results obtained are classified into different variables that had to be addressed because of the device limitations. According to our results and self-assessment by the participants the traditional way of learning is still more useful than a game for the population between 19 – 25 years old. The sample size of the studies was 13 each population with a total of 26 participants. For the game task, there was a total of seven males and six females and for the handout task; there were a total of eight males and five females. When analyzing this data, we also took in consideration the BCI device limitations. The difference between limitations will be discussed with more details in the discussion section.

4.1 Statistical Analysis

We performed a statistical analysis to determine if there is a significance difference between the gaming and the text learning styles. The statistical software R was used to perform the statistical analysis. R developed at Bell laboratories is a programming language and a software environment to perform statistical computations.

A two-tailed T test using pooled variance was calculated using the formula from illustration 1. Our t-test value is equal to -0.017273. After, the t value was obtained we found our degree of freedom which is equal to 24 and the significance level is .025. Using the degree of freedom and the significance level we found the value in the T table in which is 2.06390. Our t-test value -.01727305 in which is less than the t-table value which is 2.06390, by this results, then we fail to reject the null hypothesis which tells us that there are no statistical difference between the two groups

After, obtaining our results we confirmed it by finding the two-tailed p-value in which is equal to 0.9864. By conventional criteria, this difference is considered to be not statistically significant because 0.9864 is greater than the significance level in which is equal to .05. This also supports that there are no difference between the means of the game learners and the handout learners.

$$t = \frac{(\bar{X}_{game} - \bar{X}_{handout})}{\sqrt{{S_p}^2(\frac{1}{n_{game}} + \frac{1}{n_{handout}})}}$$

Illustration 1: T value formula for engagement levels

4.2 Game vs. Handout

Groups	N	Engagement Levels	STD	Test Scores	STD	Knowledge Increment		
Game	13	0.566889	0.037924	58.46	13.45	Pre: 2.00	Post: 5.85	Diff: 3.85
Handout	13	0.572727	0.042919	66.92	15.48	Pre: 2.08	Post: 6.08	Diff: 4.00

Table 1: Game vs. Handout: Engagement levels and test scores correlation

As shown in table 1 the handout engagement is greater than the game by 0.005838, the test score is greater by 8.46 and the increment in knowledge is greater by 0.15. By looking at table 1, we can see a correlation between self-evaluation in knowledge increment, engagement levels, and test scores.

4.3 Gender Classification

Gender	Engagement Levels	STD	Test Scores	STD	Knowledge Increment		
Female	0.551710	0.000975	63.33	13.66	Pre: 0.67	Post: 6.17	Diff: 5.50
Male	0.579900	0.049478	54.29	12.72	Pre: 3.14	Post: 5.57	Diff: 2.43

Table 2: Game Task by Gender

The table 2 shows the performance between genders and the correlation with their knowledge progress in the game task. We can observe that the engagement levels for the male were higher than the female by a total mean of 0.028190, but the female obtained a higher score in the test with a difference of total mean of 9.04. According to the self-assessment the female knowledge increased more than the male's knowledge. The female's knowledge increased by a total mean of 5.5 when for the male was a total mean of 2.43. In this task, the female population performed better and gained more knowledge.

Gender	Engagement Levels	STD	Test Scores	STD	Knowledge Increment		
Female	0.552359	0.000940	66.00	5.48	Pre: 2.60	Post: 6.80	Diff: 4.20
Male	0.585620	0.051682	67.50	19.82	Pre: 1.75	Post: 5.63	Diff: 3.88

Table 3: Handout Task by Gender

The table 3 shows the performance between genders and the correlation with their knowledge progress in the handout task. We can observe that the engagement levels for the male were higher than the female in this task as well by a total mean of 0.03326 and the test scores were higher as well by a total mean of 1.5. We can also observe the female's knowledge increased by a total mean of 4.20 and the male increased by 3.88.

4.4 Gender Performance Between Tasks

By comparing both tables 2 and 3 we can see how the female population performed in each task. It is clearly stated that the female performed better in the handout than in the game. In the handout, the engagement levels were higher by 0.000649, the test scores were higher by 2.67, and the knowledge growth is higher. Also, it is clearly stated for the males that they performed better in the handout as well.

4.5 Ethnicity Classification

This section shows how each ethnicity categories performed in each tasks. Besides the BCI device limitations we run into another limitation regarding ethnicities. As the studies were conducted during the summer at Clemson University, Clemson, SC, we were not able to have any Hispanic participants, as the number of Hispanics in this school is very low. According to the Clemson University data set, there is a total of 19, 453 students enrolled in the university and there is only 304 undergraduate Hispanic students enrolled. The school does not provide any data for the graduate students by ethnicity.

4.6 Ethnicity Performance by task

	Engagement Levels	STD	Test Scores	STD	Knowledge Increment		
African-American	0.591556	0.055481	52.00	8.37	Pre: 1.80	Post: 5.80	Diff: 4.00

Caucasian	0.551424	0.000777	68.33	11.69	Pre: 2.17	Post: 5.50	Diff: 3.33
Asian	0.551618	0.000985	45.00	7.07	Pre: 2.00	Post: 7.00	Diff: 5.00
Hispanic	N/A	N/A	N/A	N/A	N/A		

Table 4: Ethnicity Performance: Game Task

In table 4 the sample size is five African-Americans, six Caucasians, and two Asians with a total of 13 participants. From this table we can see the engagement level of the African-Americans are higher than the others, but the Caucasian had the highest score in the test even though having the lowest engagement and lowest knowledge increment. We deduced from these results that the African-American engagement levels were higher because the Emotiv had direct contact their scalp. The amount of hair that most of this ethnicity has are very low. The Caucasians and Asians have more hair on their scalp in which prevent the emotive to obtain higher engagement levels.

Another deduction we made was, even though the knowledge increment of the Caucasians was the lowest, but they knew the most previous the game, which their previous knowledge help them, achieve a higher score. The Asian population had the biggest knowledge increment, because they do not know a lot of the American History previous to the game.

	N	Engagement Levels	STD	Test Scores	STD	Knowledge Increment		
African-American	5	0.586720	0.057741	60.00	18.71	Pre: 1.40	Post: 6.00	Diff: 4.60
Caucasian	6	0.568237	0.036867	75.00	10.49	Pre: 2.83	Post: 6.17	Diff: 3.34
Asian	2	0.551868	0.001664	60.00	14.14	Pre: 1.50	Post: 6.00	Diff: 4.50
Hispanic	0	N/A	N/A	N/A	N/A	N/A		

Table 5. Ethnicity Performance: Handout Task

In table 5 we can make a similar conclusion as the game. The African-Americans were able to have the highest engagement because of their device connectivity and they were anxious to learn more about the expedition as we can see their previous knowledge was the lowest between the ethnicities. The Caucasians still scored more in the handout and knew more about the expedition just as the game population as well.

By comparing both tables we can still see clearly that by ethnicities the handout still have an advantage. For the African-Americans, even though their engagement levels were higher in the game, their test scores and knowledge increments were higher in the handout. The Caucasians had a higher engagement, test scores, and knowledge increase in the handout than in the game. The Asians had also a higher engagement and test scores in the handout, despite in the game the knowledge increment was only higher by 0.5.

Discussion

Taking in mind the data analysis of the previous section, notwithstanding the data shows that the engagement levels, test scores, and the growth in knowledge are higher in the traditional text learning style, there is not statistical difference. When Brain-Computer Interface is used for measuring the performance of two different groups, all the different possibilities has to be taken into account.

- Some of the possibilities are: what was the mood of the person prior starting the task?
- Were all the channels of the device receiving perfect connectivity?
- What was the hair texture of the participant?

	Green	Yellow	Orange	Red	Black
PG001	10	1	0	0	5
PG009	2	2	9	0	3
PH002	13	0	0	1	2
PH003	10	0	1	1	3
PH004	2	8	3	0	3
PH012	3	3	4	2	3

Table 6. Ending Signals after Task

As explained in the limitation of the device the hair texture and its amount are very important. Most of the participants had a lot of hair on their scalp in which made it harder to obtain signal. Only six participants from both groups had perfect connection, which means all the channels were receiving perfect signal. The rest did not have perfect signal, but for some of them after certain time while performing the task some of the signals lighten up. As shown in table 6, there was six participants from both populations where their signals started changing throughout the task and ended up with the signals shown in the table. The colors mean the strength of the connectivity between the user and the channels of the device. Green meaning perfect signal, yellows means very good signal, orange means poor signal, red means very poor signal, and black means no signal at all. We can see from the table that each participant's connectivity varies. It depends on the amount of hair they have, their brain waves, head shape, and head size. A deduction made was that if a participant have a lot of hair, there will always be a black signal, because there will be spots where the channels do not have direct contact with the scalp.

7. Conclusion/Future Work

A lot of questions and doubts have emerged which we will pursue for future work. The data shows us that there is not statistical difference in the engagement, knowledge retention, and knowledge growth between the game and the traditional text learning styles. By studying, the different scenarios encountered during the studies the next step is to study the brain waves of each individual with special situations while performing the studies, to determine if there is a correlation.

8. References

[1] A. Campbell, T. Choudhury, S. Hu, H. Lu, M. Mukerjee, M. Rabbi, and R. Raizada. NeuroPhone: Brain-Mobile Phone Interface using a Wireless EEG Headset. MobiHeld 2010, August 30, 2010, New Delhi, India.

[2] P. Bobrov, A. Frolov, C. Cantor, I. Fedulova, M. Bakhnyan, A. Zhavoronkov. Brain-Computer Interface Based on Generation of Visual Images. PLoS ONE 6(6): e20674. Doi:10.1371/journal.pone.0020674

[3] Emotiv Systems. Emotiv – brain computer interface technology. http://www.emotiv.com.

[4] A. Cichocki, Y. Washizawa, T. Rutkowski, H. Bakardjian, A. Phan, S. Choi, H. Lee, Q. Zhao, L. Zhang, and Y. Li. Noninvasive BCIs: Multiway Signal-Processing Array Decompositions. IEEE Computer Society. Pg. 37- 42. October, 2008.

[5] Clemson University Data Set. http://www.clemson.edu/oirweb1/FB/factBook/CommonDataSet2010.html.

[6] Statistical Software R. http://www.r-project.org/.

[7] G. Sindre, L. Natvig, and M. Jahre. Experimental Validation of the Learning Effect for a Peadgogical Game on Computer Fundamentals. IEEE Transactions on Education, vol. 52, no. 1, Pg. 10-18, February, 2009.